



HAL
open science

An adaptive test for zero mean

Cécile Durot, Yves Rozenholc

► **To cite this version:**

Cécile Durot, Yves Rozenholc. An adaptive test for zero mean. *Mathematical Methods of Statistics*, 2006, 15 (1), pp.26-60. hal-00748951

HAL Id: hal-00748951

<https://hal.science/hal-00748951v1>

Submitted on 6 Nov 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AN ADAPTIVE TEST FOR ZERO MEAN

C. DUROT¹ AND Y. ROZENHOLC²

¹Laboratoire de Mathématiques, Université Paris Sud
91405 Orsay cedex, France
E-mail: cecile.durot@math.u-psud.fr

²MAP5 – UMR CNRS 8145 – Université Paris 5
45, rue des Saints-Pères, 75270 Paris cedex 06, France
E-mail: yves.rozenholc@math-info.univ-paris5.fr

Assume we observe a random vector y of \mathbb{R}^n and write $y = f + \varepsilon$, where f is the expectation of y and ε is an unobservable centered random vector. The aim of this paper is to build a new test for the null hypothesis that $f = 0$ under as few assumptions as possible on f and ε . The proposed test is nonparametric (no prior assumption on f is needed) and nonasymptotic. It has the prescribed level α under the only assumption that the components of ε are mutually independent, almost surely different from zero and with symmetric distribution. Its power is described in a general setting and also in the regression setting, where $f_i = F(x_i)$ for an unknown regression function F and fixed design points $x_i \in [0, 1]$. The test is shown to be adaptive with respect to Hölderian smoothness in the regression setting under mild assumptions on ε . In particular, we prove adaptive properties when the ε_i 's are not assumed Gaussian nor identically distributed.

Key words: adaptive test, minimax hypothesis testing, nonparametric alternatives, symmetrization, heteroscedasticity.

2000 Mathematics Subject Classification: 62G10, 62G08.

1. Introduction

Assume we observe a random vector $y \in \mathbb{R}^n$ and write $y = f + \varepsilon$, where f is the unknown expectation of y and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$ is an unobservable random vector with mean zero. Assume the ε_i 's are mutually independent with symmetric distribution (we mean that for every i , ε_i has the same distribution as $-\varepsilon_i$). Our aim is to build a nonasymptotic test for the null hypothesis $H_0: f = 0$ against the nonparametric composite alternative $H_1: f \neq 0$, when no prior assumption is made on f and as few as possible assumptions are made on ε . In particular, we do not

©2006 by Allerton Press, Inc. Authorization to photocopy individual items for internal or personal use, or the internal or personal use of specific clients, is granted by Allerton Press, Inc. for libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that the base fee of \$50.00 per copy is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923.

assume *a priori* that f belongs to a given smoothness class, and the ε_i 's are not assumed identically distributed nor Gaussian. The proposed test is defined as a multi-test and is based on a symmetrization principle that exploits the symmetry assumption.

The problem of hypothesis testing in the general model $y = f + \varepsilon$ has rarely been addressed in the literature. Baraud *et al.* (2003) consider the problem of testing the null hypothesis that f belongs to a given linear subspace of \mathbb{R}^n . No prior information on f is required, but they assume that the components of ε are independent with the same Gaussian distribution. Under the same assumptions, Baraud *et al.* (2005) propose a test for the null hypothesis that f belongs to a given convex subset of \mathbb{R}^n .

An interesting particular case of the general model above is the regression model, which is obtained in the case where $f_i = F(x_i)$ for all $i = 1, \dots, n$. Here, the x_i 's are nonrandom design points and F is an unknown function. In this context, many tests have been proposed for the null hypothesis that F belongs to a given set \mathcal{F} against a nonparametric alternative. Typically, \mathcal{F} is a parametric set of functions or \mathcal{F} is restricted to a single function, which amounts to $\mathcal{F} = \{0\}$. These tests are often based on a distance between a nonparametric estimator for F and an estimator for F that is computed under the null hypothesis, see, e.g., Müller (1992). This requires the choice of a smoothing parameter such as a bandwidth. Other tests consider the smoothing parameter itself as a test statistic, see, e.g., Eubank and Hart (1992). Another approach consists in building a test statistic as a function of estimators for the Fourier coefficients of F , see, e.g., Chen (1994). We refer to the book of Hart (1997) for a review of these methods.

Recently, the problem of adaptive minimax testing has been addressed. Suppose that the null hypothesis is $F \equiv 0$ and consider the alternative that F is bounded away from zero in the L_2 -norm, $\|F\|_2 \geq \rho(n)$, and possesses smoothness properties. The minimal rate of testing (that is the minimal distance $\rho(n)$ for which testing with prescribed error probabilities is still possible) has been first derived in a white noise model with signal F . The result depends heavily on the kind of smoothness imposed, see Ingster (1982, 1993) and Ermakov (1991) for Sobolev smoothness and Lepski and Spokoiny (1999) for Besov smoothness. The optimal rate and the structure of optimal tests depend on the smoothness parameters, whereas these parameters are usually unknown in practical applications. In the white noise model, Spokoiny (1996) proves that the minimal rate of testing for $H_0: F \equiv 0$ is altered by a $\log \log n$ factor when F belongs to a Besov functional class with unknown parameters. He builds a rate optimal adaptive test based on wavelets. Gayraud and Pouet (2005) obtain similar results in a regression model for a composite null hypothesis under Hölderian smoothness. They prove that, in the Gaussian model, the optimal rate of testing is altered by a $\log \log n$ factor if the smoothness parameter is unknown. They build an adaptive test which achieves the optimal rate over a class of Hölderian functions with smoothness parameter $s > 1/4$ in a possibly non-Gaussian model. Other examples of adaptive tests are given by Baraud *et al.* (2003), Härdle and Kneip (1999) and by Horowitz and Spokoiny (2001). All these adaptive tests are defined as multi-tests. Roughly speaking, the authors first build a test T_s that is minimax for a fixed smoothness parameter s and reject the null hypothesis if there exists s in a given grid such that T_s rejects.

Most of non-adaptive tests for zero mean involve a smoothing parameter which is chosen in a somewhat arbitrary way (Staniswalis and Severini, 1991). Adaptive tests do not show this drawback but either they require the errors to be i.i.d. Gaussian (Härdle and Kneip, 1999, Baraud *et al.*, 2003) or they are asymptotic (Horowitz and Spokoiny, 2001, Gayraud and Pouet, 2005). Moreover, existing tests involve estimation of the unknown variance. This requires either homoscedasticity (Gayraud and Pouet, 2005) or regularity assumptions on the variance (Horowitz and Spokoiny, 2001). On the contrary, our test requires very mild assumptions on the errors, is nonasymptotic and needs neither variance estimation nor arbitrary choice of a parameter. It has the prescribed level under the only assumption that the ε_i 's are mutually independent with symmetric distributions (the ε_i 's may have different distributions). Moreover, it achieves the optimal rate of testing over the class of Hölderian functions with smoothness parameter $s > 1/4$ in the case where the ε_i 's satisfy a Bernstein-type condition, and in particular, in the homoscedastic Gaussian case. The test still achieves the optimal rate of testing for $s \geq 1/4 + 1/p$ in the case where the ε_i 's possess bounded moments of order $2p$ for some $p \geq 2$.

The paper is organized as follows. The testing procedure is described in Section 2. It is also stated in this section that the proposed test has the prescribed nonasymptotic level. In Section 3, we discuss implementation of the test. The power is studied in Section 4 under various assumptions on the ε_i 's. In Section 5, we compute the rate of testing of the test in a regression model under a Hölderian assumption. A simulation study is reported in Section 6 and the proofs are given in Section 7.

2. The Testing Procedure

Assume we observe a random vector y of \mathbb{R}^n and write

$$y = f + \varepsilon,$$

where f is an unknown vector of \mathbb{R}^n and ε is an unobservable random vector with mean zero. Assume that the components of ε are independent and possess a symmetric distribution around zero, which means that for all i , ε_i and $-\varepsilon_i$ have the same distribution. Assume furthermore that for all i , ε_i is almost surely different from zero. Our aim is to build a test with nonasymptotic level α for the hypothesis $H_0: f = 0$ against $H_1: f \neq 0$. Here, α is a fixed number in $(0, 1)$. The test is based on a symmetrization principle that exploits the symmetry assumption. Before describing the test more precisely, let us introduce some notation.

Notation:

- For every set A , let $|A|$ denote the cardinality of A and let $\mathbf{1}_A$ be the indicator function of A , which means that $\mathbf{1}_A(x)$ equals 1 if $x \in A$ and 0 otherwise.
- Let w be a random vector of \mathbb{R}^n , independent of y , with independent components w_i distributed as random signs: $P(w_i = 1) = P(w_i = -1) = 1/2$.
- For all $u, v \in \mathbb{R}^n$, let $u \times v$ be the vector of \mathbb{R}^n with i th component $(u \times v)_i = u_i v_i$.
- Let $\|\cdot\|_n$ be the Euclidean norm in \mathbb{R}^n .
- For every partition m of $\{1, \dots, n\}$ into D_m nonempty subsets $e_{m,1}, \dots, e_{m,D_m}$, let $t_{m,j}$ ($j = 1, \dots, D_m$) be the vector of \mathbb{R}^n with i th component $\mathbf{1}_{e_{m,j}}(i)$ and let

Π_m be the orthogonal projector onto the linear span of $\{t_{m,1}, \dots, t_{m,D_m}\}$. Thus for all $u \in \mathbb{R}^n$ and $i = 1, \dots, n$

$$(\Pi_m u)_i = \sum_{j=1}^{D_m} \left(\frac{1}{|e_{m,j}|} \sum_{k \in e_{m,j}} u_k \right) \mathbf{1}_{e_{m,j}}(i), \quad \|\Pi_m u\|_n^2 = \sum_{j=1}^{D_m} \frac{1}{|e_{m,j}|} \left(\sum_{k \in e_{m,j}} u_k \right)^2.$$

The test is based on the following heuristics. Under H_0 , y has the same distribution as $w \times y$. Hence for every partition m of $\{1, \dots, n\}$, $\|\Pi_m y\|_n^2$ and $\|\Pi_m(w \times y)\|_n^2$ have the same distribution. Under H_1 , as $f \neq 0$, consider an (unobservable) partition m such that for every $e \in m$ the numbers $(f_i)_{i \in e}$ all have the same sign. Then for every $e \in m$, $|\sum_{i \in e} f_i| \geq |\sum_{i \in e} w_i f_i|$ with a strict inequality if $w_i = -1$ and $f_i \neq 0$ for some $i \in e$, hence one has $\|\Pi_m f\|_n^2 > \|\Pi_m(w \times f)\|_n^2$ provided $w_i = -1$ and $f_i \neq 0$ for some $i \in \{1, \dots, n\}$. Thus under H_1 there exists m such that $\|\Pi_m y\|_n^2$ tends to be larger than $\|\Pi_m(w \times y)\|_n^2$, whereas under H_0 , $\|\Pi_m y\|_n^2$ and $\|\Pi_m(w \times y)\|_n^2$ are of the same order of magnitude for every m . Therefore, we propose to reject H_0 if there exists m such that $\|\Pi_m y\|_n^2$ exceeds a given quantile of $\|\Pi_m(w \times y)\|_n^2$. Since the quantiles of $\|\Pi_m(w \times y)\|_n^2$ cannot be computed (they depend on the unknown distribution of y) we consider conditional quantiles given y .

The precise construction of the test is as follows. Consider a collection of partitions \mathcal{M} and positive numbers α_m with $\sum_{m \in \mathcal{M}} \alpha_m = \alpha$. We reject H_0 if there exists $m \in \mathcal{M}$ such that $\|\Pi_m y\|_n^2$ exceeds the y -conditional quantile of $\|\Pi_m(w \times y)\|_n^2$ defined by

$$q_m^y(\alpha_m) = \inf \left\{ x \in \mathbb{R}, P \left[\|\Pi_m(w \times y)\|_n^2 > x \mid y \right] \leq \alpha_m \right\}.$$

The critical region of our test is thus

$$(1) \quad \sup_{m \in \mathcal{M}} \left\{ \|\Pi_m y\|_n^2 - q_m^y(\alpha_m) \right\} > 0.$$

Note that $q_m^y(\alpha_m)$ can theoretically be computed since it only depends on the known distribution of w , but its exact calculus requires about 2^n computations which cannot be performed in practice. Hence we estimate it through Monte-Carlo simulations, see Section 3. It is stated in the following theorem that the test has a nonasymptotic level α .

Theorem 1. *Assume we observe $y = f + \varepsilon$, where $f \in \mathbb{R}^n$ and the ε_i 's are independent random variables with symmetric distribution. Assume $P(\varepsilon_i = 0) = 0$ for all $i = 1, \dots, n$. Let \mathcal{M} be a collection of partitions of $\{1, \dots, n\}$, let α and $(\alpha_m)_{m \in \mathcal{M}}$ be positive numbers such that $\alpha = \sum_{m \in \mathcal{M}} \alpha_m$. Then*

$$P_{H_0} \left(\sup_{m \in \mathcal{M}} \left\{ \|\Pi_m y\|_n^2 - q_m^y(\alpha_m) \right\} > 0 \right) \leq \alpha.$$

Remarks.

- Like the adaptive tests mentioned in the introduction, our test is a multi-test: the null hypothesis is rejected if one of the tests with critical regions $\{\|\Pi_m y\|_n^2 >$

$q_m^y(\alpha_m)$ rejects. A given partition m allows us to detect alternatives with a given smoothness s , and adaptive properties arise from the use of many partitions.

- Our method allows us to test the more general null hypothesis $H_0: f \in V$, where V is a linear subspace of \mathbb{R}^n , but for this problem, we need the ε_i 's to be i.i.d. Gaussian. Let V be a linear subspace of \mathbb{R}^n with dimension $k < n$, let Π be the orthogonal projector onto V^\perp and let Δ be the diagonal matrix of which $(n - k)$ first diagonal components equal 1 and the others are zero. There exists an orthogonal matrix O with $\Pi = O^T \Delta O$. Now, let y_V be the random vector that consists of the $(n - k)$ first components of $\Delta O y = O \Pi y$, let f_V be the expectation of y_V and let $\varepsilon_V = y_V - f_V$. If the ε_i 's are i.i.d. Gaussian, so are the components of ε_V . Since testing $f \in V$ amounts to testing $f_V = 0$ within the model $y_V = f_V + \varepsilon_V$, this may be done by applying our method. Then the theoretical results we obtain in the i.i.d. Gaussian model for $H_0: f = 0$ (see Sections 4 and 5) can be generalized to $H_0: f \in V$. We do not detail these results.

3. Practical Implementation

By definition, $q_m^y(\alpha_m)$ is the $1 - \alpha_m$ quantile of the discrete distribution with support $\{\|\Pi_m(u \times y)\|^2, u \in \{-1, 1\}^n\}$, which puts mass $k2^{-n}$ at a given point x of the above set if there exist k vectors $u \in \{-1, 1\}^n$ with $\|\Pi_m(u \times y)\|^2 = x$. The exact calculation of $q_m^y(\alpha_m)$ thus requires about 2^n computations, which cannot be performed in practice. We thus suggest estimation of this quantile instead of computing its exact value. The practical implementation of the test then is as follows. Draw independent random vectors w^1, \dots, w^B which all are distributed like w and are independent of y . For every $m \in \mathcal{M}$, compute the empirical quantile defined by

$$\hat{q}_m^B(\alpha_m) = \inf \left\{ x \in \mathbb{R}, \frac{1}{B} \sum_{b=1}^B \mathbf{1}_{\|\Pi_m(w^b \times y)\|_n^2 > x} \leq \alpha_m \right\}$$

and reject H_0 if

$$(2) \quad \sup_{m \in \mathcal{M}} \{ \|\Pi_m y\|_n^2 - \hat{q}_m^B(\alpha_m) \} > 0.$$

It is stated in the following theorem that the test with critical region (2) has asymptotically level α if $B \rightarrow \infty$. Therefore, it suffices to choose B large enough so that the level of this test is close to α . But our aim is to consider quite moderate B so we also provide in Theorem 2 a control of the level in terms of α and B , under the additional assumption that the distributions of the ε_i 's are continuous. This result provides a control of what we lose in terms of first kind error probability when we consider the critical region (2) instead of (1).

Theorem 2. *Under the assumptions of Theorem 1,*

$$\lim_{B \rightarrow \infty} P_{H_0} \left(\sup_{m \in \mathcal{M}} \{ \|\Pi_m y\|_n^2 - \hat{q}_m^B(\alpha_m) \} > 0 \right) \leq \alpha.$$

If moreover the distributions of the ε_i 's are continuous, then

$$P_{H_0} \left(\sup_{m \in \mathcal{M}} \{ \|\Pi_m y\|_n^2 - \hat{q}_m^B(\alpha_m) \} > 0 \right) \leq \alpha + \sum_{m \in \mathcal{M}} 2^{D_m - n} + |\mathcal{M}| \frac{\sqrt{\pi}}{2\sqrt{2B}}.$$

By Theorem 2, it suffices to choose $B = c|\mathcal{M}|^2$ with a large enough $c > 0$ so that the level is close to α . The estimation of $q_m^y(\alpha_m)$ then requires about $n|\mathcal{M}|^2$ computations. In practical applications, one can consider, for instance, the collection of dyadic partitions \mathcal{M}_d defined in Section 5. The cardinality of \mathcal{M}_d is about $\log_2 n$ hence, using B with order of magnitude $(\log_2 n)^2$, we get a level close to the nominal level α in about $n(\log_2 n)^2$ computations.

The following theorem describes what we lose in terms of second kind error probability when we consider the critical region (2) instead of (1).

Theorem 3. *For every $m \in \mathcal{M}$, let $\delta_m < \alpha_m$. Under the assumptions of Theorem 1, we have for every $f \in \mathbb{R}^n$*

$$\lim_{B \rightarrow \infty} P_f \left(\sup_{m \in \mathcal{M}} \{ \|\Pi_m y\|_n^2 - \widehat{q}_m^B(\alpha_m) \} > 0 \right) \geq P_f \left(\sup_{m \in \mathcal{M}} \{ \|\Pi_m y\|_n^2 - q_m^y(\delta_m) \} > 0 \right).$$

4. Power

In this section, we study the power of the test with critical region (1). By Theorem 3, similar results hold for the test with critical region (2) provided B is large enough. Let \mathcal{M} be a collection of partitions of $\{1, \dots, n\}$. Let α and β be fixed numbers in $(0, 1)$. Let $(\alpha_m)_{m \in \mathcal{M}}$ be positive numbers such that $\alpha = \sum_{m \in \mathcal{M}} \alpha_m$. The aim of this section is to describe a subset $\mathcal{F}_n(\beta)$ of $\mathbb{R}^n \setminus \{0\}$ over which the power of the test is greater than $1 - \beta$, i.e., which satisfies

$$(3) \quad P_f \left(\sup_{m \in \mathcal{M}} \{ \|\Pi_m y\|_n^2 - q_m^y(\alpha_m) \} > 0 \right) \geq 1 - \beta, \quad \text{for all } f \in \mathcal{F}_n(\beta).$$

For every partition m , the subsets $e \in m$ which contain only one point do not contribute to the power of the test since their contributions to the norms of $\Pi_m y$ and $\Pi_m(w \times y)$ are identical. Hence we restrict ourselves to collections \mathcal{M} which do not contain what we call the *trivial* partition, that is the partition made up of n singletons. For every $m \in \mathcal{M}$, we set

$$J_{m,1} = \{j \in \{1, \dots, D_m\}, |e_{m,j}| = 1\}, \quad J_{m,2} = \{j \in \{1, \dots, D_m\}, |e_{m,j}| \geq 2\}$$

(hence $J_{m,2} \neq \emptyset$) and

$$I_m = \min_{j \in J_{m,2}} |e_{m,j}|.$$

We study the power of the test under two different assumptions on the ε_i 's integrability. First, we assume that the ε_i 's satisfy the following Bernstein-type condition: there exist positive real numbers γ and μ such that for all integers $p \geq 1$,

$$(4) \quad \max_{1 \leq i \leq n} E(\varepsilon_i^{2p}) \leq \gamma p! \mu^{p-2}.$$

Thus the ε_i 's possess bounded moments of any order and ε_i^2 possesses an exponential moment, hence the errors are much integrable. Note however that this assumption is less restrictive than the Gaussian assumption: if the ε_i 's are Gaussian with

bounded variance $\sigma_i^2 \leq \sigma^2$, then $E(\varepsilon_i^{2p}) \leq p!(2\sigma^2)^p$, so (4) holds with $\mu = 2\sigma^2$ and $\gamma = \mu^2$. It is also less restrictive than the boundedness assumption: if $|\varepsilon_i| \leq c$ for all i , then (4) holds with $\mu = c^2$ and $\gamma = \mu^2$. Secondly, we assume there exist $p \geq 2$ and $\mu > 0$ such that

$$(5) \quad \max_{1 \leq i \leq n} E(\varepsilon_i^{2p}) \leq \mu^p.$$

Note that under both assumptions, the ε_i 's possess bounded variances and fourth moments: $E(\varepsilon_i^2) \leq \sigma^2$ and $E(\varepsilon_i^4) \leq \mu_4$ for all i and some positive σ and μ_4 . It is also assumed in the sequel that $|f_i| \leq L$ for all $i = 1, \dots, n$ and a possibly unknown $L > 0$.

4.1. POWER UNDER THE BERNSTEIN-TYPE CONDITION.

Theorem 4. *Assume we observe $y = f + \varepsilon$, where the ε_i 's are independent variables with symmetric distribution. Assume $P(\varepsilon_i = 0) = 0$ and $|f_i| \leq L$ for a positive number L and all i . Assume moreover (4) holds for all integers $p \geq 1$ and some positive γ and μ . Let \mathcal{M} be a collection of partitions which does not contain the trivial one. Let α, β in $(0, 1)$ and $(\alpha_m)_{m \in \mathcal{M}}$ be positive numbers with $\sum_{m \in \mathcal{M}} \alpha_m = \alpha$. For every $A = (A_1, \dots, A_5) \in (0, \infty)^5$, $m \in \mathcal{M}$ and $f \in \mathbb{R}^n$ let*

$$(6) \quad \begin{aligned} \Delta_1(m, f, A) &= A_1 \|f - \Pi_m f\|_n^2 + A_2 \sqrt{\frac{\gamma D_m}{\beta}} \\ &+ A_3 (\sqrt{\gamma} + \mu + L^2) \left[1 + \frac{1}{I_m} \log \left(\frac{2D_m}{\beta} \right) \right] \log \left(\frac{2}{\beta \alpha_m} \right) \\ &+ A_4 (\sqrt{\gamma} + \mu + L^2) \sqrt{D_m \left[1 + \frac{1}{I_m} \log \left(\frac{2D_m}{\beta} \right) \right] \log \left(\frac{2}{\beta \alpha_m} \right)} + A_5 L^2 |J_{m,1}|. \end{aligned}$$

Then there exists an absolute A such that (3) holds with

$$\mathcal{F}_n(\beta) = \left\{ f \in \mathbb{R}^n, \|f\|_n^2 \geq \inf_{m \in \mathcal{M}} \Delta_1(m, f, A) \right\}.$$

Hence our test is powerful over $\mathcal{F}_n(\beta)$ provided the constants A_1, \dots, A_5 are large enough. This set is large if there exists $m \in \mathcal{M}$ such that $\Pi_m f$ is close to f , D_m and $|J_{m,1}|$ are small, while I_m is large enough.

This result applies to the i.i.d. Gaussian model. Assume that the ε_i 's are i.i.d. $\mathcal{N}(0, \sigma^2)$ and in order to make appear the signal to noise ratio, assume that there exists $L > 0$ such that $|f_i| \leq \sigma L$ for all i . Then (4) holds with $\mu = 2\sigma^2$ and $\gamma = \mu^2$. Theorem 4 consequently applies with $\sqrt{\gamma}$ and $\sqrt{\gamma} + \mu + L^2$ replaced by σ^2 and $\sigma^2(1 + L^2)$ respectively. However (see Section 7.7), one can obtain a slightly sharper result by using Cochran's theorem: the power of the test is greater than $1 - \beta$ as soon as

$$(7) \quad \|f\|_n^2 \geq \inf_{m \in \mathcal{M}} \left\{ A_1 \|f - \Pi_m f\|_n^2 \right. \\ \left. + A_3 \sigma^2 (1 + L^2) \left[1 + \frac{1}{I_m} \log \left(\frac{2D_m}{\beta} \right) \right] \log \left(\frac{2}{\beta \alpha_m} \right) \right. \\ \left. + A_4 \sigma^2 (1 + L^2) \sqrt{D_m \left[1 + \frac{1}{I_m} \log \left(\frac{2D_m}{\beta} \right) \right] \log \left(\frac{2}{\beta \alpha_m} \right)} + A_5 \sigma^2 L^2 |J_{m,1}| \right\}$$

for large enough A_k 's. This condition reduces to

$$(8) \quad \|f\|_n^2 \geq \inf_{m \in \mathcal{M}} \left\{ C_1 \|f - \Pi_m f\|_n^2 + C_2 \sigma^2 \left[\log \left(\frac{2}{\beta \alpha_m} \right) + \sqrt{D_m \log \left(\frac{2}{\alpha_m \beta} \right)} \right] \right\},$$

where C_1 is an absolute constant and C_2 only depends on L , provided this infimum is achieved at a partition $m^* \in \mathcal{M}$ with

$$(9) \quad \frac{1}{I_{m^*}} \log \left(\frac{2D_{m^*}}{\beta} \right) \leq 1 \quad \text{and, e.g.,} \quad |J_{m^*,1}| \leq \sqrt{D_{m^*}}.$$

Instead of (9), Baraud *et al.* (2003) assume that $(D_m + \log(1/\alpha))/(n - D_m)$ remains bounded to prove that the power of their test is greater than $1 - \beta$ under condition (8). Hence, both tests are powerful on similar sets. However, unlike the test of Baraud *et al.* our test requires neither Gaussian nor homoscedasticity assumptions.

4.2. POWER UNDER THE BOUNDED MOMENTS ASSUMPTION.

Theorem 5. *Assume we observe $y = f + \varepsilon$, where the ε_i 's are independent variables with symmetric distribution. Assume $P(\varepsilon_i = 0) = 0$ and $|f_i| \leq L$ for a positive number L and all i . Assume moreover (5) holds for some $p \geq 2$ and $\mu > 0$. Let \mathcal{M} be a collection of partitions which does not contain the trivial one. Let α, β in $(0, 1)$ and $(\alpha_m)_{m \in \mathcal{M}}$ be positive numbers with $\sum_{m \in \mathcal{M}} \alpha_m = \alpha$. For every $A = (A_1, \dots, A_6) \in (0, \infty)^6$, $m \in \mathcal{M}$, and $f \in \mathbb{R}^n$ let*

$$(10) \quad \begin{aligned} \Delta_2(m, f, A) &= A_1 \|f - \Pi_m f\|_n^2 + A_2 (\mu + L^2) \sqrt{\frac{D_m}{\beta}} \\ &+ A_3 p (\mu + L^2) \left[1 + \frac{1}{\sqrt{I_m}} \left(\frac{D_m}{\beta} \right)^{1/p} \right] \log \left(\frac{2}{\alpha_m} \right) \\ &+ A_4 (\mu + L^2) \sqrt{p D_m} \left[1 + \frac{1}{\sqrt{I_m}} \left(\frac{D_m}{\beta} \right)^{1/p} \right] \log \left(\frac{2}{\alpha_m} \right) + A_5 L^2 |J_{m,1}| + A_6 \frac{\mu}{\beta}. \end{aligned}$$

Then there exists an absolute A such that (3) holds with

$$\mathcal{F}_n(\beta) = \left\{ f \in \mathbb{R}^n, \|f\|_n^2 \geq \inf_{m \in \mathcal{M}} \Delta_2(m, f, A) \right\}.$$

Let us compare this with Theorem 4. There are two main differences between Δ_1 and Δ_2 . First, the absolute constants A_3 and A_4 in Δ_1 are replaced in Δ_2 by $A_3 p$ and $A_4 \sqrt{p}$ respectively. Next, the term

$$\frac{1}{I_m} \log \left(\frac{2D_m}{\beta} \right)$$

in Δ_1 is replaced in Δ_2 by the greater term

$$\frac{1}{\sqrt{I_m}} \left(\frac{D_m}{\beta} \right)^{1/p},$$

which depends on the error integrability. Hence, as expected, the set where our test is powerful is larger under the Bernstein-type condition than under the bounded moments condition.

5. Rate of Testing

In this section we assume

$$f_i = F(x_i), \quad i = 1, \dots, n,$$

for fixed numbers $x_i \in [0, 1]$ and an unknown $F: [0, 1] \rightarrow \mathbb{R}$. We aim at testing

$$H'_0: F \equiv 0 \quad \text{against} \quad H'_1: F \neq 0.$$

For this task, we consider the so-called *collection of dyadic partitions* $\mathcal{M}_d = \{m_k, k \in \mathcal{I}\}$, where

$$\mathcal{I} = \{2^l, l \in \mathbb{N}, 2^l \leq n/2\}$$

and m_k consists of the nonempty sets among

$$\left\{ i, x_i \in \left(\frac{j-1}{k}, \frac{j}{k} \right] \right\}, \quad j = 1, \dots, k.$$

Let α and β in $(0, 1)$ and $\alpha_m = \alpha/|\mathcal{M}_d|$. As in Section 4, we restrict our attention to the test with critical region (1), which still has level α in this setting. The aim of this section is to study the power of this test when the ε_i 's satisfy either (4) or (5) and F is assumed Hölderian. In particular, we aim at proving adaptive properties of the test with respect to Hölderian smoothness. We distinguish between the cases where the Hölderian smoothness of F is at most one or greater than one.

Assume the Hölderian smoothness of F to be at most one: there exist $s \in (0, 1]$, $R > 0$, and $L > 0$ such that

$$(11) \quad \forall (u, v) \in [0, 1]^2, \quad |F(u) - F(v)| \leq R|u - v|^s, \quad \text{and} \quad \sup_{u \in [0, 1]} |F(u)| \leq L.$$

Assume moreover

$$(12) \quad |J_{m_k, 1}| \leq a_0 \sqrt{D_{m_k}} \quad \text{and} \quad I_{m_k} \geq \frac{an}{k}, \quad \text{all } k \in \mathcal{I},$$

for absolute constants $a_0 \geq 0$ and $a > 0$. This means that the design points are almost equidistant. In particular, the condition is satisfied with $a_0 = 0$ and $a = 1/2$ if $x_i = i/n$ for all i . The following corollary of Theorems 4 and 5 provides conditions on f under which the test is powerful, that is

$$(13) \quad P_f \left(\sup_{m \in \mathcal{M}} \{ \|\Pi_m y\|_n^2 - q_m^y(\alpha_m) \} > 0 \right) \geq 1 - \beta.$$

For the sake of simplicity, it is assumed that $\alpha_m = \alpha/|\mathcal{M}_d|$ for every m but it is worth noticing that the results remain true if $\sum_{m \in \mathcal{M}} \alpha_m = \alpha$ and $\alpha_m \geq (\log_2 n)^{-c_\alpha}$ for all $m \in \mathcal{M}$ and a positive c_α , which only depends on α .

Corollary 1. *Assume we observe $y = f + \varepsilon$, where the ε_i 's are independent variables with symmetric distribution and $f_i = F(x_i)$ for fixed $x_i \in [0, 1]$ and an unknown F . Assume $P(\varepsilon_i = 0) = 0$ for all i and F satisfies (11) for unknown $s \in (0, 1]$, $R > 0$, and $L > 0$. Let $\mathcal{M} = \mathcal{M}_d$ be the collection of dyadic partitions. Let α and β in $(0, 1)$. For every $m \in \mathcal{M}$, let $\alpha_m = \alpha/|\mathcal{M}|$. Assume furthermore (12).*

1. *If (4) holds for all integers $p \geq 1$ and some positive numbers γ and μ , then (13) holds whenever n is large enough and one of the three following conditions is fulfilled for a large enough C (here, $\delta = \sqrt{\gamma} + \mu + L^2$ and C only depends on a, a_0, α, β):*

- (a) $s > 1/4$ and $\frac{1}{\sqrt{n}}\|f\|_n \geq CR^{1/(1+4s)} \left(\frac{\delta\sqrt{\log \log n}}{n} \right)^{2s/(1+4s)}$;
- (b) $s = 1/4$ and $\frac{1}{\sqrt{n}}\|f\|_n \geq CR^{2/3}n^{-1/4}\delta^{1/6}(\log n \times \log \log n)^{1/12}$;
- (c) $s < 1/4$ and $\frac{1}{\sqrt{n}}\|f\|_n \geq CRn^{-s}$.

2. *If (5) holds for some $p \geq 2$ and $\mu > 0$, then (13) holds whenever n is large enough and one of the three following conditions is fulfilled for a large enough C (here, $\delta_p = \sqrt{p}(\mu + L^2)$ and C only depends on a, a_0, α, β):*

- (a) $s \geq 1/4 + 1/p$ and $\frac{1}{\sqrt{n}}\|f\|_n \geq CR^{1/(1+4s)} \left(\frac{\delta_p\sqrt{\log \log n}}{n} \right)^{2s/(1+4s)}$;
- (b) $s \in [1/4 - 1/p, 1/4 + 1/p]$ and

$$\frac{1}{\sqrt{n}}\|f\|_n \geq CR^{(2+3p)/(2+3p+8ps)} \left(\frac{\delta_p\sqrt{\log \log n}}{n^{5/4}} \right)^{4ps/(2+3p+8ps)} ;$$

- (c) $s \leq 1/4 - 1/p$ and $\frac{1}{\sqrt{n}}\|f\|_n \geq CRn^{-s}$.

If $x_i = i/n$ for all i and F satisfies (11) then

$$\frac{1}{\sqrt{n}}\|f\|_n \geq \|F\|_2 - Rn^{-s}.$$

Hence in that case, Corollary 1 holds with $\|f\|_n/\sqrt{n}$ replaced by $\|F\|_2$. In particular, if $s > 1/4$ and (4) holds with $\mu = 2\sigma^2$ and $\gamma = \mu^2$ (which is indeed the case if the ε_i 's are i.i.d. $\mathcal{N}(0, \sigma^2)$), then the power of the test is greater than a prescribed $1 - \beta$ whenever n is large enough and

$$\|F\|_2 \geq CR^{1/(1+4s)} \left(\frac{\sigma^2\sqrt{\log \log n}}{n} \right)^{2s/(1+4s)}$$

for a positive C , which only depends on $L/\sigma, \alpha$, and β . This rate is precisely the minimal rate of testing obtained by Spokoiny (1996) in a white noise model. It is also the minimal rate of testing obtained by Gayraud and Pouet (2005) in the i.i.d. Gaussian regression model (but they do not describe the role of R and σ in this setting). This proves that our test achieves the optimal rate of testing under the

Bernstein-type condition (4) if $s > 1/4$, and also under the less restrictive moment condition (5) if $s \geq 1/4 + 1/p$. In the case where $s \leq 1/4$, the optimal rate of testing is not known, even in the i.i.d. Gaussian model. Note however that the rate Rn^{-s} (obtained when $s < 1/4$ and (4) holds, and also when $s \leq p - 1/4$ and (5) holds) was already obtained by Baraud *et al.* (2003) in the i.i.d. Gaussian setting for $s < 1/4$.

Now, assume the Hölderian smoothness of F to be greater than one: there exist $d \in \mathbb{N} \setminus \{0\}$, $\kappa \in (0, 1]$, R , and L such that

$$(14) \quad \forall (u, v) \in [0, 1]^2, \quad |F^{(d)}(u) - F^{(d)}(v)| \leq R|u - v|^\kappa, \quad \text{and} \quad \sup_{u \in [0, 1]} |F(u)| \leq L.$$

Here, $F^{(d)}$ denotes the d th derivative of F (which is assumed to exist) and we denote by $s = d + \kappa$ the Hölderian smoothness of F . For technical reasons, we restrict ourselves to the case where $x_i = i/n$ for all i .

Corollary 2. *Assume we observe $y = f + \varepsilon$, where the ε_i 's are independent variables with symmetric distribution and $f_i = F(i/n)$ for an unknown F . Assume $P(\varepsilon_i = 0) = 0$ for all i and F satisfies (14) for unknown $d \in \mathbb{N} \setminus \{0\}$, $\kappa \in (0, 1]$, $R > 0$, and $L > 0$. Let $s = d + \kappa > 1$ and let $\mathcal{M} = \mathcal{M}_d$ be the collection of dyadic partitions. Let α and β in $(0, 1)$. For every $m \in \mathcal{M}$, let $\alpha_m = \alpha/|\mathcal{M}|$. Assume furthermore that either (4) holds for all integers $p \geq 1$ and some positive numbers γ and μ , or (5) holds for some $p \geq 2$ and $\mu > 0$. Then there exists $C > 0$, which only depends on α , β , and s , such that (13) holds whenever n is large enough and*

$$(15) \quad \|F\|_2 \geq CR^{1/(1+4s)} \left(\frac{\delta \sqrt{\log \log n}}{n} \right)^{2s/(1+4s)}.$$

Here, $\delta = \sqrt{\gamma} + \mu + L$ under assumption (4) and $\delta = \sqrt{p}(\mu + L^2)$ under assumption (5).

Thus our test achieves the optimal rate of testing when $s > 1$ under mild assumptions on the ε_i 's.

6. Simulations

We carried out a simulation study to demonstrate the behavior of our test. We also simulated the test $T_{P_2, \mathcal{M}_{\text{dyad}}}$ proposed by Baraud *et al.* (2003) in order to compare the performance of the two tests.

6.1. THE SIMULATION EXPERIMENT. We generated a random vector $\varepsilon \in \mathbb{R}^n$ from a given distribution \mathcal{G} and computed $y = f + \varepsilon$, where $n \in \{64, 128, 256\}$ and $f_i = F(x_i)$ for a given function F and fixed x_i 's. We considered the collection of dyadic partitions $\mathcal{M} = \mathcal{M}_d$ as defined in Section 5, we set $\alpha = 5\%$ and $\alpha_m = \alpha/|\mathcal{M}|$ for every $m \in \mathcal{M}$. Then, we generated $B = 2500$ independent copies of w and computed $\hat{q}_m^B(\alpha_m)$ from these copies. Finally, we rejected the null hypothesis that $F \equiv 0$ if (2) holds. For given \mathcal{G} , F and n , we repeated this step 1500 times in order to get 1500 independent tests, and we computed the percentage of rejections among these tests. We thus obtained an estimate for the level (if $F \equiv 0$) or for the power of our test. In order to assess the accuracy of the estimate, we repeated the above

computations independently 100 times and computed the mean and the standard deviation of the percentages of rejection obtained over these 100 simulations. For every choice of \mathcal{G} , F , and n , we obtained a small standard deviation (about 2.10^{-3} to 5.10^{-3}), which means that the mean of the percentages of rejection is a good estimate for the actual level of the test. For each simulation, we treated likewise the $T_{P_2, \mathcal{M}_{dya}}$ test of Baraud *et al.* (2003).

We considered five different distributions \mathcal{G} . In all cases, the variables ε_i are mutually independent, with mean zero. The first distribution (called *Gaussian*) is the one under which our test and the test $T_{P_2, \mathcal{M}_{dya}}$ both have prescribed level: the ε_i 's are identically distributed and standard Gaussian. Then we considered two distributions, called *Mixture* and *Heteroscedastic*, under which our test has prescribed level, while $T_{P_2, \mathcal{M}_{dya}}$ does not: under *Mixture* the ε_i 's are i.i.d. mixture of Gaussian distributions as defined in Baraud *et al.* (2003) ((b), p. 236); under *Heteroscedastic*, ε_i is a centered Gaussian variable with variance

$$v(i) = \sin^4 \left(4\pi \left(\frac{i-1}{n-1} \right) \left(\frac{n-i}{n-1} \right) \right).$$

Finally, we considered two distributions, called *Type I* and *Asymmetric* under which none of the two tests have prescribed level: in both cases, the ε_i 's are i.i.d., *Type I* is defined as in Baraud *et al.* (2003) ((c), p. 236) and under *Asymmetric*, the ε_i 's are distributed as

$$\frac{3}{2\sqrt{5}} \left(U^{-2/5} - \frac{5}{3} \right),$$

where U is uniformly distributed on $[0, 1]$. The distribution of the ε_i 's is weakly asymmetric under *Type I* and strongly asymmetric under *Asymmetric*.

6.2. THE LEVEL. We first considered the case where $F \equiv 0$ in order to estimate the level of the tests. For simplicity, the mean of 100 percentages of rejection (as described in Section 6.1) is called here the *estimated level*. The estimated levels obtained for different \mathcal{G} and n are given in Table 1. Under *Gaussian*, *Mixture* and *Heteroscedastic* the estimated level of our test is, as expected, no greater than the nominal level. Under *Type I*, the estimated level remains less than the nominal level which means that the method is robust against a slight departure from symmetry. Under *Asymmetric*, the estimated level is greater than α when n is small and, despite asymmetry, it is smaller than α for $n = 256$. Even if we cannot explain rigorously this phenomenon, it seems interesting to us to explain it heuristically. Recall that the test consists in selecting a good partition m^* and then comparing the distributions of $\|\Pi_{m^*} y\|_n^2$ and $\|\Pi_{m^*}(w \times y)\|_n^2$. If n is large, then either each subset in m^* contains a large number of points or the number of subsets in m^* is large (or even, the two properties hold simultaneously). If each subset in m^* contains a large number of points, then by the central limit theorem the projections of y and $w \times y$ on these subsets have distributions, which are close to Gaussian; if the number of sets in m^* is large, then the squared norm of the projections of y and $w \times y$ on m^* (which are sums of projections on the subsets in m^*) have distributions close to Gaussian. It thus seems that the combination of two central limit theorems forces $\|\Pi_{m^*} y\|_n$ and $\|\Pi_{m^*}(w \times y)\|_n$ to have distributions, which are close to each other when $f = 0$ and n is large, even if the distributions

of the ε_i 's are not symmetric. This explains why the level of our test is less than α under *Asymmetric* for large n .

We now compare the performance of our test with that of $T_{P_2, \mathcal{M}_{dya}}$. Under *Gaussian*, *Mixture*, and *Type I*, the two tests have similar estimated levels and are conservative. Under *Heteroscedastic*, our test is still conservative while the estimated level of $T_{P_2, \mathcal{M}_{dya}}$ is much larger than α , even for large n . Finally, under *Asymmetric*, both tests have estimated level larger than α when n is small. But our test performs better than $T_{P_2, \mathcal{M}_{dya}}$, all the more so that n increases. In particular, our test has the prescribed level when $n = 256$, whereas $T_{P_2, \mathcal{M}_{dya}}$ has not.

TABLE 1. Estimated level of our test (Roman) and $T_{P_1, \mathcal{M}_{dya}}$ (italic) for the five different distributions \mathcal{G} when $\alpha = 5\%$ and $n \in \{64, 128, 256\}$

Distribution \mathcal{G}	$n = 64$	$n = 128$	$n = 256$
<i>Gaussian</i>	3.71	3.80	3.72
	<i>3.53</i>	<i>3.54</i>	<i>3.46</i>
<i>Mixture</i>	3.70	3.79	3.76
	<i>2.96</i>	<i>3.20</i>	<i>3.31</i>
<i>Heteroscedastic</i>	3.15	3.18	3.19
	<i>16.4</i>	<i>19.3</i>	<i>21.9</i>
<i>Type I</i>	3.67	3.46	3.33
	<i>3.94</i>	<i>3.89</i>	<i>3.70</i>
<i>Asymmetric</i>	9.10	6.52	4.75
	<i>10.1</i>	<i>8.62</i>	<i>7.60</i>

6.3. THE POWER. In this section, we consider cases where $F \neq 0$, so the mean of 100 percentages of rejection (as described in Section 6.1) is called here *estimated power*. To choose regression functions F , we were inspired by Baraud *et al.* (2003), where simulations of a test for linearity were performed. To adapt their alternatives to the case of testing for zero regression, we removed the linear part of their functions, so we considered the following functions:

$$\begin{aligned} F_{1k}(x) &= c_{1k} \cos(10\pi x), & k &= 1, 2, 3, 4, \\ F_{2k}(x) &= 5 \phi(x/c_{2k})/c_{2k}, & k &= 1, 2, \\ F_{3k}(x) &= -c_{3k} (x - 0.1) \mathbf{1}_{x \leq 0.1}, & k &= 1, 2, 3, \end{aligned}$$

where ϕ is the standard Gaussian density and $c_{11}, c_{12}, \dots, c_{33}$ are equal to 0.25, 0.5, 0.75, 1, 0.25, 1, 20, 30, 40 respectively (see Figure 1). Thus for $j \in \{1, 3\}$, the larger k the farther F_{jk} is from the null function. In the case of alternatives F_{1k} and F_{3k} we set $x_i = (i - 0.5)/n$, while for F_{2k} the x_i 's are simulated once for all as i.i.d. centered Gaussian variables with variance 25 in the range $[\Phi^{-1}(0.05), \Phi^{-1}(0.95)]$, where Φ is the standard Gaussian distribution function.

Under *Gaussian* (see Table 2), the two tests have similar power against alternatives F_{1k} and F_{2k} : the estimated powers are good against F_{2k} and quite small

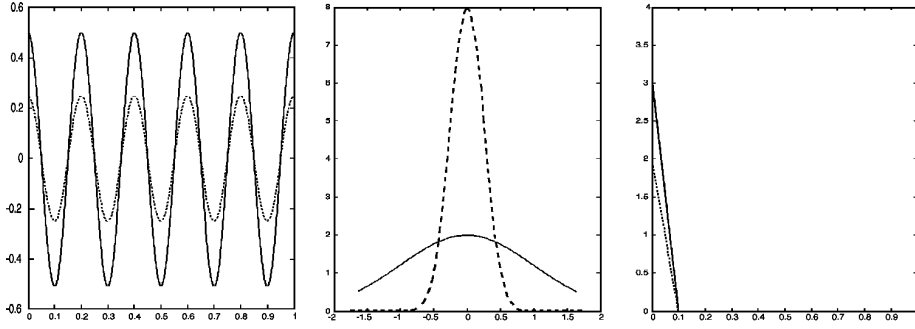


FIGURE 1. Alternatives F_{11} , F_{12} (left), F_{21} , F_{22} (center) and F_{31} , F_{32} (right). Alternatives F_{j1} and F_{j2} are drawn in dotted and plain lines respectively

against F_{1k} , especially for small values of k and n . The estimated power of $T_{P_2, \mathcal{M}_{dya}}$ is greater than that of our test against F_{3k} , all the more so n is small and k is large (that is, the signal/noise ratio is large). This is certainly due to the small number of indices i such that $F_{3k}(x_i) \neq 0$. Indeed, our test can detect an alternative if $\|\Pi_{m^*} y\|_n^2$ is significantly larger than $\|\Pi_{m^*}(w \times y)\|_n^2$ for a well chosen partition m^* , but $\|\Pi_{m^*} y\|_n^2$ remains close to $\|\Pi_{m^*}(w \times y)\|_n^2$ if there are only a few indices i such that y_i is significantly different from zero.

TABLE 2. Estimated power of our test (Roman) and $T_{P_2, \mathcal{M}_{dya}}$ (italic) under Gaussian distribution. The nominal level is $\alpha = 5\%$

Distribution \mathcal{G}	Gaussian								
	F_{11}	F_{12}	F_{13}	F_{14}	F_{21}	F_{22}	F_{31}	F_{32}	F_{33}
$n = 64$	4.41	8.00	19.4	41.7	99.9	100	12.2	20.2	27.5
	<i>4.29</i>	<i>7.78</i>	<i>18.5</i>	<i>39.5</i>	<i>99.8</i>	<i>100</i>	<i>16.2</i>	<i>40.4</i>	<i>72.4</i>
$n = 128$	6.46	25.0	69.4	96.7	100	100	34.4	70.3	91.2
	<i>6.12</i>	<i>24.2</i>	<i>68.2</i>	<i>96.3</i>	<i>100</i>	<i>100</i>	<i>43.3</i>	<i>89.4</i>	<i>99.8</i>
$n = 256$	12.3	70.1	99.5	100	100	100	84.8	99.9	100
	<i>11.7</i>	<i>68.9</i>	<i>99.5</i>	<i>100</i>	<i>100</i>	<i>100</i>	<i>87.8</i>	<i>100</i>	<i>100</i>

Under *Mixture* or *Type I* (see Table 3), the estimated powers of the two tests are smaller than under *Gaussian*, which is due to a greater variance of the errors: under *Mixture* (resp. *Type I*), the common variance of the ε_i 's equals 4.678 (resp. 4). One can notice however that the signal/noise ratio is the same under *Gaussian* with F_{11} (resp. F_{12} , resp. F_{31}) as under *Type I* with F_{12} (resp. F_{14} , resp. F_{33}), and is close to that under *Mixture* with F_{12} (resp. F_{14} , resp. F_{33}). Comparing the estimated powers in these cases shows that, the signal/noise ratio being fixed, both tests have similar powers under these three distributions.

Finally, we studied the power of the two tests under *Heteroscedastic* and *Asymmetric*. We give in Table 4 the estimated powers only in the cases where the

TABLE 3. Estimated power of our test (Roman) and $T_{P_2, \mathcal{M}_{dy\alpha}}$ (italic) under *Mixture* and *Type I* distributions. The nominal level is $\alpha = 5\%$

Distribution \mathcal{G}	<i>Mixture</i>								
F	F_{11}	F_{12}	F_{13}	F_{14}	F_{21}	F_{22}	F_{31}	F_{32}	F_{33}
$n = 64$	3.93	4.61	6.24	9.32	88.0	98.0	6.15	9.11	12.9
	<i>3.12</i>	<i>3.63</i>	<i>4.63</i>	<i>6.80</i>	<i>87.1</i>	<i>97.5</i>	<i>5.06</i>	<i>8.43</i>	<i>14.4</i>
$n = 128$	4.31	6.63	13.0	25.6	99.8	100	9.74	19.5	34.2
	<i>3.60</i>	<i>5.32</i>	<i>10.1</i>	<i>20.5</i>	<i>99.8</i>	<i>100</i>	<i>8.46</i>	<i>18.7</i>	<i>37.3</i>
$n = 256$	5.16	11.9	31.7	63.4	100	100	18.9	47.1	78.6
	<i>4.41</i>	<i>10.0</i>	<i>27.6</i>	<i>59.0</i>	<i>100</i>	<i>100</i>	<i>17.0</i>	<i>45.7</i>	<i>79.5</i>

Distribution \mathcal{G}	<i>Type I</i>								
F	F_{11}	F_{12}	F_{13}	F_{14}	F_{21}	F_{22}	F_{31}	F_{32}	F_{33}
$n = 64$	3.88	4.50	5.86	8.68	92.8	99.9	4.99	7.27	10.6
	<i>4.15</i>	<i>4.69</i>	<i>5.93</i>	<i>8.38</i>	<i>92.4</i>	<i>99.9</i>	<i>5.67</i>	<i>9.08</i>	<i>15.6</i>
$n = 128$	3.99	6.21	12.4	26.4	100	100	8.04	17.5	33.8
	<i>4.41</i>	<i>6.52</i>	<i>12.3</i>	<i>25.2</i>	<i>100</i>	<i>100</i>	<i>9.25</i>	<i>21.0</i>	<i>43.0</i>
$n = 256$	4.58	11.8	35.1	70.8	100	100	18.3	51.4	87.3
	<i>4.98</i>	<i>11.9</i>	<i>33.9</i>	<i>69.1</i>	<i>100</i>	<i>100</i>	<i>19.3</i>	<i>53.3</i>	<i>88.7</i>

TABLE 4. Estimated power of our test (Roman) and $T_{P_2, \mathcal{M}_{dy\alpha}}$ (italic) under *Heteroscedastic* and *Asymmetric* distributions. The nominal level is $\alpha = 5\%$

Distribution \mathcal{G}	<i>Heteroscedastic</i>								
F	F_{11}	F_{12}	F_{13}	F_{14}	F_{21}	F_{22}	F_{31}	F_{32}	F_{33}
$n = 64$	3.94	7.30	15.8	30.8	99.6	100	11.4	17.8	23.1
$n = 128$	4.32	10.2	30.0	64.4	100	100	19.3	45.5	73.4
$n = 256$	5.33	20.5	70.0	98.4	100	100	41.6	95.1	100

Distribution \mathcal{G}	<i>Asymmetric</i>								
F	F_{11}	F_{12}	F_{13}	F_{14}	F_{21}	F_{22}	F_{31}	F_{32}	F_{33}
$n = 128$	20.9	75.0	93.6	97.8	100	100	74.5	96.8	99.6
	<i>17.0</i>	<i>59.9</i>	<i>85.9</i>	<i>94.0</i>	<i>100</i>	<i>100</i>	<i>76.4</i>	<i>93.8</i>	<i>97.4</i>
$n = 256$	39.2	93.6	98.9	99.6	100	100	99.2	100	100
	<i>30.9</i>	<i>85.2</i>	<i>96.1</i>	<i>98.3</i>	<i>100</i>	<i>100</i>	<i>93.3</i>	<i>98.2</i>	<i>99.2</i>

estimated level of the test is not much larger than the nominal level. In particular, under *Heteroscedastic* we only give the estimated power of our test. We can see that the estimated power of our test is smaller under *Heteroscedastic* than under *Gaussian*, but remains reasonable. Under *Asymmetric*, the estimated power of our test is in most cases greater than that of $T_{P_2, \mathcal{M}_{dy\alpha}}$, although the estimated level of $T_{P_2, \mathcal{M}_{dy\alpha}}$ is greater than α .

In conclusion, our test is powerful against various alternatives (provided the number of indices i such that y_i is significantly different from zero is large enough) and robust against departures from Gaussian distribution and homoscedasticity. It is also robust against slight departures from symmetry.

7. Proofs

Some of the proofs require lemmas. Lemmas are stated when needed and proved in Section 7.8.

7.1. PROOF OF THEOREM 1. Under H_0 , we have $y = \varepsilon$, so our aim here is to prove that

$$(16) \quad P\left(\sup_{m \in \mathcal{M}} \{\|\Pi_m \varepsilon\|_n^2 - q_m^\varepsilon(\alpha_m)\} > 0\right) \leq \alpha.$$

For every $u \in \mathbb{R}^n$, let $|u|$ and $\text{sign}(u)$ be the vectors of \mathbb{R}^n with i th component $|u_i|$ and

$$(17) \quad (\text{sign}(u))_i = \mathbf{1}_{\{u_i \geq 0\}} - \mathbf{1}_{\{u_i < 0\}},$$

respectively. The distribution of ε_i is symmetric about zero and $P(\varepsilon_i = 0) = 0$, hence $\text{sign}(\varepsilon)$ has the same distribution as w and is independent of $|\varepsilon|$. Moreover, $w \times \text{sign}(\varepsilon)$ has the same distribution as w . Since

$$w \times \varepsilon = w \times \text{sign}(\varepsilon) \times |\varepsilon|,$$

it follows that the distribution of $w \times \varepsilon$ conditionally on ε is identical to its distribution conditionally on $|\varepsilon|$. In particular, $q_m^\varepsilon(\alpha_m) = q_m^{|\varepsilon|}(\alpha_m)$, where

$$q_m^{|\varepsilon|}(\alpha_m) = \inf \{x \in \mathbb{R}, P[\|\Pi_m(w \times \varepsilon)\|_n^2 > x \mid |\varepsilon|] \leq \alpha_m\}.$$

But conditionally on $|\varepsilon|$, $w \times \varepsilon$ has the same distribution as $\text{sign}(\varepsilon) \times |\varepsilon| = \varepsilon$. Therefore for every $m \in \mathcal{M}$,

$$P(\|\Pi_m \varepsilon\|_n^2 > q_m^\varepsilon(\alpha_m) \mid |\varepsilon|) = P(\|\Pi_m(w \times \varepsilon)\|_n^2 > q_m^{|\varepsilon|}(\alpha_m) \mid |\varepsilon|) \leq \alpha_m.$$

Integrating the latter inequality yields

$$P(\|\Pi_m \varepsilon\|_n^2 > q_m^\varepsilon(\alpha_m)) \leq \alpha_m.$$

By assumption $\sum_{m \in \mathcal{M}} \alpha_m = \alpha$, hence we get (16). \square

7.2. PROOF OF THEOREM 2. The first kind error probability of the test with critical region (2) satisfies

$$P\left(\sup_{m \in \mathcal{M}} \{\|\Pi_m \varepsilon\|_n^2 - \widehat{q}_m^B(\alpha_m)\} > 0\right) \leq P[\exists m, q_m^\varepsilon(\alpha_m) < \|\Pi_m \varepsilon\|_n^2] + \mathbb{P}_n^B,$$

where

$$\mathbb{P}_n^B = P[\exists m, \|\Pi_m \varepsilon\|_n^2 > \widehat{q}_m^B(\alpha_m) \text{ and } q_m^\varepsilon(\alpha_m) \geq \|\Pi_m \varepsilon\|_n^2].$$

It thus follows from Theorem 1 that

$$(18) \quad P\left(\sup_{m \in \mathcal{M}} \{\|\Pi_m \varepsilon\|_n^2 - \widehat{q}_m^B(\alpha_m)\} > 0\right) \leq \alpha + \mathbb{P}_n^B.$$

We can have $q_m^\varepsilon(\alpha_m) \geq \|\Pi_m \varepsilon\|_n^2$ if and only if $p_m^\varepsilon > \alpha_m$, where

$$p_m^\varepsilon = P(\|\Pi_m(\omega \times \varepsilon)\|_n^2 \geq \|\Pi_m \varepsilon\|_n^2 \mid \varepsilon).$$

Likewise, we can have $\widehat{q}_m^B(\alpha_m) < \|\Pi_m \varepsilon\|_n^2$ if and only if

$$\frac{1}{B} \sum_{b=1}^B \mathbf{1}_{\|\Pi_m(w^b \times \varepsilon)\|_n^2 \geq \|\Pi_m \varepsilon\|_n^2} \leq \alpha_m.$$

Conditioning with respect to ε thus yields

$$\mathbb{P}_n^B \leq \sum_{m \in \mathcal{M}} E \left[P \left(\frac{1}{B} \sum_{b=1}^B \mathbf{1}_{\|\Pi_m(w^b \times \varepsilon)\|_n^2 \geq \|\Pi_m \varepsilon\|_n^2} \leq \alpha_m \mid \varepsilon \right) \mathbf{1}_{p_m^\varepsilon > \alpha_m} \right].$$

For every $m \in \mathcal{M}$, let S_m^ε be a random variable, which is distributed conditionally on ε as a binomial variable with parameter B and probability of success p_m^ε . The Hoeffding inequality yields

$$(19) \quad \begin{aligned} \mathbb{P}_n^B &\leq \sum_{m \in \mathcal{M}} E [P(S_m^\varepsilon \leq B\alpha_m \mid \varepsilon) \mathbf{1}_{p_m^\varepsilon > \alpha_m}] \\ &\leq \sum_{m \in \mathcal{M}} E [\exp(-2B(p_m^\varepsilon - \alpha_m)^2) \mathbf{1}_{p_m^\varepsilon > \alpha_m}]. \end{aligned}$$

By dominated convergence,

$$\lim_{B \rightarrow \infty} E [\exp(-2B(p_m^\varepsilon - \alpha_m)^2) \mathbf{1}_{p_m^\varepsilon > \alpha_m}] = 0,$$

so the first part of the theorem follows from (18). In order to prove the second part, we compute the distribution of p_m^ε in the case where the distributions of the ε_i 's are continuous (see Section 7.8 for a proof).

Lemma 1. *Under the assumptions of Theorem 1 with the additional assumption that the distributions of the ε_i 's are continuous, p_m^ε has a discrete uniform distribution on the set*

$$\mathcal{E} = \{k2^{D_m-n}, k = 1, \dots, 2^{n-D_m}\}.$$

Combining (19) and Lemma 1 we get

$$\begin{aligned} \mathbb{P}_n^B &\leq \sum_{m \in \mathcal{M}} \left(2^{D_m-n} \sum_{x \in \mathcal{E}, x > \alpha_m} \exp(-2B(x - \alpha_m)^2) \right) \\ &\leq \sum_{m \in \mathcal{M}} \left(2^{D_m-n} + \int_{\alpha_m}^1 \exp(-2B(x - \alpha_m)^2) dx \right). \end{aligned}$$

The result now follows from (18) and straightforward computations. \square

7.3. PROOF OF THEOREM 3. The second kind error probability of the test with critical region (2) satisfies

$$P_f\left(\sup_{m \in \mathcal{M}} \{\|\Pi_m y\|_n^2 - \widehat{q}_m^B(\alpha_m)\} \leq 0\right) \leq P_f[\forall m, q_m^y(\delta_m) \geq \|\Pi_m y\|_n^2] + \mathbb{Q}_n^B,$$

where

$$\mathbb{Q}_n^B = P_f[\exists m, \|\Pi_m y\|_n^2 \leq \widehat{q}_m^B(\alpha_m) \text{ and } q_m^y(\delta_m) < \|\Pi_m y\|_n^2].$$

Using the same arguments as in the proof of Theorem 2 (and also the same notation) we get

$$\mathbb{Q}_n^B \leq \sum_{m \in \mathcal{M}} E_f[P_f(S_m^y > B\alpha_m \mid y)\mathbf{1}_{p_m^y \leq \delta_m}],$$

where S_m^y is a random variable distributed conditionally on y as a binomial variable with parameter B and probability of success p_m^y . By assumption, $\delta_m < \alpha_m$ for all m , so by dominated convergence, \mathbb{Q}_n^B tends to zero as $B \rightarrow \infty$, which proves the result. \square

7.4. PROOF OF THEOREMS 4 AND 5. In this section, we first describe the common line of the proof for the two theorems and then describe the specific arguments for each of them. The lemmas stated in this section are proved in Section 7.8. The following three inequalities are repeatedly used throughout the proof: for all positive real numbers a and b

$$(20) \quad \sqrt{a+b} \leq \sqrt{a} + \sqrt{b};$$

for all positive real numbers a , b and θ ,

$$(21) \quad 2\sqrt{ab} \leq \theta a + \theta^{-1}b;$$

for all positive numbers a , b and k we have

$$(22) \quad (a+b)^k \leq 2^k(a^k \vee b^k).$$

Line of proof. Fix $\beta \in (0, 1)$ and $f \in \mathcal{F}_n(\beta)$. The second kind error probability of the test at f satisfies

$$P_f\left(\sup_{m \in \mathcal{M}} \{\|\Pi_m y\|_n^2 - q_m^y(\alpha_m)\} \leq 0\right) \leq \inf_{m \in \mathcal{M}} P_f(\|\Pi_m y\|_n^2 \leq q_m^y(\alpha_m)).$$

Therefore, the power of the test is at least $1 - \beta$ whenever there exists $m \in \mathcal{M}$ such that

$$(23) \quad P_f(\|\Pi_m y\|_n^2 \leq q_m^y(\alpha_m)) \leq \beta.$$

Since $f \in \mathcal{F}_n(\beta)$, there exists a partition $m \in \mathcal{M}$ such that

$$(24) \quad \|f\|_n^2 \geq \Delta(m, f, A),$$

where Δ denotes either Δ_1 or Δ_2 . In the sequel, m denotes such a partition. We aim to prove that (23) holds for this partition, provided the A_k 's are large enough. Now we show that the subsets $e_{m,j}$ which contain only one point cannot contribute to the power of the test. Let $\tilde{y} \in \mathbb{R}^n$ with i th component $\tilde{y}_i = 0$ if $i \in e_{m,j}$ for an index $j \in J_{m,1}$ and $\tilde{y}_i = y_i$ otherwise. For every $j \in J_{m,1}$, let $i(j)$ be the unique element of $e_{m,j}$. By definition of Π_m ,

$$\|\Pi_m y\|_n^2 = \sum_{j \in J_{m,2}} \frac{1}{|e_{m,j}|} \left(\sum_{i \in e_{m,j}} y_i \right)^2 + \sum_{j \in J_{m,1}} y_{i(j)}^2 = \|\Pi_m \tilde{y}\|_n^2 + \sum_{j \in J_{m,1}} y_{i(j)}^2.$$

Let

$$\tilde{q}_m^y(\alpha_m) = \inf \{x \in \mathbb{R}, P_f(\|\Pi_m(w \times \tilde{y})\|_n^2 > x \mid y) \leq \alpha_m\}.$$

We have $w_i = \pm 1$, so

$$\|\Pi_m(w \times y)\|_n^2 = \|\Pi_m(w \times \tilde{y})\|_n^2 + \sum_{j \in J_{m,1}} y_{i(j)}^2,$$

and we get

$$q_m^y(\alpha_m) = \tilde{q}_m^y(\alpha_m) + \sum_{j \in J_{m,1}} y_{i(j)}^2.$$

Hence (23) amounts to

$$(25) \quad P_f(\|\Pi_m \tilde{y}\|_n^2 \leq \tilde{q}_m^y(\alpha_m)) \leq \beta,$$

which means that the sets with only one point can be removed from the condition. Thus we aim to prove that (25) holds provided the A_k 's are large enough. We set

$$Z_m = \|\Pi_m(w \times \tilde{y})\|_n^2 = \sum_{j \in J_{m,2}} \frac{1}{|e_{m,j}|} \left(\sum_{i \in e_{m,j}} w_i y_i \right)^2$$

and

$$M_m = \max_{j \in J_{m,2}} \left\{ \frac{1}{|e_{m,j}|} \sum_{i \in e_{m,j}} y_i^2 \right\}.$$

We first give an upper bound for $\tilde{q}_m^y(\alpha_m)$ in order to control the probability in (25).

Lemma 2.

$$\tilde{q}_m^y(\alpha_m) \leq E_f(Z_m \mid y) + 8M_m \log(1/\alpha_m) + 4\sqrt{2E_f(Z_m \mid y)M_m \log(1/\alpha_m)}.$$

The obtained upper bound depends on $E_f(Z_m \mid y)$ and M_m , which we have to control. It is proved in the following two lemmas that with high probability, these variables are not much greater than their expectation. The control we obtain depends on the assumptions on ε .

Lemma 3. *Under the assumptions of Theorem 1,*

$$P_f\left(M_m \geq 18(\sqrt{\gamma} \vee \mu \vee L^2) \left[1 + \frac{1}{I_m} \log\left(\frac{3D_m}{\beta}\right)\right]\right) \leq \frac{\beta}{3}$$

and

$$P_f\left(E_f(Z_m | y) \geq E_f(Z_m) + 8(\sqrt{\gamma} \vee L^2) \sqrt{D_m \log(3/\beta)} + 2(\mu \vee L^2) \log(3/\beta)\right) \leq \frac{\beta}{3}.$$

Lemma 4. *Under the assumptions of Theorem 5, there exists an absolute constant $C \geq 1$ such that*

$$P_f\left(M_m \geq Cp(\mu \vee L^2) \left[1 + \frac{1}{\sqrt{I_m}} \left(\frac{3D_m}{\beta}\right)^{1/p}\right]\right) \leq \frac{\beta}{3}.$$

Moreover,

$$P_f\left(E_f(Z_m | y) \geq E_f(Z_m) + 4(\mu \vee L^2) \sqrt{6D_m/\beta}\right) \leq \frac{\beta}{3}.$$

The main issue to prove Theorems 4 and 5 is to derive from the three lemmas above that

$$(26) \quad P_f(\tilde{q}_m^y(\alpha_m) \leq E_f(Z_m) + R_m) \geq 1 - \frac{2}{3}\beta,$$

where R_m is a positive real number to be chosen later. Then,

$$(27) \quad P_f(\|\Pi_m \tilde{y}\|_n^2 \leq \tilde{q}_m^y(\alpha_m)) \leq \frac{2}{3}\beta + P_f(\|\Pi_m \tilde{y}\|_n^2 \leq E_f(Z_m) + R_m),$$

and it remains to prove that the right-hand side probability is less than or equal to $\beta/3$. In order to do that, we state a concentration inequality which proves that with high probability, $\|\Pi_m \tilde{y}\|_n^2$ is not much smaller than its expectation. Here again, the obtained control depends on the assumptions on ε . In the sequel, \tilde{f} denotes the expectation of \tilde{y} .

Lemma 5. *Under the assumptions of Theorem 5,*

$$P_f\left(\|\Pi_m \tilde{y}\|_n^2 \leq E_f \|\Pi_m \tilde{y}\|_n^2 - 6\sqrt{\frac{\gamma D_m}{\beta}} - \frac{1}{3} \|\Pi_m \tilde{f}\|_n^2 - 14(\sqrt{\gamma} \vee \mu \vee L^2) \log\left(\frac{6}{\beta}\right)\right) \leq \frac{\beta}{3}.$$

Lemma 6. *Under the assumptions of Theorem 5,*

$$P_f\left(\|\Pi_m \tilde{y}\|_n^2 \leq E_f \|\Pi_m \tilde{y}\|_n^2 - 3\mu \sqrt{\frac{D_m}{\beta}} - \frac{1}{3} \|\Pi_m \tilde{f}\|_n^2 - \frac{9\mu}{\beta}\right) \leq \frac{\beta}{3}.$$

In both cases, there is $R'_m \geq 0$ such that

$$(28) \quad P_f\left(\|\Pi_m \tilde{y}\|_n^2 \leq E_f \|\Pi_m \tilde{y}\|_n^2 - R'_m - \frac{1}{3} \|\Pi_m \tilde{f}\|_n^2\right) \leq \frac{\beta}{3}.$$

Thus the right-hand side probability in (27) is less than or equal to $\beta/3$ as soon as

$$(29) \quad E_f(Z_m) + R_m \leq E_f \|\Pi_m \tilde{y}\|_n^2 - R'_m - \frac{1}{3} \|\Pi_m \tilde{f}\|_n^2,$$

hence it suffices to prove (29). By assumption, the w_i 's and the y_i 's are mutually independent, and the distribution of w_i is symmetric about zero. Therefore, $E(w_i) = 0$ and the random variables $w_i y_i$ are mutually independent with zero mean and variance $E_f(y_i^2) = E(\varepsilon_i^2) + f_i^2$. Hence,

$$(30) \quad E_f(Z_m) = \sum_{j \in J_{m,2}} \frac{1}{|e_{m,j}|} \sum_{i \in e_{m,j}} (E(\varepsilon_i^2) + f_i^2).$$

We have $|e_{m,j}| \geq I_m$ for all $j \in J_{m,2}$, so

$$\begin{aligned} & E_f \|\Pi_m \tilde{y}\|_n^2 - E_f(Z_m) \\ &= \sum_{j \in J_{m,2}} \frac{1}{|e_{m,j}|} \left(\sum_{i \in e_{m,j}} E(\varepsilon_i^2) + \left(\sum_{i \in e_{m,j}} f_i \right)^2 - \sum_{i \in e_{m,j}} (E(\varepsilon_i^2) + f_i^2) \right) \\ &\geq \sum_{j \in J_{m,2}} \frac{1}{|e_{m,j}|} \left(\sum_{i \in e_{m,j}} f_i \right)^2 - \frac{1}{I_m} \sum_{j \in J_{m,2}} \sum_{i \in e_{m,j}} f_i^2. \end{aligned}$$

Hence

$$E_f \|\Pi_m \tilde{y}\|_n^2 - E_f(Z_m) \geq \|\Pi_m f\|_n^2 - |J_{m,1}| \max_{1 \leq i \leq n} f_i^2 - \frac{1}{I_m} \|f\|_n^2.$$

In order to prove (29), it suffices to prove

$$(31) \quad \frac{2}{3} \|\Pi_m f\|_n^2 \geq \frac{1}{I_m} \|f\|_n^2 + |J_{m,1}| \max_{1 \leq i \leq n} f_i^2 + R_m + R'_m.$$

By definition, $I_m \geq 2$ and by the Pythagoras equality,

$$(32) \quad \|\Pi_m f\|_n^2 = \|f\|_n^2 - \|f - \Pi_m f\|_n^2.$$

Thus it suffices to prove

$$\frac{1}{6} \|f\|_n^2 \geq \frac{2}{3} \|f - \Pi_m f\|_n^2 + |J_{m,1}| \max_{1 \leq i \leq n} f_i^2 + R_m + R'_m.$$

But f satisfies (24), so it suffices to check that for large enough A_k ,

$$(33) \quad \Delta(m, f, A) \geq 4 \|f - \Pi_m f\|_n^2 + 6 \{ |J_{m,1}| \max_{1 \leq i \leq n} f_i^2 + R_m + R'_m \}.$$

Now to prove Theorems 4 and 5, it remains to compute R_m and R'_m and to check (33) with Δ replaced by Δ_1 and Δ_2 respectively.

Proof of Theorem 4. By (30),

$$E_f(Z_m) \leq D_m(\sqrt{2\gamma} + L^2).$$

Using (21) we thus get

$$\begin{aligned} E_f(Z_m) + 8(\sqrt{\gamma} \vee L^2)\sqrt{D_m \log(3/\beta)} + 2(\mu \vee L^2) \log(3/\beta) \\ \leq 7(\sqrt{\gamma} \vee \mu \vee L^2)(D_m + \log(3/\beta)). \end{aligned}$$

Combining Lemmas 2 and 3 proves that with probability greater than $1 - 2\beta/3$,

$$\begin{aligned} \tilde{q}_m^y(\alpha_m) &\leq E_f(Z_m) + 8(\sqrt{\gamma} \vee L^2)\sqrt{D_m \log(3/\beta)} + 2(\mu \vee L^2) \log(3/\beta) \\ &\quad + 144(\sqrt{\gamma} \vee \mu \vee L^2) \left[1 + \frac{1}{I_m} \log\left(\frac{3D_m}{\beta}\right) \right] \log\left(\frac{1}{\alpha_m}\right) \\ &\quad + 24(\sqrt{\gamma} \vee \mu \vee L^2) \sqrt{7\left(D_m + \log\left(\frac{3}{\beta}\right)\right) \left(1 + \frac{1}{I_m} \log\left(\frac{3D_m}{\beta}\right)\right) \log\left(\frac{1}{\alpha_m}\right)}. \end{aligned}$$

Using (20) and (21), we get that with probability greater than $1 - 2\beta/3$,

$$\begin{aligned} \tilde{q}_m^y(\alpha_m) &\leq E_f(Z_m) + 8(\sqrt{\gamma} \vee L^2)\sqrt{D_m \log(3/\beta)} + 151(\sqrt{\gamma} \vee \mu \vee L^2) \log(3/\beta) \\ &\quad + 151(\sqrt{\gamma} \vee \mu \vee L^2) \left[1 + \frac{1}{I_m} \log\left(\frac{3D_m}{\beta}\right) \right] \log\left(\frac{1}{\alpha_m}\right) \\ &\quad + 24(\sqrt{\gamma} \vee \mu \vee L^2) \sqrt{7D_m \left[1 + \frac{1}{I_m} \log\left(\frac{3D_m}{\beta}\right) \right] \log\left(\frac{1}{\alpha_m}\right)}. \end{aligned}$$

By assumption, $\log(1/\alpha_m)$ and $\log(3/\beta)$ are positive, so we obtain (26) with

$$\begin{aligned} R_m &= 151(\sqrt{\gamma} \vee \mu \vee L^2) \left[1 + \frac{1}{I_m} \log\left(\frac{3D_m}{\beta}\right) \right] \log\left(\frac{3}{\beta\alpha_m}\right) \\ &\quad + 72(\sqrt{\gamma} \vee \mu \vee L^2) \sqrt{D_m \left[1 + \frac{1}{I_m} \log\left(\frac{3D_m}{\beta}\right) \right] \log\left(\frac{3}{\beta\alpha_m}\right)}. \end{aligned}$$

By Lemma 5, we have (28) with

$$R'_m = 6\sqrt{\gamma D_m/\beta} + 14(\sqrt{\gamma} \vee \mu \vee L^2) \log(6/\beta),$$

so (33) holds with Δ replaced by Δ_1 provided A_1, \dots, A_5 are large enough. \square

Proof of Theorem 5. By (30),

$$E_f(Z_m) \leq D_m(\mu + L^2).$$

Combining Lemmas 2 and 4 proves that with probability greater than $1 - 2\beta/3$,

$$\begin{aligned} \tilde{q}_m^y(\alpha_m) &\leq E_f(Z_m) + 4(\mu \vee L^2) \sqrt{\frac{6D_m}{\beta}} \\ &\quad + 8Cp(\mu \vee L^2) \left[1 + \frac{1}{\sqrt{I_m}} \left(\frac{3D_m}{\beta} \right)^{1/p} \right] \log \left(\frac{1}{\alpha_m} \right) \\ &\quad + 18(\mu \vee L^2) \sqrt{Cp \left(D_m + \sqrt{\frac{D_m}{\beta}} \right) \left[1 + \frac{1}{\sqrt{I_m}} \left(\frac{3D_m}{\beta} \right)^{1/p} \right] \log \left(\frac{1}{\alpha_m} \right)}. \end{aligned}$$

Using (20) and (21), we get (26) with

$$\begin{aligned} R_m &= 19(\mu \vee L^2) \sqrt{\frac{D_m}{\beta}} + 17Cp(\mu \vee L^2) \left[1 + \frac{1}{\sqrt{I_m}} \left(\frac{3D_m}{\beta} \right)^{1/p} \right] \log \left(\frac{1}{\alpha_m} \right) \\ &\quad + 18(\mu \vee L^2) \sqrt{CpD_m \left[1 + \frac{1}{\sqrt{I_m}} \left(\frac{3D_m}{\beta} \right)^{1/p} \right] \log \left(\frac{1}{\alpha_m} \right)}. \end{aligned}$$

By Lemma 6, we have (28) with

$$R'_m = 3\mu \sqrt{D_m/\beta} + 9\mu/\beta,$$

so (33) holds with Δ replaced by Δ_2 provided A_1, \dots, A_6 are large enough. \square

7.5. PROOF OF COROLLARY 1. We define δ and δ' in the following way. If (4) holds for all $p \geq 1$ and some γ and μ , we set $\delta = \delta' = \sqrt{\gamma} + \mu + L^2$. If (5) holds for some $p \geq 2$ and some μ , we set $\delta = \sqrt{p}(\mu + L^2)$ and $\delta' = p(\mu + L^2)$. By Step 2 in the proof of Corollary 1 of Baraud *et al.* (2003),

$$\|f - \Pi_{m_k} f\|_n^2 \leq nR^2 k^{-2s}$$

for all $k \in \mathcal{I}$. Moreover, $|\mathcal{M}| \leq \log n / \log 2$, so there exists $c_\alpha > 0$, which only depends on α , such that

$$\log(1/\alpha_m) \leq c_\alpha \log \log n.$$

We have (12) and $D_{m_k} \leq k$ for all $k \in \mathcal{I}$, so it follows from Theorem 4 that the power of the test is greater than $1 - \beta$ whenever (4) holds and

$$(34) \quad \begin{aligned} \|f\|_n^2 &\geq A \inf_{k \in \mathcal{I}} \left\{ nR^2 k^{-2s} + \delta \left(1 + \frac{k}{n} \log n \right) \log \log n \right. \\ &\quad \left. + \delta \sqrt{k \left(1 + \frac{k}{n} \log n \right) \log \log n} \right\}, \end{aligned}$$

for a large enough A . Likewise, it follows from Theorem 5 that the power of the test is greater than $1 - \beta$ whenever (5) holds and

$$(35) \quad \begin{aligned} \|f\|_n^2 &\geq A \inf_{k \in \mathcal{I}} \left\{ nR^2 k^{-2s} + \delta' \left(1 + \frac{k^{1/p+1/2}}{\sqrt{n}} \right) \log \log n \right. \\ &\quad \left. + \delta \sqrt{k \left(1 + \frac{k^{1/p+1/2}}{\sqrt{n}} \right) \log \log n} \right\} \end{aligned}$$

for a large enough A . In (34) and (35), $A > 0$ only depends on a , a_0 , α , and β . In the sequel, \mathcal{I}' denotes the subset of \mathcal{I} defined as follows. Under (4), \mathcal{I}' is the set of those $k \in \mathcal{I}$, which satisfy $k \log n \leq n$ and under (5), \mathcal{I}' is the set of those $k \in \mathcal{I}$, which satisfy $k^{1/p+1/2} \leq \sqrt{n}$, that is $k \leq n^{p/(p+2)}$.

• Assume first that (4) and 1 (a), resp. (5) and 2 (a), hold for a large enough C . By (34) and (35), the power of the test is greater than $1 - \beta$ whenever

$$\|f\|_n^2 \geq 2A \left[\inf_{k \in \mathcal{I}'} \{nR^2 k^{-2s} + \delta \sqrt{k \log \log n}\} + \delta' \log \log n \right].$$

Let

$$(36) \quad k^* = \left(\frac{nR^2}{\delta \sqrt{\log \log n}} \right)^{2/(1+4s)}.$$

We have

$$nR^2 k^{-2s} \leq \delta \sqrt{k \log \log n}$$

if and only if $k \geq k^*$ and for large enough n , there exists $k' \in \mathcal{I}'$ such that $k^* \leq k' \leq 2k^*$. Therefore,

$$\begin{aligned} \inf_{k \in \mathcal{I}'} \{nR^2 k^{-2s} + \delta \sqrt{k \log \log n}\} &\leq 2\delta \sqrt{k' \log \log n} \\ &\leq 4\delta \sqrt{k^* \log \log n} \leq 4nR^{2/(1+4s)} \left(\frac{\delta \sqrt{\log \log n}}{n} \right)^{4s/(1+4s)}. \end{aligned}$$

The power of the test is thus greater than $1 - \beta$ whenever n is large enough and

$$\|f\|_n^2 \geq 2A \left[4nR^{2/(1+4s)} \left(\frac{\delta \sqrt{\log \log n}}{n} \right)^{4s/(1+4s)} + \delta' \log \log n \right].$$

This indeed holds if n is large enough and either 1 (a) or 2 (a) is fulfilled for a large enough C .

• Assume (4). By (34), the power is greater than $1 - \beta$ whenever n is large enough and

$$(37) \quad \|f\|_n^2 \geq 3A \inf_{k \in \mathcal{I} \setminus \mathcal{I}'} \left\{ nR^2 k^{-2s} + \delta k \sqrt{\frac{1}{n} \log n \times \log \log n} \right\}$$

for a large enough A . Let

$$k_* = \left(\frac{n^{3/2} R^2}{\delta \sqrt{\log n \times \log \log n}} \right)^{1/(1+2s)}.$$

If $s = 1/4$ and n is large enough, there exists $k' \in \mathcal{I} \setminus \mathcal{I}'$ such that $k_* \leq k' \leq 2k_*$. Therefore,

$$\inf_{k \in \mathcal{I} \setminus \mathcal{I}'} \left\{ nR^2 k^{-2s} + \delta k \sqrt{\frac{1}{n} \log n \times \log \log n} \right\} \leq 4\delta k_* \sqrt{\frac{1}{n} \log n \times \log \log n}.$$

The power of the test is thus greater than $1 - \beta$ whenever n is large enough and 1 (b) holds for a large enough C . Let k_0 be a point in \mathcal{I} with $k_0 \geq n/8$. Then (37) holds whenever

$$\|f\|_n^2 \geq 3A \left\{ nR^2 k_0^{-2s} + \delta k_0 \sqrt{\frac{1}{n} \log n \times \log \log n} \right\}.$$

Since $k_0 \leq n/2$, this indeed holds if n is large enough and 1 (c) is fulfilled for a large enough C .

- Assume (5) holds for some $p \geq 2$. If n is large enough, we have for all $k \geq 1$

$$\delta' k^{1/p+1/2} n^{-1/2} \log \log n \leq \delta k^{1/2p+3/4} n^{-1/4} \sqrt{\log \log n}.$$

By (35), the power is thus greater than $1 - \beta$ whenever n is large enough and

$$(38) \quad \|f\|_n^2 \geq 4A \inf_{k \in \mathcal{I} \setminus \mathcal{I}'} \left\{ nR^2 k^{-2s} + \delta k^{1/2p+3/4} n^{-1/4} \sqrt{\log \log n} \right\}.$$

Let

$$k_* = \left(\frac{n^{5/4} R^2}{\delta \sqrt{\log \log n}} \right)^{4p/(2+3p+8ps)}.$$

If $s \in [1/4 - 1/p, 1/4 + 1/p)$ and n is large enough, there exists $k' \in \mathcal{I} \setminus \mathcal{I}'$ such that $k_* \leq k' \leq 2k_*$. Therefore,

$$\inf_{k \in \mathcal{I} \setminus \mathcal{I}'} \left\{ nR^2 k^{-2s} + \delta k^{1/2p+3/4} n^{-1/4} \sqrt{\log \log n} \right\} \leq 4\delta k_*^{1/2p+3/4} n^{-1/4} \sqrt{\log \log n}.$$

The power of the test is thus greater than $1 - \beta$ whenever n is large enough and 2 (b) holds for a large enough C . Let k_0 be a point in \mathcal{I} with $k_0 \geq n/8$. Then (38) holds whenever

$$\|f\|_n^2 \geq 4A \left\{ nR^2 k_0^{-2s} + \delta k_0^{1/2p+3/4} n^{-1/4} \sqrt{\log \log n} \right\}.$$

This indeed holds if n is large enough and 2 (c) is fulfilled for a large enough C . \square

7.6. PROOF OF COROLLARY 2. Let k^* be given by (36) and let $k' \in \mathcal{I}$ be such that $k^* \leq k' \leq 2k^*$ (such a k' exists provided n is large enough). By Theorems 4 and 5, it suffices to prove that $\|f\|_n^2$ exceeds $\Delta(m_{k'}, f, A)$ for large enough n and a fixed A , where Δ denotes either Δ_1 or Δ_2 . By (31), one can choose

$$A_1 = \frac{2/3}{2/3 - 1/I_{m_{k'}}}.$$

Moreover, there exists $A_0 > 0$, which only depends on α , β , and A such that

$$\Delta(m_{k'}, f, A) \leq A_1 \|f - \Pi_m f\|_n^2 + A_0 \delta \sqrt{k' \log \log n}.$$

By (32), it thus suffices to show that for large enough n ,

$$(39) \quad \|\Pi_{m_{k'}} f\|_n^2 \geq \frac{3}{2I_{m_{k'}}} \|f\|_n^2 + A_0 \delta \sqrt{k' \log \log n}.$$

We need further notation. Let F_n be the function defined by $F_n(t) = F(x_i)$ for all $t \in (x_{i-1}, x_i]$, $i = 1, \dots, n$, where we set $x_i = i/n$ for all $i = 0, \dots, n$. For every $j = 0, \dots, k'$, let

$$t_j = \frac{1}{n} \sum_{i \leq j} |e_{m_{k'}, i}| = \frac{1}{n} \left[\frac{nj}{k'} \right]$$

(here, $[x]$ denotes the integer part of x). Let Q be the orthogonal projector from $L_2[0, 1]$ onto the set of step functions, which are constant on each interval $(t_{j-1}, t_j]$, and let Q_r be the orthogonal projector from $L_2[0, 1]$ onto the set of step functions, which are constant on each interval $((j-1)/k', j/k']$, $j = 1, \dots, k'$. Since $t_{j-1} - t_j \leq 2/k'$ and $|t_j - j/k'| \leq 1/n$, we have

$$\|QF\|_2^2 = \sum_{j=1}^{k'} \frac{1}{t_j - t_{j-1}} \left(\int_{t_{j-1}}^{t_j} F(x) dx \right)^2 \geq \frac{1}{4} \|Q_r F\|_2^2 - \frac{2k'^2 \|F\|_\infty^2}{n^2}$$

(recall that $(a+b)^2 \geq a^2/2 - b^2$ for all real numbers a and b). Now,

$$\|Q(F - F_n)\|_2 \leq \|F - F_n\|_2 \leq \|F - F_n\|_\infty \leq \|F'\|_\infty / n$$

and we have

$$\|\Pi_{m_{k'}} f\|_n = \sqrt{n} \|QF_n\|_2.$$

Therefore,

$$\|\Pi_{m_{k'}} f\|_n \geq \sqrt{n} \|QF\|_2 - \|F'\|_\infty / \sqrt{n} \geq \frac{\sqrt{n}}{2} \|Q_r F\|_2 - \frac{\sqrt{2}k' \|F\|_\infty}{\sqrt{n}} - \|F'\|_\infty / \sqrt{n}.$$

The partition of $[0, 1]$ associated with Q_r is equispaced (all intervals have the same length $1/k'$), so one can prove that there exist positive numbers C_1 and C_2 , which only depend on s , such that

$$\|Q_r F\|_2 \geq C_1 \|F\|_2 - C_2 R k'^{-s},$$

see Proposition 2.16 of Ingster and Suslina (2003). It then follows from the definition of k' and the assumption $s > 1$ that there exist positive numbers C'_1 and C'_2 , which only depend on s , such that for large enough n ,

$$\|\Pi_{m_{k'}} f\|_n \geq C'_1 \sqrt{n} \|F\|_2 - C'_2 \sqrt{n} R k'^{-s}.$$

Note that

$$\|f\|_n = \sqrt{n} \|F_n\|_2 \leq \sqrt{n} \|F\|_2 + \|F\|_\infty / \sqrt{n}$$

and that $I_{m_{k'}}$ tends to infinity as n goes to infinity. Thus in order to prove (39), it suffices to prove that for large enough n and some $A'_0 > 0$,

$$\|F\|_2^2 \geq A'_0 \left\{ R^2 k'^{-2s} + \frac{\delta}{n} \sqrt{k' \log \log n} \right\}.$$

But this is indeed the case if F satisfies (15) for a large enough $C > 0$. \square

7.7. POWER IN A GAUSSIAN HOMOSCEDASTIC MODEL. In this section, we assume that the ε_i 's are i.i.d. $\mathcal{N}(0, \sigma^2)$ and $|f_i| \leq \sigma L$ for all i . We will prove that the power of the test is greater than $1 - \beta$ whenever f satisfies (7) for large enough A_k 's. Under the above assumptions, Lemmas 2 and 3 are valid with $\mu = 2\sigma^2$, $\gamma = \mu^2$, and L^2 replaced by $\sigma^2 L^2$. In particular, we have (26) with

$$\begin{aligned} R_m &= 302\sigma^2(1 + L^2) \left[1 + \frac{1}{I_m} \log \left(\frac{3D_m}{\beta} \right) \right] \log \left(\frac{6}{\beta\alpha_m} \right) \\ &\quad + 144\sigma^2(1 + L^2) \sqrt{D_m \left[1 + \frac{1}{I_m} \log \left(\frac{3D_m}{\beta} \right) \right] \log \left(\frac{6}{\beta\alpha_m} \right)}. \end{aligned}$$

Moreover, one can improve the result given in Lemma 5 under the Gaussian assumption, due to the Cochran theorem. Indeed, let D'_m denote the cardinality of $J_{m,2}$. By the Cochran Theorem, $\|\Pi_m \tilde{y}\|_n^2 / \sigma^2$ is a non-central χ^2 variable with D'_m degrees of freedom and non-centrality parameter $\|\Pi_m \tilde{f}\|_n^2 / \sigma^2$. By Lemma 1 of Birgé (2001), we thus have for all positive x

$$P_f \left[\frac{1}{\sigma^2} \|\Pi_m \tilde{y}\|_n^2 \leq \frac{1}{\sigma^2} E_f \|\Pi_m \tilde{y}\|_n^2 - 2 \sqrt{\left(D'_m + \frac{2}{\sigma^2} \|\Pi_m \tilde{f}\|_n^2 \right) x} \right] \leq \exp(-x).$$

Using (20) and (21) one obtains

$$P_f \left[\|\Pi_m \tilde{y}\|_n^2 \leq E_f \|\Pi_m \tilde{y}\|_n^2 - 2\sigma^2 \sqrt{D'_m x} - \frac{1}{3} \|\Pi_m \tilde{f}\|_n^2 - 6\sigma^2 x \right] \leq \exp(-x),$$

for all $x > 0$. Setting $x = \log(3/\beta)$ in this inequality yields

$$P_f \left[\|\Pi_m \tilde{y}\|_n^2 \leq E_f \|\Pi_m \tilde{y}\|_n^2 - 2\sigma^2 \sqrt{D_m \log \left(\frac{3}{\beta} \right)} - \frac{1}{3} \|\Pi_m \tilde{f}\|_n^2 - 6\sigma^2 \log \left(\frac{3}{\beta} \right) \right] \leq \frac{\beta}{3}.$$

Therefore, we have (28) with

$$R'_m = 2\sigma^2 \sqrt{D_m \log(3/\beta)} + 6\sigma^2 \log(3/\beta).$$

Let $\Delta(m, f, A)$ denote the bracketed term in (7). Then (32) holds provided the A_k 's are large enough, which proves the announced result. \square

7.8. PROOF OF THE LEMMAS. We first recall two inequalities, which will be used in this section.

Bernstein's Inequality. Let X_1, \dots, X_n be independent real-valued random variables. Assume that there exist positive numbers v and c such that for all integers $k \geq 2$

$$\sum_{i=1}^n E[|X_i|^k] \leq k! \frac{v}{2} c^{k-2}.$$

Let $S = \sum_{i=1}^n (X_i - E[X_i])$. Then for every $x > 0$, $P[S \geq \sqrt{2vx} + cx] \leq \exp(-x)$.

Rosenthal's Inequality. Let X_1, \dots, X_n be independent centered real-valued random variables with finite t -th moment, $2 \leq t < \infty$. Then there exists an absolute constant L such that

$$E\left|\sum_{i=1}^n X_i\right|^t \leq L^t t^t \max\left(\sum_{i=1}^n E|X_i|^t, \left(\sum_{i=1}^n E|X_i|^2\right)^{t/2}\right).$$

Proof of Lemma 1. For notational convenience, we omit subscript m . For every $j = 1, \dots, D$ and $z \in \mathbb{R}^n$, define the set $\mathcal{A}_j(z)$ by $\mathcal{A}_j(z) = \{0\}$ if $D = 1$ and

$$\mathcal{A}_j(z) = \left\{ \sum_{k \neq j} \frac{1}{|e_k|} \left(\sum_{i \in e_k} u_i z_i \right)^2, u \in \{\pm 1\}^n \right\}$$

if $D > 1$. For every $j = 1, \dots, D$ let

$$\mathcal{E}_j = \{e \subset e_j \text{ s.t. } e \neq \emptyset \text{ and } e \neq e_j\}.$$

Finally, let

$$\mathcal{Y} = \bigcap_{j=1}^D \left\{ z \in \mathbb{R}^n \text{ s.t. } \forall u \in \{\pm 1\}^n, \forall a, a' \in \mathcal{A}_j(z), \forall e \in \mathcal{E}_j, \right. \\ \left. 4 \sum_{i \in e} u_i z_i \sum_{i \in e_j \setminus e} u_i z_i \neq (a' - a)|e_j| \right\}.$$

The ε_i 's are independent and all have a continuous distribution, so the event $\{\varepsilon \in \mathcal{Y}\}$ is the intersection of a finite number of events having probability one. We thus assume in the sequel without loss of generality that $\varepsilon \in \mathcal{Y}$. Let u and u' be elements of $\{\pm 1\}^n$ such that

$$(40) \quad \|\Pi(u \times \varepsilon)\|_n^2 = \|\Pi(u' \times \varepsilon)\|_n^2.$$

Fix $j \in \{1, \dots, D\}$. We have

$$0 = \|\Pi(u \times \varepsilon)\|_n^2 - \|\Pi(u' \times \varepsilon)\|_n^2 = \frac{1}{|e_j|} \left[\left(\sum_{i \in e_j} u_i \varepsilon_i \right)^2 - \left(\sum_{i \in e_j} u'_i \varepsilon_i \right)^2 \right] + a - a'$$

for some $a, a' \in \mathcal{A}_j(\varepsilon)$. Setting $e = \{i \in e_j \text{ s.t. } u_i = u'_i\}$ we get

$$\left(\sum_{i \in e_j} u_i \varepsilon_i \right)^2 - \left(\sum_{i \in e_j} u'_i \varepsilon_i \right)^2 = \sum_{i \in e_j} (u_i + u'_i) \varepsilon_i \sum_{i \in e_j} (u_i - u'_i) \varepsilon_i = 4 \sum_{i \in e} u_i \varepsilon_i \sum_{i \in e_j \setminus e} u_i \varepsilon_i,$$

since $u_i = -u'_i$ for every $i \in e_j \setminus e$. As $\varepsilon \in \mathcal{Y}$, either $e = \emptyset$ or $e = e_j$. Thus for all $j = 1, \dots, D$, either $u_i = u'_i$ for all $i \in e_j$ or $u_i = -u'_i$ for all $i \in e_j$. This implies that the cardinality of the set

$$\{\|\Pi(u \times \varepsilon)\|_n^2, u \in \{\pm 1\}^n\}$$

is 2^{n-D} and that for every element a of this set, the cardinality of the set

$$\{u \in \{\pm 1\}^n \text{ s.t. } a = \|\Pi(u \times \varepsilon)\|_n^2\}$$

is equal to 2^D . This proves that conditionally on ε , $\|\Pi(w \times \varepsilon)\|_n^2$ has a discrete uniform distribution on a set with 2^{n-D} distinct values. Clearly, $\|\Pi\varepsilon\|_n^2$ belongs to this set, so $p^\varepsilon = k2^{D-n}$ for some $k = 1, \dots, 2^{n-D}$. More precisely, we have $p^\varepsilon = k2^{D-n}$ if and only if the cardinality of the following set is equal to k :

$$\{\|\Pi(u \times \varepsilon)\|_n^2 \text{ s.t. } \|\Pi(u \times \varepsilon)\|_n^2 \geq \|\Pi\varepsilon\|_n^2, u \in \{\pm 1\}^n\}.$$

But this set has the same cardinality as the set

$$\{\|\Pi(u \times |\varepsilon|)\|_n^2 \text{ s.t. } \|\Pi(u \times |\varepsilon|)\|_n^2 \geq \|\Pi(\text{sign}(\varepsilon) \times |\varepsilon|)\|_n^2, u \in \{\pm 1\}^n\},$$

where we recall that $\text{sign}(\varepsilon)$ is the vector in \mathbb{R}^n defined by (17). Since $\text{sign}(\varepsilon)$ has the same distribution as w and is independent of $|\varepsilon|$ (see the proof of Theorem 1) we get

$$P(p^\varepsilon = k2^{D-n} \mid |\varepsilon|) = P\left(\sum_{u \in \{\pm 1\}^n} \mathbf{1}_{\|\Pi(u \times |\varepsilon|)\|_n^2 \geq \|\Pi(w \times |\varepsilon|)\|_n^2} = k \mid |\varepsilon|\right).$$

Moreover conditionally on $|\varepsilon|$, $\|\Pi(w \times |\varepsilon|)\|_n^2$ has a uniform discrete distribution on a set with 2^{n-D} distinct values. Therefore,

$$P(p^\varepsilon = k2^{D-n} \mid |\varepsilon|) = 2^{D-n}.$$

Integrating this inequality yields the result. \square

Proof of Lemma 2. First, note that Lemma 2 is trivial whenever $\tilde{y}_i = 0$ for all i since in that case $\tilde{q}_m^y(\alpha_m) = 0$. We thus assume in the sequel that there exists i such that $\tilde{y}_i \neq 0$. In particular, $M_m > 0$. Now, note that Z_m can be written as the sum of squared random variables:

$$Z_m = \sum_{j \in J_{m,2}} X_j^2,$$

where for every $j \in J_{m,2}$,

$$X_j = \sum_{i \in \mathcal{J}} \frac{w_i y_i}{\sqrt{|e_{m,j}|}} \mathbf{1}_{e_{m,j}}(i).$$

Here,

$$(41) \quad \mathcal{J} = \bigcup_{j \in J_{m,2}} e_{m,j}.$$

Denote by S_m the unit sphere in $\mathbb{R}^{|J_{m,2}|}$. Then

$$Z_m^{1/2} = \sup_{a \in S_m} \sum_{j \in J_{m,2}} a_j X_j = \sup_{a \in S_m} \sum_{i \in \mathcal{J}} w_i \frac{y_i a_{j(i)}}{\sqrt{|e_{m,j(i)}|}},$$

where for every i , $j(i)$ denotes the unique integer in $J_{m,2}$ that satisfies $i \in e_{m,j(i)}$. If S'_m denotes a finite subset of S_m , then it follows from a result of Massart (2006) that for all $x \geq 0$,

$$P_f \left[\sup_{a \in S'_m} \sum_{j \in J_{m,2}} a_j X_j \geq E_f \left(\sup_{a \in S'_m} \sum_{j \in J_{m,2}} a_j X_j \mid y \right) + x \mid y \right] \leq \exp \left(- \frac{x^2}{8\sigma^2(S'_m)} \right),$$

where

$$\sigma^2(S'_m) = \sup_{a \in S'_m} \sum_{i \in \mathcal{J}} \frac{y_i^2 a_{j(i)}^2}{|e_{m,j(i)}|} = \sup_{a \in S'_m} \sum_{j \in J_{m,2}} \frac{a_j^2}{|e_{m,j}|} \sum_{i \in e_{m,j}} y_i^2.$$

But S_m is separable and for every subset S'_m of S_m , $\sigma^2(S'_m) \leq M_m$. Hence for all $x \geq 0$,

$$P_f [Z_m^{1/2} \geq E_f(Z_m^{1/2} \mid y) + x \mid y] \leq \exp(-x^2/8M_m).$$

In particular,

$$P_f [Z_m^{1/2} \geq E_f(Z_m^{1/2} \mid y) + \sqrt{8M_m \log(1/\alpha_m)} \mid y] \leq \alpha_m.$$

Squaring the inequality yields

$$P_f \left[Z_m \geq E_f^2(Z_m^{1/2} \mid y) + 8M_m \log(1/\alpha_m) + 2E_f(Z_m^{1/2} \mid y) \sqrt{8M_m \log(1/\alpha_m)} \mid y \right] \leq \alpha_m.$$

By definition of $\tilde{q}_m^y(\alpha_m)$ we thus have

$$\tilde{q}_m^y(\alpha_m) \leq E_f^2(Z_m^{1/2} \mid y) + 8M_m \log(1/\alpha_m) + 2E_f(Z_m^{1/2} \mid y) \sqrt{8M_m \log(1/\alpha_m)},$$

and the result follows from the Jensen inequality, which implies

$$E_f^2(Z_m^{1/2} \mid y) \leq E_f(Z_m \mid y). \quad \square$$

Proof of Lemma 3. By definition of M_m , we have for all real numbers x and c

$$P_f [M_m \geq c + x] \leq \sum_{j \in J_{m,2}} P_f \left[\frac{1}{|e_{m,j}|} \sum_{i \in e_{m,j}} y_i^2 \geq c + x \right].$$

In particular, if c is large enough so that $E_f(y_i^2) \leq c$ for all i , then for all x

$$(42) \quad P_f[M_m \geq c + x] \leq \sum_{j \in J_{m,2}} P_f \left[\frac{1}{|e_{m,j}|} \sum_{i \in e_{m,j}} (y_i^2 - E_f(y_i^2)) \geq x \right].$$

Let $p \geq 1$. By (22), $y_i^{2p} \leq 2^{2p}(\varepsilon_i^{2p} \vee L^{2p})$ and in particular, $y_i^{2p} \leq 2^{2p}(\varepsilon_i^{2p} + L^{2p})$. Therefore,

$$(43) \quad E_f(y_i^{2p}) \leq 2^{2p+1} \max \{ \gamma p! \mu^{p-2}, L^{2p} \} \leq \gamma_0 p! \mu_0^{p-2},$$

where $\gamma_0 = 2^5(\gamma \vee L^4)$ and $\mu_0 = 4(\mu \vee L^2)$. By Bernstein's inequality, we thus have for all $x \geq 0$

$$P_f \left[\sum_{i \in e_{m,j}} (y_i^2 - E_f(y_i^2)) \geq 2\sqrt{\gamma_0 |e_{m,j}| x} + \mu_0 x \right] \leq \exp(-x).$$

Moreover, $E_f(y_i^2) \leq L^2 + \sqrt{2\gamma}$, so by (42),

$$P_f \left[M_m \geq \sqrt{2\gamma} + L^2 + 2\sqrt{\frac{\gamma_0 x}{I_m}} + \frac{\mu_0 x}{I_m} \right] \leq D_m \exp(-x).$$

We have

$$\begin{aligned} \sqrt{2\gamma} + L^2 + 2\sqrt{\frac{\gamma_0 x}{I_m}} + \frac{\mu_0 x}{I_m} &\leq \sqrt{2\gamma} + L^2 + (2\sqrt{\gamma_0} + \mu_0) \left(1 + \frac{x}{I_m} \right) \\ &\leq 18(\sqrt{\gamma} \vee \mu \vee L^2) \left(1 + \frac{x}{I_m} \right). \end{aligned}$$

Hence for all $x > 0$,

$$P_f \left[M_m \geq 18(\sqrt{\gamma} \vee \mu \vee L^2) \left(1 + \frac{x}{I_m} \right) \right] \leq D_m \exp(-x).$$

Setting here $x = \log(3D_m/\beta)$ yields the first assertion in Lemma 3. The variables w_i are independent with mean zero and variance 1, and these variables are independent from y . Hence conditionally on y , they are still independent with mean zero and variance 1, and we derive from the definition of Z_m that

$$(44) \quad E_f(Z_m | y) = \sum_{j \in J_{m,2}} \frac{1}{|e_{m,j}|} \sum_{i \in e_{m,j}} y_i^2.$$

By (43),

$$\sum_{j \in J_{m,2}} \sum_{i \in e_{m,j}} E_f \left[\left(\frac{y_i^2}{|e_{m,j}|} \right)^p \right] \leq \frac{D_m \gamma_0}{2} p! \left(\frac{\mu_0}{2} \right)^{p-2},$$

so Bernstein's inequality yields for all $x \geq 0$

$$P_f \left[E_f(Z_m | y) \geq E_f(Z_m) + \sqrt{2D_m \gamma_0 x} + \frac{1}{2} \mu_0 x \right] \leq \exp(-x).$$

Setting here $x = \log(3/\beta)$ completes the proof of the lemma. \square

Proof of Lemma 4. By Rosenthal's inequality, there exists an absolute constant $C' > 0$ such that for all j

$$\begin{aligned} E_f \left| \sum_{i \in e_{m,j}} (y_i^2 - E_f(y_i^2)) \right|^p \\ \leq (C'p)^p \max \left\{ \sum_{i \in e_{m,j}} E_f |y_i^2 - E_f(y_i^2)|^p, \left(\sum_{i \in e_{m,j}} \text{var}(y_i^2) \right)^{p/2} \right\}. \end{aligned}$$

Moreover, by Jensen's inequality, $E\varepsilon_i^4 \leq \mu^2$ for all i , so by (22),

$$\text{var}_f(y_i^2) \leq E_f(y_i^4) \leq 2^4(\mu^2 + L^4).$$

Hence,

$$\left(\sum_{i \in e_{m,j}} \text{var}_f(y_i^2) \right)^{p/2} \leq 2^{5p/2} |e_{m,j}|^{p/2} (\mu^p \vee L^{2p}).$$

On the other hand,

$$\sum_{i \in e_{m,j}} E_f |y_i^2 - E_f(y_i^2)|^p \leq 2^p \sum_{i \in e_{m,j}} [E_f(y_i^{2p}) + |E_f(y_i^2)|^p] \leq 2^{p+1+2p} |e_{m,j}| (\mu^p + L^{2p}),$$

so we get

$$E_f \left| \sum_{i \in e_{m,j}} (y_i^2 - E_f(y_i^2)) \right|^p \leq (Cp)^p |e_{m,j}|^{p/2} (\mu^p \vee L^{2p}),$$

where for instance $C = 2^4 C'$. We have $E_f(y_i^2) \leq \mu + L^2$, so (42) and Markov's inequality yield

$$P_f[M_m \geq \mu + L^2 + x] \leq (Cp)^p (\mu^p \vee L^{2p}) \sum_{j \in J_{m,2}} \frac{1}{x^p |e_{m,j}|^{p/2}} \leq (Cp)^p (\mu^p \vee L^{2p}) \frac{D_m}{x^p I_m^{p/2}}$$

for all $x > 0$. In particular,

$$P_f \left[M_m \geq \mu + L^2 + Cp(\mu \vee L^2) \frac{1}{\sqrt{I_m}} \left(\frac{3D_m}{\beta} \right)^{1/p} \right] \leq \frac{\beta}{3}.$$

We can assume without loss of generality that $C \geq 1$, so the first assertion of Lemma 4 follows. We have (44), where the y_i 's are independent. Therefore,

$$\text{var}_f[E_f(Z_m | y)] \leq \max_{1 \leq i \leq n} E_f(y_i^4) D_m \leq 2^4(\mu^2 + L^4) D_m.$$

Now the second assertion of the lemma follows from the Bienaymé–Chebyshev inequality. \square

Proof of Lemma 5. Let \mathcal{J} be defined by (41). We have

$$\|\Pi_m \tilde{y}\|_n^2 = \|\Pi_m \tilde{\varepsilon}\|_n^2 + \|\Pi_m \tilde{f}\|_n^2 + 2 \sum_{j \in J_{m,2}} \frac{1}{|e_{m,j}|} \left(\sum_{i \in e_{m,j}} \varepsilon_i \right) \left(\sum_{i \in e_{m,j}} f_i \right),$$

where $\tilde{\varepsilon} = \tilde{y} - \tilde{f}$. For every $i \in \mathcal{J}$, let $j(i)$ denote the integer in $J_{m,2}$ which satisfies $i \in e_{m,j(i)}$ and let

$$a_i = \frac{1}{|e_{m,j(i)}|} \sum_{i \in e_{m,j(i)}} f_i.$$

Then,

$$\sum_{j \in J_{m,2}} \frac{1}{|e_{m,j}|} \left(\sum_{i \in e_{m,j}} \varepsilon_i \right) \left(\sum_{i \in e_{m,j}} f_i \right) = \sum_{i \in \mathcal{J}} a_i \varepsilon_i.$$

Since $|a_i| \leq L$, we have for all integers $p \geq 1$

$$\sum_{i \in \mathcal{J}} E|a_i \varepsilon_i|^p \leq L^{p-2} \sum_{i \in \mathcal{J}} a_i^2 \sqrt{\gamma} \sqrt{p!} \mu^{p/2-1} \leq L^{p-2} \|\Pi_m \tilde{f}\|_n^2 \sqrt{\gamma} p! \mu^{p/2-1}.$$

Using Bernstein's inequality, we obtain that for all $x \geq 0$

$$P_f \left[\sum_{j \in J_{m,2}} \frac{1}{|e_{m,j}|} \left(\sum_{i \in e_{m,j}} \varepsilon_i \right) \left(\sum_{i \in e_{m,j}} f_i \right) \leq -2\sqrt{\|\Pi_m \tilde{f}\|_n^2 \sqrt{\gamma} x} - L\sqrt{\mu} x \right] \leq \exp(-x).$$

We have $L\sqrt{\mu} \leq L^2 \vee \mu$ and by (21),

$$2\sqrt{\|\Pi_m \tilde{f}\|_n^2 \sqrt{\gamma} x} \leq \frac{1}{6} \|\Pi_m \tilde{f}\|_n^2 + 6\sqrt{\gamma} x.$$

Setting $x = \log(6/\beta)$ thus yields

$$\begin{aligned} P_f \left[\sum_{j \in J_{m,2}} \frac{1}{|e_{m,j}|} \left(\sum_{i \in e_{m,j}} \varepsilon_i \right) \left(\sum_{i \in e_{m,j}} f_i \right) \right. \\ \left. \leq -\frac{1}{6} \|\Pi_m \tilde{f}\|_n^2 - 7(\sqrt{\gamma} \vee \mu \vee L^2) \log \left(\frac{6}{\beta} \right) \right] \leq \frac{\beta}{6}. \end{aligned}$$

On the other hand, the ε_i 's are mutually independent and ε_i and ε_i^3 have mean zero, so

$$(45) \quad \text{var} \left(\sum_{i \in e_{m,j}} \varepsilon_i \right)^2 \leq E \left(\sum_{i \in e_{m,j}} \varepsilon_i \right)^4 \leq 3|e_{m,j}|^2 \max_{1 \leq i \leq n} E(\varepsilon_i^4).$$

Therefore,

$$\text{var} (\|\Pi_m \tilde{\varepsilon}\|_n^2) \leq 6\gamma D_m$$

and it follows from the Bienaymé–Chebyshev inequality that

$$P_f \left[\|\Pi_m \tilde{\varepsilon}\|_n^2 \leq E_f \|\Pi_m \tilde{\varepsilon}\|_n^2 - 6\sqrt{\gamma D_m / \beta} \right] \leq \beta/6.$$

The result follows. \square

Proof of Lemma 6. The distribution of ε_i is symmetric, so ε_i and ε_i^3 have mean zero. Moreover, the ε_i 's are mutually independent, so

$$\text{var}_f(\|\Pi_m \tilde{y}\|_n^2) = \sum_{j \in J_{m,2}} \frac{1}{|e_{m,j}|^2} \left[\text{var} \left(\sum_{i \in e_{m,j}} \varepsilon_i \right)^2 + 4 \left(\sum_{i \in e_{m,j}} f_i \right)^2 \sum_{i \in e_{m,j}} E(\varepsilon_i^2) \right].$$

It then follows from (45) that

$$\text{var}_f(\|\Pi_m \tilde{y}\|_n^2) \leq 3D_m \mu^2 + 4\|\Pi_m \tilde{f}\|_n^2 \mu.$$

By the Bienaymé–Chebyshev inequality,

$$P_f \left[\|\Pi_m \tilde{y}\|_n^2 \leq E_f \|\Pi_m \tilde{y}\|_n^2 - \sqrt{\frac{9D_m \mu^2 + 12\|\Pi_m \tilde{f}\|_n^2 \mu}{\beta}} \right] \leq \frac{\beta}{3}.$$

By (20) and (21) we have

$$\sqrt{\frac{9D_m \mu^2 + 12\|\Pi_m \tilde{f}\|_n^2 \mu}{\beta}} \leq 3\mu \sqrt{\frac{D_m}{\beta}} + \frac{1}{3} \|\Pi_m \tilde{f}\|_n^2 + \frac{9\mu}{\beta}$$

and therefore,

$$P_f \left[\|\Pi_m \tilde{y}\|_n^2 \leq E_f \|\Pi_m \tilde{y}\|_n^2 - 3\mu \sqrt{\frac{D_m}{\beta}} - \frac{1}{3} \|\Pi_m \tilde{f}\|_n^2 - \frac{9\mu}{\beta} \right] \leq \frac{\beta}{3}.$$

This completes the proof of the lemma. \square

Acknowledgments. The second author started to work on this subject while visiting the department *Industrial Engineering and Management* of the Technion of Haïfa, Israel. During this period, he had several valuable discussions on the subject with Felix Abramovitch from Tel Aviv University. The authors thank Felix Abramovitch for his contribution at that time.

References

- [1] Y. Baraud, S. Huet, and B. Laurent (2003), *Adaptive tests of linear hypotheses by model selection*, Ann. Statist., **31**, 225–251.
- [2] Y. Baraud, S. Huet, and B. Laurent (2005), *Testing convex hypotheses on the mean of a Gaussian vector. Application to testing qualitative hypotheses on a regression function*, Ann. Statist., **33**, 214–257.
- [3] L. Birgé (2001), *An alternative point of view on Lepski's method*. In: *State of the art in probability and statistics (Leiden, 1999)*, Inst. Math. Statist., Beachwood, OH, pp. 113–133.
- [4] J.-C. Chen (1994), *Testing for no effect in nonparametric regression via spline smoothing techniques*, Ann. Inst. Statist. Math., **46**, 251–265.
- [5] M. S. Ermakov (1991), *Minimax detection of a signal in a Gaussian white noise*, Theory Probab. Appl., **35**, 667–679.
- [6] R. L. Eubank and J. D. Hart (1992), *Testing goodness-of-fit in regression via order selection criteria*, Ann. Statist., **20**, 1412–1425.

- [7] G. Gayraud and C. Pouet (2005), *Adaptive minimax testing in the discrete regression scheme*, Probab. Theory Rel. Fields., **133**, 531–558.
- [8] W. Härdle and A. Kneip (1999), *Testing a regression model when we have smooth alternatives in mind*, Scand. J. Statist., **26**, 221–238.
- [9] J. D. Hart (1997), *Nonparametric Smoothing and Lack-of-Fit Tests*, Springer Series in Statistics, New York.
- [10] J. L. Horowitz and V. G. Spokoiny (2001), *An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative*, Econometrica, **69**, 599–631.
- [11] Yu. I. Ingster (1982), *Minimax nonparametric detection of signals in white Gaussian noise*, Problems Inform. Transmission, **18**, 130–140.
- [12] Yu. I. Ingster (1993), *Asymptotically minimax hypothesis testing for nonparametric alternatives*. I, Math. Methods Statist., **2**, 85–114.
- [13] Yu. I. Ingster and I. A. Suslina (2003), *Nonparametric Goodness-of-Fit testing under Gaussian Models*, Springer Lecture Notes in Statistics, Vol. 169, New York.
- [14] O. V. Lepski and V. G. Spokoiny (1999), *Minimax nonparametric hypothesis testing: the case of an inhomogeneous alternative*, Bernoulli, **5**, 333–358.
- [15] P. Massart (2006), *Concentration Inequalities and Model Selection*, Lectures on Probability Theory and Statistics (Saint-Flour, 2003), Springer, to appear.
- [16] H.-G. Müller (1992), *Goodness-of-fit diagnostics for regression models*, Scand. J. Statist., **19**, 157–172.
- [17] J. G. Staniswalis and T. A. Severini (1991), *Diagnostics for assessing regression models*, J. Amer. Statist. Assoc., **86**, 684–692.
- [18] V. G. Spokoiny (1996), *Adaptive hypothesis testing using wavelets*, Ann. Statist., **24**, 2477–2498.

[Received August 2005; revised February 2006]