

ADAPTIVE TESTS OF QUALITATIVE HYPOTHESES

YANNICK BARAUD¹, SYLVIE HUET² AND BÉATRICE LAURENT³

Abstract. We propose a test of a qualitative hypothesis on the mean of a n -Gaussian vector. The testing procedure is available when the variance of the observations is unknown and does not depend on any prior information on the alternative. The properties of the test are non-asymptotic. For testing positivity or monotonicity, we establish separation rates with respect to the Euclidean distance, over subsets of \mathbb{R}^n which are related to Hölderian balls in functional spaces. We provide a simulation study in order to evaluate the procedure when the purpose is to test monotonicity in a functional regression model and to check the robustness of the procedure to non-Gaussian errors.

Mathematics Subject Classification. 62G10, 62G20.

Received April 17, 2002.

1. INTRODUCTION

We consider the statistical model

$$\mathbf{Y} = \mathbf{f} + \boldsymbol{\varepsilon}, \quad (1)$$

where \mathbf{f} is an unknown vector of \mathbb{R}^n and $\boldsymbol{\varepsilon}$ is a Gaussian vector with i.i.d. components of mean 0 and unknown variance σ^2 . Let \mathcal{C} be some set in \mathbb{R}^n containing 0 and satisfying the following condition:

Condition (A): For all $\mathbf{g}, \mathbf{h} \in \mathcal{C}$, $\mathbf{g} + \mathbf{h} \in \mathcal{C}$.

We call such a set \mathcal{C} a cone.

We want to test that \mathbf{f} belongs to \mathcal{C} against that it does not. For illustration of sets \mathcal{C} , one can take

$$\mathcal{C}_{\geq 0} = \{\mathbf{f} \in \mathbb{R}^n, \forall i \in \{1, \dots, n\}, f_i \geq 0\},$$

to test that the coordinates of \mathbf{f} are nonnegative, or

$$\mathcal{C}_{\nearrow} = \{\mathbf{f} \in \mathbb{R}^n, f_1 \leq f_2 \leq \dots \leq f_n\},$$

to test that the sequence of the coordinates of \mathbf{f} is nondecreasing.

Keywords and phrases: Adaptive test, test of monotonicity, test of positivity, qualitative hypothesis testing, nonparametric alternative, nonparametric regression.

¹ École Normale Supérieure, DMA, 45 rue d'Ulm, 75230 Paris Cedex 05, France; e-mail: yannick.baraud@ens.fr

² Unité BIA, 78352 Jouy-en-Josas Cedex, France; e-mail: Sylvie.Huet@jouy.inra.fr

³ bâtiment 425, Université Paris-Sud, 91405 Orsay Cedex, France; e-mail: beatrice.laurent@math.u-psud.fr

Our test relies on the following idea: we consider a suitable collection $\{S_m, m \in \mathcal{M}\}$ of linear subspaces of \mathbb{R}^n and, denoting by Π_m the orthogonal projector onto S_m , we reject the null hypothesis if there exists some m in \mathcal{M} such that the Euclidean distance $d(\cdot, \cdot)$ between $\Pi_m \mathbf{Y}$ and \mathcal{C} is large. Our results are non asymptotic: under the assumption that the ε_i 's are Gaussian, for each n and for each $\alpha \in]0, 1[$, we propose a level- α test and we characterise a class of vectors over which the test achieves some prescribed power.

As a particular case, we consider the functional regression model

$$Y_i = F(x_i) + \varepsilon_i, i = 1 \dots, n,$$

where the x_i 's are deterministic points in $[0, 1]$ and F is an unknown function from $[0, 1]$ into \mathbb{R} . Model (1) includes this model by taking $\mathbf{f} = (F(x_1), \dots, F(x_n))'$. For each $s \in]0, 1]$ and $L > 0$, we define the Hölderian ball $H_s(L)$ by

$$H_s(L) = \{F : [0, 1] \rightarrow \mathbb{R}, \forall (x, y) \in [0, 1]^2, |F(x) - F(y)| \leq L|x - y|^s\}. \quad (2)$$

In the case where $\mathcal{C} = \mathcal{C}_{\geq 0}$ or $\mathcal{C} = \mathcal{C}_{\nearrow}$ and under the *a posteriori* assumption that F belongs to $H_s(L)$ for some $s \in]0, 1]$ and $L > 0$, our test rejects the null hypothesis with probability larger than $1 - \beta$ when $d(\mathbf{f}, \mathcal{C})/\sqrt{n}$ is larger than $CL^{1/1+2s}(\sigma^2/n)^{s/1+2s}$ where C denotes some constant depending on α and β only. This result is free from any assumption on the design points x_i 's.

Given some estimator $\hat{\mathbf{f}}$ of \mathbf{f} , a natural idea to test the hypothesis " $\mathbf{f} \in \mathcal{C}$ " is to consider the Euclidean distance between $\hat{\mathbf{f}}$ and \mathcal{C} , denoted by $d(\hat{\mathbf{f}}, \mathcal{C})$, and to reject the null when this quantity is large. The problem is to find a threshold, say t_α , such that the test is of level α , that is

$$\sup_{\mathbf{f} \in \mathcal{C}} P_{\mathbf{f}} \left(d(\hat{\mathbf{f}}, \mathcal{C}) > t_\alpha \right) \leq \alpha.$$

The difficulty is to control the probability of error when \mathbf{f} runs among the whole set \mathcal{C} . Under the condition

$$\kappa^2 = \sup_{\mathbf{f} \in \mathcal{C}} E \left[d^2(\hat{\mathbf{f}}, \mathbf{f}) \right] < +\infty, \quad (3)$$

the problem is solved by taking $t_\alpha = \kappa/\sqrt{\alpha}$ and using Markov's inequality. Such a test is usually called a *plug-in procedure*. Unfortunately, Condition (3) cannot be satisfied when \mathcal{C} is large, namely when $\mathcal{C} = \mathcal{C}_{\geq 0}$ or $\mathcal{C} = \mathcal{C}_{\nearrow}$, making thus the use of plug-in procedures impossible in these cases.

Our test is not based on a plug-in procedure, but rather on multiple testing. The test statistic we consider has the property to be stochastically bounded under the null independently of \mathbf{f} . It is thus possible to derive uniform separation rates without any prior knowledge of both s and L . Moreover, our test does not depend on any estimation of σ .

Testing that \mathbf{f} belongs to a cone against that the Euclidean distance between \mathbf{f} and the cone is large, had never been treated before. In the Gaussian white noise model, under *a priori* assumptions, both under the null and the alternative, on the regularity of the signal, Juditsky and Nemiroski [10] established minimax rates of testing.

The problem of testing qualitative hypotheses (monotonicity, positivity, convexity) has also been treated by Dümbgen and Spokoiny [5] in the Gaussian white noise model. Their test is asymptotically optimal in the sense that for any Lipschitz class $H_1(L)$, with $L > 0$, the optimal separation rate $L^{1/3}(\log(n)/n)^{1/3}$ (with respect to the supremum norm) is achieved.

In the literature, the problem of testing monotonicity in the functional regression model has been mostly studied. Let us mention the work of Gijbels *et al.* [7], Hall and Heckman [8], Ghosal *et al.* [6] and Baraud *et al.* [3]. A common feature of those papers lies on the fact that the proposed tests are sensitive to local discrepancies to the null. In contrast, the procedure we propose aims at detecting discrepancies to the null with respect to the Euclidean distance.

The paper is organised as follows. Section 2 presents the testing procedure. Section 3 contains our main result on the theoretical performances of the test. Section 4 contains simulations showing the practical implementation of the procedure, its performances and the robustness with respect to non-Gaussian errors. Finally, Section 5 is devoted to the proofs.

2. THE TESTING PROCEDURE

Let us consider the regression model given by (1) and let \mathcal{C} be some set in \mathbb{R}^n satisfying Condition (A) and containing 0. We aim at testing the hypothesis $H_0: \mathbf{f} \in \mathcal{C}$. For that purpose, we introduce a collection $\{S_m, m \in \mathcal{M}\}$ of linear subspaces of \mathbb{R}^n , and for each $m \in \mathcal{M}$, we denote by D_m the dimension of S_m . We assume that the following property holds:

Property (S): For any $m \in \mathcal{M}$, $\Pi_m(\mathcal{C}) \subset \mathcal{C}$, where Π_m denotes the orthogonal projector onto S_m .

Let \mathcal{S}_n be some linear subspace of \mathbb{R}^n of dimension denoted by $\mathcal{D}_n < n$, such that for any $m \in \mathcal{M}$, $S_m \subseteq \mathcal{S}_n$. We denote by $\Pi_{\mathcal{S}_n}$ the orthogonal projector onto \mathcal{S}_n .

Let us define $d_n(\cdot, \cdot) = d(\cdot, \cdot)/\sqrt{n}$ where $d(\cdot, \cdot)$ denotes the Euclidean distance in \mathbb{R}^n and let $d_n(\mathbf{f}, \mathcal{C})$ be the distance between \mathbf{f} and \mathcal{C} defined by

$$d_n(\mathbf{f}, \mathcal{C}) = \inf_{\mathbf{g} \in \mathcal{C}} d_n(\mathbf{f}, \mathbf{g}).$$

For some $\alpha \in]0, 1[$, we consider the test statistic T_α defined by

$$T_\alpha = \sup_{m \in \mathcal{M}} \left(\frac{d_n^2(\Pi_m \mathbf{Y}, \mathcal{C})}{d_n^2(\mathbf{Y}, \mathcal{S}_n)/(n - \mathcal{D}_n)} - q_m(u_\alpha) \right), \quad (4)$$

where for each $u \in]0, 1[$, $q_m(u)$ is defined as the $1 - u$ quantile of the random variable

$$Z_m = \frac{d_n^2(\Pi_m \boldsymbol{\varepsilon}, \mathcal{C})}{d_n^2(\boldsymbol{\varepsilon}, \mathcal{S}_n)/(n - \mathcal{D}_n)},$$

and u_α is defined by

$$u_\alpha = \sup \left\{ u \in]0, 1[, P \left(\sup_{m \in \mathcal{M}} (Z_m - q_m(u)) > 0 \right) \leq \alpha \right\}.$$

We reject the null hypothesis if T_α is positive.

In the particular case where \mathcal{C} is a linear subspace of \mathbb{R}^n the procedure is similar to that described in Baraud *et al.* [2].

3. NON ASYMPTOTIC PERFORMANCES OF THE TEST

3.1. Level and power of the test

In this section, we show that our test is of level α for all n , and we describe a subset $\mathcal{E}_n(\beta)$ of \mathbb{R}^n over which the power of the test is greater than $1 - \beta$. In the sequel we denote by $P_{\mathbf{f}}$ the law of the observation \mathbf{Y} drawn from the regression model (1) and we denote respectively by $\bar{\chi}_D(u)$ and $\bar{\Phi}(u)$ the probability for a chi-square statistic with D degrees of freedom and for a standard normal variable to be larger than u .

Theorem 1. *Let \mathcal{C} be some set in \mathbb{R}^n containing 0 and satisfying Condition (A), \mathcal{S}_n a linear subspace of \mathbb{R}^n and $\{S_m, m \in \mathcal{M}\}$ a collection of linear subspaces of \mathcal{S}_n , which satisfies Property (S).*

For $\alpha \in]0, 1[$, we define the test statistic T_α by (4) and we decide to reject the null hypothesis when T_α is positive. Then we have

$$\forall \mathbf{f} \in \mathcal{C}, \quad P_{\mathbf{f}}(T_\alpha > 0) \leq \alpha.$$

For $\beta \in]0, 1[$, we define

$$v_m^2(\beta) = \left[\bar{\chi}_{D_m}^{-1}(\beta/2) + 2q_m(u_\alpha) \frac{\bar{\chi}_{(n-D_n)}^{-1}(\beta/2)}{n - D_n} \right] \frac{\sigma^2}{n} + 2d_n^2(\mathbf{f}, \mathcal{S}_n) \frac{q_m(u_\alpha)}{n - D_n},$$

and

$$\rho_n^2(\mathbf{f}, \beta) = 3 \inf_{m \in \mathcal{M}} [d_n^2(\mathbf{f}, S_m) + v_m^2(\beta)].$$

If $\mathbf{f} \in \mathbb{R}^n$ satisfies that $d_n^2(\mathbf{f}, \mathcal{C}) \geq \rho_n^2(\mathbf{f}, \beta)$, then

$$P_{\mathbf{f}}(T_\alpha \leq 0) \leq \beta.$$

Comments. We immediately deduce from the theorem that

- the size of the test, $\sup_{\mathbf{f} \in \mathcal{C}} P_{\mathbf{f}}(T_\alpha > 0)$, is exactly α , this supremum being achieved for $\mathbf{f} = 0 \in \mathcal{C}$;
- the test is powerful over the class of vectors,

$$\mathcal{E}_n(\beta) = \{\mathbf{f} \in \mathbb{R}^n, d_n^2(\mathbf{f}, \mathcal{C}) \geq \rho_n^2(\mathbf{f}, \beta)\}.$$

Moreover, let us give some comments on the order of magnitude of $\rho_n^2(\mathbf{f}, \beta)$. We shall see that under suitable conditions on the S_m 's (see the comment in Sect. 5.3), $v_m^2(\beta) + d_n^2(\mathbf{f}, S_m)$ is of order $D_m/n + d_n^2(\mathbf{f}, S_m)$. For the sake of simplicity, let us now assume that the collection of spaces $\{S_m, m \in \mathcal{M}\}$ is totally ordered for the inclusion, so that $d_n^2(\mathbf{f}, S_m)$ decreases with the dimension D_m . Therefore, $\rho_n^2(\mathbf{f}, \beta)$ is of order D_{m^*}/n where m^* realizes among \mathcal{M} the best trade-off between $d_n^2(\mathbf{f}, S_m)$ and $v_m^2(\beta)$. For each \mathbf{f} , the quantity $\rho_n^2(\mathbf{f}, \beta)$ provides an upper bound for the (squared) separation rate of our test. This rate is of the same order as the quadratic risk of the penalised estimator of \mathbf{f} proposed by Baraud [1].

In the particular case where \mathcal{C} is a linear subspace of \mathbb{R}^n , adequate calculations allow to improve this rate of testing, namely the quantity $v_m^2(\beta) + d_n^2(\mathbf{f}, S_m)$ is of order $\sqrt{D_m}/n + d_n^2(\mathbf{f}, S_m)$, we refer to the work of Baraud *et al.* [2]. Then the best compromise between the terms $d_n^2(\mathbf{f}, S_m)$ and $v_m^2(\beta)$ leads to separation rates that are similar to the estimation rates of $\|\mathbf{f}\|_n^2$, see Laurent and Massart [11].

3.2. Uniform rates of testing

In this section we establish uniform rates of testing over Hölderian balls of functions. More precisely, let $s \in]0, 1[$, $L > 0$ and define $\mathcal{H}_s(L)$ as follows:

$$\mathcal{H}_s(L) = \{(F(x_1), \dots, F(x_n))' / F \in H_s(L)\},$$

where $H_s(L)$ is defined by equation (2) and the x_i 's are deterministic points in $[0, 1]$. The aim of this section is to compute some quantity $\rho_{s,L}^2(\beta)$ such that for all $\mathbf{f} \in \mathcal{H}_s(L)$ satisfying $d_n^2(\mathbf{f}, \mathcal{C}) \geq \rho_{s,L}^2(\beta)$, our test rejects the null hypothesis with probability greater than $1 - \beta$. We derive such $\rho_{s,L}^2(\beta)$ by computing an upper bound for the supremum of $\rho_n^2(\mathbf{f}, \beta)$ over $\mathbf{f} \in \mathcal{H}_s(L)$ and by applying Theorem 1.

Since we deal with a collection of vectors \mathbf{f} such that each component f_i is defined as the value of a function F at point x_i , we consider a collection of linear spaces defined from functional spaces. We apply Theorem 1 with the collection $\{S_m, m \in \mathcal{M}\}$ defined as follows. For $n > 4$, let $\mathcal{M} = \{2^l, l \leq l_n\}$, where $l_n = \sup\{l, 2^l \leq n/2 + 2\}$. For each $m \in \mathcal{M}$, S_m is the linear space spanned by the vectors

$$\left\{ \left(\mathbb{1}_{\lfloor \frac{j-1}{m}, \frac{j}{m} \rfloor}(x_1), \dots, \mathbb{1}_{\lfloor \frac{j-1}{m}, \frac{j}{m} \rfloor}(x_n) \right)', j = 1, \dots, m \right\}.$$

We choose $\mathcal{S}_n = S_{2^{l_n}}$ which satisfies $D_n < n$.

Corollary 1. Assume that the set \mathcal{C} satisfies the assumptions of Theorem 1 and that Property (S) holds for the collection of S_m 's defined above. Let $\alpha \in]0, 1[$, $s \in]0, 1[$, $L > 0$ and assume that n satisfies $n \geq 4(2L/\sigma)^{1/s}$. For all $\beta \in]0, 1[$, let

$$\rho_{s,L}^2(\beta) = C(\alpha, \beta) \left[L^{2/1+2s} \left(\frac{\sigma^2}{n} \right)^{2s/1+2s} + \frac{\log \log(n)}{n} \sigma^2 \right],$$

where $C(\alpha, \beta)$ is some constant depending on α and β only. Then for all $\mathbf{f} \in \mathcal{H}_s(L)$ such that $d_n^2(\mathbf{f}, \mathcal{C}) \geq \rho_{s,L}^2(\beta)$, we have

$$P_{\mathbf{f}}(T_{\alpha} > 0) \geq 1 - \beta.$$

Comment. When $\mathcal{C} = \mathcal{C}_{\geq 0}$ and $\mathcal{C} = \mathcal{C}_{\nearrow}$, the S_m 's satisfy Property (S), and Corollary 1 applies.

By taking $\sigma^2 = 1$, fixing the value of L and letting n grow towards infinity, we obtain the asymptotic separation rate $r_n(s, L) = L^{1/(1+2s)} n^{-s/(1+2s)}$. This rate corresponds to the minimax estimation rate over $\mathcal{H}_s(L)$.

In the Gaussian white noise model, when $\mathcal{C} = \mathcal{C}_{\geq 0}$, Juditsky and Nemirovski [10] considered the test of $F \in \mathcal{C}_{\geq 0} \cap \mathcal{H}_s(L, R)$ where

$$\mathcal{H}_s(L, R) = \mathcal{H}_s(L) \cap \left\{ \sup_{x \in [0,1]} |F(x)| \leq R \right\}, \quad \text{with } R \geq 2L$$

against $F \in \mathcal{H}_s(L, R)$. They showed that the minimax separation rate for this problem is of the form $r_n(s, L)(\log(n))^{-\theta}$, for some positive θ depending on s .

Let us now turn to the case $\mathcal{C} = \mathcal{C}_{\nearrow}$.

We first consider the case where $s = 1$. By taking $\sigma^2 = 1$, fixing the value of L and letting n grow towards infinity, we obtain the asymptotic separation rate $r_n(1, L) = L^{1/3} n^{-1/3}$.

In the Gaussian white noise model, for the problem of testing $F \in \mathcal{C}_{\nearrow} \cap \mathcal{H}_1(L, R)$ against $F \in \mathcal{H}_1(L, R)$, Juditsky and Nemirovski [10] showed that the minimax separation rate is bounded from below by $r_n(1, L)(\log(n))^{-\theta}$, for some positive θ .

In the case where $s < 1$, the minimax separation rate and *a fortiori* the adaptive rate of testing are unknown.

4. EXAMPLES AND SIMULATIONS

In this section we describe how to implement the test for testing $\mathbf{f} \in \mathcal{C}_{\nearrow}$ and we carry out a simulation study in order to evaluate the performances of our test both when the errors are Gaussian and when they are not. We first describe how the testing procedure is performed, then we present the simulation experiment and finally the results of the simulation study.

4.1. The testing procedure

Let \mathcal{S}_n be the linear space with dimension $\mathcal{D}_n < n$, spanned by the vectors $\{\mathbf{V}_l, l = 1, \dots, \mathcal{D}_n\}$ defined as follows:

$$\mathbf{V}_l = \sum_{i \in I_l} \mathbf{e}_i \quad \text{where } I_l = \left\{ i \in \{1, \dots, n\}, \frac{l-1}{\mathcal{D}_n} < \frac{i}{n} \leq \frac{l}{\mathcal{D}_n} \right\}$$

and $(\mathbf{e}_1, \dots, \mathbf{e}_n)$ denotes the canonical basis of \mathbb{R}^n .

Next, let $\mathcal{M} = \{2, \dots, M_n\}$ for some $M_n \leq \mathcal{D}_n$. For each $m \in \mathcal{M}$ we choose S_m as the linear space spanned by the vectors $\{\mathbf{W}_{m,j}, j = 1, \dots, m\}$ defined as

$$\mathbf{W}_{m,j} = \sum_{l \in J_{m,j}} \mathbf{V}_l \quad \text{where } J_{m,j} = \left\{ l \in \{1, \dots, \mathcal{D}_n\}, \frac{j-1}{m} < \frac{l}{\mathcal{D}_n} \leq \frac{j}{m} \right\}.$$

TABLE 1. Testing monotonicity: simulated functions F ; the ratio *signal/noise* equals $d_n^2(\mathbf{f}, \mathcal{C})/\sigma^2$.

F	σ^2	Ratio <i>signal/noise</i>
$F_0(x) = 0$	0.01	0
$F_1(x) = -0.15x$	0.01	0.18
$F_2(x) = -0.2 \exp(-50(x - 0.5)^2)$	0.01	0.36
$F_3(x) = 0.1x + F_2(x)$	0.01	0.23
$F_4(x) = 0.1 \cos(6\pi x)$	0.01	0.44
$F_5(x) = 0.1x + F_4(x)$	0.01	0.36
$F_6(x) = 0.5x$	0.01	0
$F_7(x) = 0.5x + F_2(x)$	0.0006	0.47
$F_8(x) = 0.5x + F_4(x)$	0.003	0.47

It is easy to check that for such a collection the Property (S) is fulfilled and that for any $m \in \mathcal{M}$, $S_m \subseteq \mathcal{S}_n$.

To compute the test statistic, we need to compute $d_n(\mathbf{y}, \mathcal{C})$ for some $\mathbf{y} \in \mathbb{R}^n$. This is done by calculating the orthogonal projection, $\Pi_{\mathcal{C}} \mathbf{y}$, of \mathbf{y} onto the convex set \mathcal{C} . For further details on the computation of $\Pi_{\mathcal{C}} \mathbf{y}$ we refer to Brunk [4].

For each m the quantiles $q_m(u)$ are calculated by simulations for u varying on a grid of values u_1, \dots, u_l . Then for each u_j , we compute the quantity

$$p(u_j) = P \left(\sup_{m \in \mathcal{M}} \left\{ \frac{d_n^2(\Pi_m \boldsymbol{\varepsilon}, \mathcal{C})}{d_n^2(\boldsymbol{\varepsilon}, \mathcal{S}_n)} - q_m(u_j) \right\} > 0 \right),$$

by simulations, $\boldsymbol{\varepsilon}$ being a n -sample of $\mathcal{N}(0, 1)$ and we take u_α as

$$u_\alpha = \max \{u_j, p(u_j) \leq \alpha\}.$$

4.2. The simulation experiment

We consider three distributions of the errors ε_i , with expectation zero and variance σ^2 (that were already considered by Horowitz and Spokoiny, 2000).

1. The Gaussian distribution: $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$.
2. The mixture of Gaussian distributions: ε_i is distributed as $\pi X_1 + (1 - \pi) X_2$ where π is distributed as a Bernoulli variable with expectation 0.9, X_1 and X_2 are centred Gaussian variables with variance respectively equal to 0.039σ and 0.625σ . π , X_1 and X_2 are independent. This distribution has heavy tails.
3. The type I distribution: ε_i has density $(s/\sigma) f_X(\mu + (s/\sigma)x)$ where $f_X(x) = \exp\{-x - \exp(-x)\}$, and where μ and s^2 are the expectation and the variance of a variable X with density f_X . This distribution is asymmetrical.

The number of observations n equals 100, and $x_i = i/(n+1)$, for $i = 1, \dots, n$. We choose $\mathcal{D}_n = 50$ and $M_n = 20$. We consider several functions F presented in Table 1, and for each of them, we set $f_i = F(x_i)$ and simulate the observations $Y_i = f_i + \varepsilon_i$. The variance of the errors as well as the ratio *signal/noise* equal to $d_n^2(\mathbf{f}, \mathcal{C})/\sigma^2$ are displayed in Table 1.

In Figure 1 we have displayed the functions F_ℓ for $\ell = 0, \dots, 8$ and for each of them one sample simulated with Gaussian errors and the corresponding value of the test statistic T_α for $\alpha = 5\%$.

4.3. Results of the simulation study

The results of the simulation experiment, based on 1000 simulations, are presented in Table 2. The calculation of u_α is based on $B = 40\,000$ simulations: $B/2$ simulations are used for calculating the values of the $p(u_j)$'s, the u_j 's, $j = 1, \dots, l$ being chosen uniformly between 99.2% and 99.6% for $l = 20$, and $B/2$ simulations are used for calculating u_α . Note that u_α does not depend on $(x_i, i = 1, \dots, n)$, but only on the number of observations n .

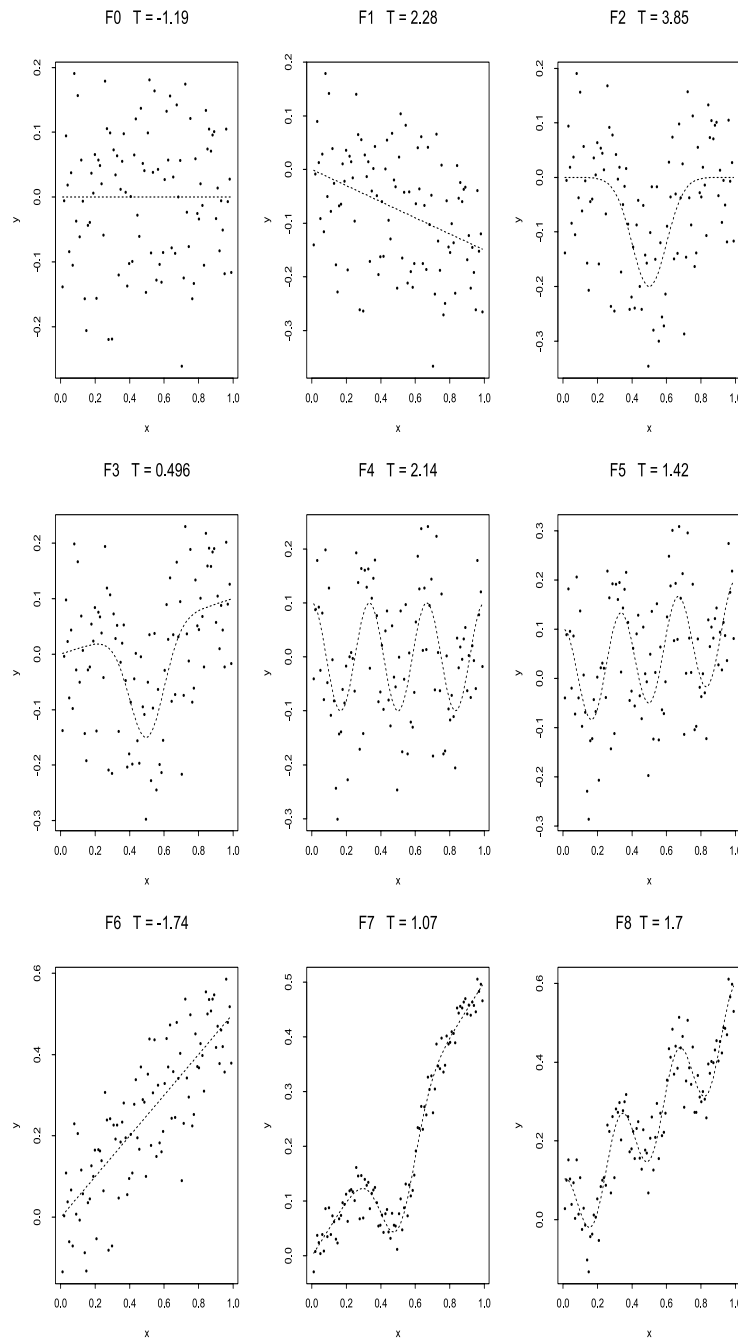


FIGURE 1. For each function $F_\ell, \ell = 0, \dots, 8$, the simulated data $Y_i = F_\ell(x_i) + \varepsilon_i$ for $i = 1, \dots, n$ are displayed. The errors ε_i are Gaussian centred variables with variance σ^2 . The value of the test statistic T_α , with $\alpha = 5\%$, is given for each example. The hypothesis $\mathbf{f} \in \mathcal{C}_\nearrow$ is rejected for all functions F , except for the functions F_0 and F_6 that belong to the null hypothesis.

TABLE 2. Testing monotonicity: percentage of rejection.

F	errors distribution		
	Normal	Mixture	Type I
$F_0(x)$	0.044	0.044	0.058
$F_1(x)$	0.938	0.941	0.944
$F_2(x)$	0.997	0.991	0.993
$F_3(x)$	0.909	0.912	0.914
$F_4(x)$	0.975	0.964	0.965
$F_5(x)$	0.909	0.906	0.929
$F_6(x)$	0	0.001	0
$F_7(x)$	0.966	0.945	0.950
$F_8(x)$	0.955	0.927	0.948

As expected, when the distribution of the errors is Gaussian, the level of the test calculated for the function $F_0(x) = 0$ is (nearly) equal to α .

For the distributions “Mixture” and “Type I”, it has also the desired level, showing the robustness of the procedure with respect to non-Gaussian errors.

Looking at the results corresponding to the functions F_1 to F_5 , it appears that the test is powerful for detecting discrepancies to a constant function or to a function that increases slowly.

The quantiles $q_m(u_\alpha)$ introduced in the test statistic defined in equation (4) are calculated when the vector of expectation \mathbf{f} equals zero. This corresponds to the worse case under the null hypothesis “ \mathbf{f} belongs to \mathcal{C}_\nearrow ”. It follows that if the vector in \mathcal{C}_\nearrow is strongly increasing, the size of the test is zero. This fact is confirmed by the simulation study for the function F_6 . As a consequence, the test is powerful for detecting a discrepancy to a strongly increasing function if the variance of the errors is small. This is exactly what we get in the simulation study for the functions F_7 and F_8 .

The tests proposed by Baraud *et al.* [3] aim at detecting local discrepancies to the null. Since the procedure proposed in this paper aims at detecting a “global” discrepancy, it is worth to provide an example where it performs better. In the case of the function F_8 , the tests proposed by Baraud *et al.* [3] have a power that is not larger than 89%, while the above test achieves a power larger than 95%.

5. PROOF OF THEOREM 1

5.1. Level of the test

In this section we show that the level of the test is smaller than α . We start with the following lemma which compares the repartition functions of a central and a non central χ^2 random variables with the same degrees of freedom.

Lemma 1. *Let T be a χ^2 variable with D degrees of freedom and let T' be a non central χ^2 variable with D degrees of freedom and non centrality parameter $a > 0$. Then for all $u > 0$*

$$P(T' \leq u) \leq P(T \leq u).$$

The proof of the lemma is postponed to the end of the section.

To start with, note that by Property (S) and under H_0 , $\Pi_m \mathbf{f}$ belongs to \mathcal{C} for all $m \in \mathcal{M}$. Now, thanks to Condition (A) for all $\mathbf{g} \in \mathcal{C}$,

$$d_n(\Pi_m \mathbf{Y}, \mathcal{C}) \leq d_n(\Pi_m \mathbf{Y}, \Pi_m \mathbf{f} + \mathbf{g}) = d_n(\Pi_m \boldsymbol{\varepsilon}, \mathbf{g}),$$

which leads to

$$d_n^2(\Pi_m \mathbf{Y}, \mathcal{C}) \leq d_n^2(\Pi_m \boldsymbol{\varepsilon}, \mathcal{C}).$$

It follows that for all $\mathbf{f} \in \mathcal{C}$,

$$\begin{aligned} P_{\mathbf{f}}(T_\alpha > 0) &= P_{\mathbf{f}} \left(\sup_{m \in \mathcal{M}} \left\{ \frac{d_n^2(\Pi_m \mathbf{Y}, \mathcal{C})}{d_n^2(\mathbf{Y}, \mathcal{S}_n)/(n - \mathcal{D}_n)} - q_m(u_\alpha) \right\} > 0 \right) \\ &\leq P_{\mathbf{f}} \left(\sup_{m \in \mathcal{M}} \left\{ \frac{d_n^2(\Pi_m \boldsymbol{\varepsilon}, \mathcal{C})}{d_n^2(\mathbf{Y}, \mathcal{S}_n)/(n - \mathcal{D}_n)} - q_m(u_\alpha) \right\} > 0 \right) \\ &\leq P_{\mathbf{f}} \left(d_n^2(\mathbf{Y}, \mathcal{S}_n) < (n - \mathcal{D}_n) \sup_{m \in \mathcal{M}} \frac{d_n^2(\Pi_m \boldsymbol{\varepsilon}, \mathcal{C})}{q_m(u_\alpha)} \right). \end{aligned}$$

Note that the random variable $nd_n^2(\mathbf{Y}, \mathcal{S}_n)/\sigma^2$ is distributed as a non central χ^2 random variable with $n - \mathcal{D}_n$ degrees of freedom and non centrality parameter $\sqrt{n}d_n(\mathbf{f}, \mathcal{S}_n)/\sigma$. We deduce from Lemma 1 that for all $u > 0$,

$$P(d_n^2(\mathbf{Y}, \mathcal{S}_n) \leq u) \leq P(d_n^2(\boldsymbol{\varepsilon}, \mathcal{S}_n) \leq u).$$

Since for all $m \in \mathcal{M}$, $S_m \subseteq \mathcal{S}_n$, it follows from Cochran's theorem that $d_n^2(\mathbf{Y}, \mathcal{S}_n)$ is independent of the random variable $\sup_{m \in \mathcal{M}} (d_n^2(\Pi_m \boldsymbol{\varepsilon}, \mathcal{C})/q_m(u_\alpha))$. Then by conditioning and applying Lemma 1 we obtain

$$\begin{aligned} P_{\mathbf{f}}(T_\alpha > 0) &\leq P \left(d_n^2(\boldsymbol{\varepsilon}, \mathcal{S}_n) < (n - \mathcal{D}_n) \sup_{m \in \mathcal{M}} \frac{d_n^2(\Pi_m \boldsymbol{\varepsilon}, \mathcal{C})}{q_m(u_\alpha)} \right) \\ &\leq P \left(\sup_{m \in \mathcal{M}} \left(\frac{d_n^2(\Pi_m \boldsymbol{\varepsilon}, \mathcal{C})}{d_n^2(\boldsymbol{\varepsilon}, \mathcal{S}_n)/(n - \mathcal{D}_n)} - q_m(u_\alpha) \right) > 0 \right) \\ &\leq \alpha, \end{aligned}$$

by definition of u_α .

Proof of Lemma 1. This result has already been shown by Anderson in a more general setting [9] (p. 155). In the particular case of χ^2 variables, the proof is very simple and is given below.

Let U_1, \dots, U_D be i.i.d. $\mathcal{N}(0, 1)$ random variables. T is distributed as $U_1^2 + \dots + U_D^2$ and T' as $(U_1 + a)^2 + \dots + U_D^2$. We just have to prove that $\forall u > 0$,

$$P(|U_1 + a| \leq u) \leq P(|U_1| \leq u).$$

Let $h_u(a) = P(|U_1 + a| \leq u) = \Phi(u - a) - \Phi(-a - u)$, where Φ denotes the repartition function of U_1 . Denoting by ϕ the density of U_1 , we have $h'_u(a) = -\phi(u - a) + \phi(-a - u)$, $h'_u(0) = 0$ and $h'_u(a) < 0$ for $a > 0$, which proves that h_u admits a maximum for $a = 0$ and concludes the proof.

5.2. The power of the test

Let us evaluate the risk of second kind for \mathbf{f} such that $d_n(\mathbf{f}, \mathcal{C}) \geq \rho_n(\mathbf{f}, \beta)$.

$$\begin{aligned} P_{\mathbf{f}}(T_\alpha \leq 0) &= P_{\mathbf{f}} \left(\forall m \in \mathcal{M}, \frac{d_n^2(\Pi_m \mathbf{Y}, \mathcal{C})}{d_n^2(\mathbf{Y}, \mathcal{S}_n)/(n - \mathcal{D}_n)} \leq q_m(u_\alpha) \right) \\ &\leq \inf_{m \in \mathcal{M}} P_{\mathbf{f}} \left(\frac{d_n^2(\Pi_m \mathbf{Y}, \mathcal{C})}{d_n^2(\mathbf{Y}, \mathcal{S}_n)/(n - \mathcal{D}_n)} \leq q_m(u_\alpha) \right). \end{aligned}$$

We decompose this probability into two terms, controlling the fluctuations of the random variable $nd_n^2(\mathbf{Y}, \mathcal{S}_n)/(n - \mathcal{D}_n)$ around σ^2 . Let

$$t^2 = \frac{2}{n - \mathcal{D}_n} \left(d_n^2(\mathbf{f}, \mathcal{S}_n) + \bar{\chi}_{n - \mathcal{D}_n}^{-1}(\beta/2) \frac{\sigma^2}{n} \right).$$

Setting,

$$P_1(m) = P_{\mathbf{f}} \left(d_n^2(\Pi_m \mathbf{Y}, \mathcal{C}) \leq t^2 q_m(u_\alpha) \right)$$

and

$$P_2 = P_{\mathbf{f}} \left(d_n^2(\mathbf{Y}, \mathcal{S}_n) \geq (n - \mathcal{D}_n)t^2 \right),$$

we get

$$P_{\mathbf{f}} \left(\frac{d_n^2(\Pi_m \mathbf{Y}, \mathcal{C})}{d_n^2(\mathbf{Y}, \mathcal{S}_n)/(n - \mathcal{D}_n)} \leq q_m(u_\alpha) \right) \leq P_1(m) + P_2.$$

In the sequel we choose m among \mathcal{M} such that

$$\begin{aligned} \rho_n^2(\mathbf{f}, \beta) &= 3 \left(d_n^2(\mathbf{f}, S_m) + v_m^2(\beta) \right) \\ &= 3 \left(d_n^2(\mathbf{f}, S_m) + t^2 q_m(u_\alpha) + \bar{\chi}_{D_m}^{-1}(\beta/2) \frac{\sigma^2}{n} \right) \\ &\geq \left(d_n(\mathbf{f}, S_m) + t\sqrt{q_m(u_\alpha)} + \sqrt{\bar{\chi}_{D_m}^{-1}(\beta/2) \frac{\sigma^2}{n}} \right)^2. \end{aligned} \quad (5)$$

Upper bound for P_1

Since the function $d_n(\cdot, \mathcal{C})$ is 1-Lipschitz, namely

$$\forall \mathbf{h}, \mathbf{h}' \in \mathbb{R}^n, |d_n(\mathbf{h}, \mathcal{C}) - d_n(\mathbf{h}', \mathcal{C})| \leq d_n(\mathbf{h}, \mathbf{h}'),$$

we have that

$$d_n(\Pi_m \mathbf{Y}, \mathcal{C}) \geq d_n(\Pi_m \mathbf{f}, \mathcal{C}) - d_n(\Pi_m \boldsymbol{\varepsilon}, 0)$$

and as $d_n(\mathbf{f}, \mathcal{C}) \geq \rho_n(\mathbf{f}, \beta)$ we get

$$\begin{aligned} d_n(\Pi_m \mathbf{f}, \mathcal{C}) &\geq d_n(\mathbf{f}, \mathcal{C}) - d_n(\mathbf{f}, S_m) \\ &\geq \rho_n(\mathbf{f}, \beta) - d_n(\mathbf{f}, S_m). \end{aligned}$$

This leads to

$$\begin{aligned} P_1(m) &\leq P_{\mathbf{f}} \left(d_n(\Pi_m \boldsymbol{\varepsilon}, 0) \geq d_n(\Pi_m \mathbf{f}, \mathcal{C}) - t\sqrt{q_m(u_\alpha)} \right) \\ &\leq P_{\mathbf{f}} \left(d_n(\Pi_m \boldsymbol{\varepsilon}, 0) \geq \rho_n(\mathbf{f}, \beta) - d_n(\mathbf{f}, S_m) - t\sqrt{q_m(u_\alpha)} \right) \\ &\leq P_{\mathbf{f}} \left(d_n(\Pi_m \boldsymbol{\varepsilon}, 0) \geq \sqrt{\bar{\chi}_{D_m}^{-1}(\beta/2) \frac{\sigma^2}{n}} \right) \end{aligned}$$

thanks to (5), which implies that $P_1(m) \leq \beta/2$.

Upper bound for P_2

Since $d_n^2(\mathbf{Y}, \mathcal{S}_n) \leq 2 \left(d_n^2(\mathbf{f}, \mathcal{S}_n) + d_n^2(\boldsymbol{\varepsilon}, \mathcal{S}_n) \right)$, we have that

$$\begin{aligned} P_2 &\leq P \left(2d_n^2(\boldsymbol{\varepsilon}, \mathcal{S}_n) \geq (n - \mathcal{D}_n)t^2 - 2d_n^2(\mathbf{f}, \mathcal{S}_n) \right) \\ &= \beta/2. \end{aligned}$$

This concludes the proof of Theorem 1.

5.3. Proof of Corollary 1

We prove the corollary showing that for all $\mathbf{f} \in \mathcal{H}_s(L)$

$$\rho_n^2(\mathbf{f}, \beta) \leq C(\alpha, \beta) \left[L^{2/(1+2s)} \left(\frac{\sigma^2}{n} \right)^{2s/(1+2s)} + \frac{\log \log(n)}{n} \sigma^2 \right]. \quad (6)$$

Then the result follows from Theorem 1.

To do so, we use the following lemma that is proved at the end of this section.

Lemma 2. *Under the assumptions of Corollary 1 the following holds: for each $m \in \mathcal{M}$, there exists some constant C depending on α and β only, such that*

$$v_m^2(\beta) \leq C \frac{D_m + \log \log(n)}{n} (\sigma^2 + d_n^2(\mathbf{f}, \mathcal{S}_n)). \quad (7)$$

For all $s \in]0, 1]$, $L > 0$, $\mathbf{f} \in \mathcal{H}_s(L)$ and $m \in \mathcal{M}$

$$d_n(\mathbf{f}, \mathcal{S}_m) \leq LD_m^{-s}. \quad (8)$$

Since for all $m \in \mathcal{M}$, $\mathcal{S}_m \subset \mathcal{S}_n$, the following inequalities hold

$$\frac{D_m + \log \log(n)}{n} d_n^2(\mathbf{f}, \mathcal{S}_n) \leq 2d_n^2(\mathbf{f}, \mathcal{S}_n) \leq 2d_n^2(\mathbf{f}, \mathcal{S}_m) \quad (9)$$

and we derive from (7) and (8) that

$$\rho_n^2(\mathbf{f}, \beta) = 3 \inf_{m \in \mathcal{M}} (d_n^2(\mathbf{f}, \mathcal{S}_m) + v_m^2(\beta)) \leq C' \inf_{m \in \mathcal{M}} \left(\frac{D_m + \log \log(n)}{n} \sigma^2 + L^2 D_m^{-2s} \right)$$

for some constant C' depending on α and β only.

Then, we choose m among \mathcal{M} to satisfy

$$1 + [(L^2 n / \sigma^2)^{1/(1+2s)}] \leq D_m = 2^l < 2 + 2[(L^2 n / \sigma^2)^{1/(1+2s)}]$$

which is possible under the assumption $n \geq 4(2L/\sigma)^{1/s}$ (then $2 + 2[(L^2 n / \sigma^2)^{1/(1+2s)}] \leq 2 + n/2$). Finally, we obtain (6) by noting that

$$(L^2 n / \sigma^2)^{1/(1+2s)} \leq D_m \leq 2(1 + (L^2 n / \sigma^2)^{1/(1+2s)}).$$

Comment. Note that if $D_m \geq \log \log(n)$ then we deduce from (7) and (9) that $v_m^2(\beta) + d_n^2(\mathbf{f}, \mathcal{S}_m)$ is of order $D_m/n + d_n^2(\mathbf{f}, \mathcal{S}_m)$.

Let us now prove Lemma 2.

Proof of (7). We start with proving the following inequality

$$q_m(u_\alpha) \leq D_m \bar{F}_{D_m, n - \mathcal{D}_n}^{-1}(\alpha / |\mathcal{M}|).$$

On the one hand, since 0 belongs to \mathcal{C} , we have that

$$Z_m \leq \frac{(n - \mathcal{D}_n) d_n^2(\Pi_m \boldsymbol{\varepsilon}, 0)}{d_n^2(\boldsymbol{\varepsilon}, \mathcal{S}_n)} = X_m^2.$$

The random variable X_m^2/D_m being distributed as a Fisher random variable with D_m and $n - D_n$ degrees of freedom we get that for all $u > 0$,

$$q_m(u) \leq D_m \bar{F}_{D_m, n-D_n}^{-1}(u).$$

On the other hand, since

$$\begin{aligned} P\left(\sup_{m \in \mathcal{M}} \{Z_m - q_m(\alpha/|\mathcal{M}|)\} > 0\right) &\leq \sum_{m \in \mathcal{M}} P(Z_m > q_m(\alpha/|\mathcal{M}|)) \\ &\leq \alpha, \end{aligned}$$

we have that $\alpha/|\mathcal{M}| \leq u_\alpha$ and thus

$$q_m(u_\alpha) \leq D_m \bar{F}_{D_m, n-D_n}^{-1}(u_\alpha) \leq D_m \bar{F}_{D_m, n-D_n}^{-1}(\alpha/|\mathcal{M}|).$$

In the sequel we use the following inequality proved in Baraud *et al.* [2]: for all $u \in]0, 1[$

$$D \bar{F}_{D, N}^{-1}(u) \leq D + 2\sqrt{D \left(1 + \frac{D}{N}\right) \log\left(\frac{1}{u}\right)} + \left(1 + 2\frac{D}{N}\right) \frac{N}{2} \left[\exp\left(\frac{4}{N} \log\left(\frac{1}{u}\right)\right) - 1\right].$$

Taking $D = D_m$, $N = n - D_m$ which is of order n , and $u = u_\alpha \geq \alpha/|\mathcal{M}|$ which is of order $\alpha/\ln(n)$, we get

$$q_m(u_\alpha) \leq D_m \bar{F}_{D_m, n-D_n}^{-1}(u_\alpha) \leq C(\alpha) (D_m + \log \log(n)).$$

We now repeatedly use the following upper bound on the quantiles of a \mathcal{X}^2 random variable which can be found in Laurent and Massart [11]: for all $u \in]0, 1[$,

$$\bar{\chi}_D^{-1}(u) \leq D + 2\sqrt{-D \log(u)} - 2 \log(u).$$

We derive that

$$\bar{\chi}_{D_m}^{-1}(\beta/2) \leq C(\beta) D_m \quad \text{and} \quad \frac{\bar{\chi}_{n-D_n}^{-1}(\beta/2)}{n - D_n} \leq C(\beta),$$

which leads to the result.

Proof of (8). For each $j \in \{1, \dots, m\}$, let us set $I_{j,m} = \{i/x_i \in [(j-1)/m, j/m]\}$, $\ell(j) = \inf I_{j,m}$ when $I_{j,m} \neq \emptyset$ and let \mathbf{f}_m be the \mathbb{R}^n vector defined by: for each $i \in \{1, \dots, n\}$, $(\mathbf{f}_m)_i = f_{\ell(j)} = F(x_{\ell(j)})$ if $i \in I_{j,m}$. Clearly \mathbf{f}_m belongs to S_m and we have for all $\mathbf{f} \in \mathcal{H}_s(L)$,

$$\begin{aligned} d_n^2(\mathbf{f}, S_m) &\leq \frac{1}{n} \sum_{j=1}^m \sum_{i \in I_{j,m}} |F(x_i) - F(x_{\ell(j)})|^2 \\ &\leq L^2 m^{-2s} = L^2 D_m^{-2s}, \end{aligned}$$

since $D_m = m$.

REFERENCES

[1] Y. Baraud, Model selection for regression on a fixed design. *Probab. Theory Related Fields* **117** (2000) 467-493.
 [2] Y. Baraud, S. Huet and B. Laurent, Adaptive tests of linear hypotheses by model selection. *Ann. Statist.* **31** (2003).
 [3] Y. Baraud, S. Huet and B. Laurent, *Tests for convex hypotheses*, Technical Report 2001-66. University of Paris XI, France (2001).
 [4] H.D. Brunk, On the estimation of parameters restricted by inequalities. *Ann. Math. Statist.* **29** (1958) 437-454.
 [5] L. Dümbgen and V.G. Spokoiny, Multiscale testing of qualitative hypotheses. *Ann. Statist.* **29** (2001) 124-152.

- [6] S. Ghosal, A. Sen and A. van der Vaart, Testing monotonicity of regression. *Ann. Statist.* **28** (2000) 1054-1082.
- [7] I. Gijbels, P. Hall, M.C. Jones and I. Koch, Tests for monotonicity of a regression mean with guaranteed level. *Biometrika* **87** (2000) 663-673.
- [8] P. Hall and N. Heckman, Testing for monotonicity of a regression mean by calibrating for linear functions. *Ann. Statist.* **28** (2000) 20-39.
- [9] I.A. Ibragimov and R.Z. Has'minskii, *Statistical estimation. Asymptotic theory*. Springer-Verlag, New York-Berlin, *Appl. Math.* **16** (1981).
- [10] A. Juditsky and A. Nemirovski, On nonparametric tests of positivity/monotonicity/convexity. *Ann. Statist.* **30** (2002) 498-527.
- [11] B. Laurent and P. Massart, Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.* **28** (2000) 1302-1338.