Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 136 (2006) 4326–4348
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Testing in mixed-effects FANOVA models

Felix Abramovich[a,*], Claudia Angelini[b,1]

[a]*Department of Statistics & Operations Research, Tel Aviv University, Israel*
[b]*Istituto per le Applicazioni del Calcolo "Mauro Picone"—Sezione di Napoli, Consiglio Nazionale delle Ricerche, Italy*

Received 14 October 2003; accepted 3 June 2005
Available online 1 July 2005

Abstract

We consider the testing problem in the mixed-effects functional analysis of variance models. We develop asymptotically optimal (minimax) testing procedures for testing the significance of functional global trend and the functional fixed effects based on the empirical wavelet coefficients of the data. Wavelet decompositions allow one to characterize various types of assumed smoothness conditions on the response function under the nonparametric alternatives. The distribution of the functional random-effects component is defined in the wavelet domain and captures the sparseness of wavelet representation for a wide variety of functions. The simulation study presented in the paper demonstrates the finite sample properties of the proposed testing procedures. We also applied them to the real data from the physiological experiments.

MSC: Primary 62G10; Secondary 62G20

Keywords: Besov spaces; Functional analysis of variance; Functional hypothesis testing; Mixed effects; Wavelets

1. Introduction

Modern recording equipment enables researchers to gather a large number of observations on individuals over time that can be modelled essentially as continuous curves (functions). Such situation is typical today, for example, in the analysis of seismic, meteorological,

* Corresponding author.
E-mail address: felix@post.tau.ac.il (F. Abramovich).
[1] Supported by "CNR Fellowship 2002".

0378-3758/$ - see front matter © 2005 Published by Elsevier B.V.
doi:10.1016/j.jspi.2005.06.002

medical or financial data. A series of high-resolution images is an example of functional data in the two-dimensional domain. Last years much progress have been made in developing statistical techniques to deal with functional data (see Ramsay and Silverman, 1997, 2002 for review).

One of the new statistical challenges arising in functional data analysis is comparison between curves or sets of curves. For example, analysing electro-encephalogram (EEG) data of men and women in treatment and control groups one is naturally interested in the effects of treatment and gender on the shape of EEG measurements. Such problems are considered within *functional analysis of variance* (FANOVA) framework. Much efforts have been applied to extending the traditional ANOVA methods for FANOVA. There is a wide list of publications on *fitting* various FANOVA models and *estimating* their components (e.g., Wahba et al., 1995; Stone et al., 1997; Huang, 1998; Lin, 2000; Gu, 2002; Angelini et al., 2003). However, much less attention has been paid to the functional *inference* or *hypothesis testing*.

A somewhat naive approach to testing in FANOVA models by performing a series of standard *univariate* ANOVA tests to compare a set of curves at each specific time causes a serious multiplicity problem due to an enormous number of simultaneous tests. Ignoring multiplicity leads to an uncontrolled overall Type I error while, for example, Bonferroni-type procedures are known to yield an extremely low power. Another approach to FANOVA testing treats functional data as multivariate vectors and applies traditional multivariate ANOVA techniques combined sometimes with various initial dimensionality-reduction procedures (e.g., Raz, 1990; Eubank and La Riccia, 1993; Chen, 1994). However, the "curse of dimensionality" makes these attempts also problematic (see Faraway, 1997). Fan and Lin (1998) proposed a powerful overall test for functional hypothesis testing based on the adaptive Neyman and wavelet thresholding procedures of Fan (1996) applied to the empirical Fourier and wavelet coefficients of the data, respectively. It is well known that a large variety of different functions have a *sparse* representation in the Fourier and especially wavelet domain that allows significant reduction in dimensionality of the original functional data. Somewhat related approaches were considered in Eubank (2000) and Dette and Derbort (2001). However, the above works did not investigate the optimality of the proposed procedures. Abramovich et al. (2004) applied the asymptotically minimax functional hypothesis testing framework originated by Ingster (1982) for testing in the *fixed*-effects FANOVA. In particular, they adapted the corresponding wavelet-based testing procedures of Spokoiny (1996) for testing a zero signal in a "signal + white noise" model and showed their asymptotic optimality (in the minimax sense) for testing in the fixed-effects FANOVA models for the wide class of alternatives.

In various applications the data on individuals is usually grouped according to some factors where one is interested in the differences between groups rather than between particular individuals. Individuals are then treated as a *random* effect associated with a sample randomly drawn from a population. This also allows to model correlations between observations over the same individual which is typical, for example, for longitudinal and repeated measurements data. Recently, Guo (2002) proposed a maximum likelihood ratio based test in the mixed-effects smoothing spline FANOVA models. Spitzner et al. (2003) applied the procedures of Fan and Lin (1998) mentioned above to the mixed-effects FANOVA. In this paper, we extend the results of Abramovich et al. (2004) for the optimal testing in the mixed-effects FANOVA model and derive the corresponding rate-optimal tests.

The paper is organized as follows. Section 2 presents the mixed-effects FANOVA model and the hypotheses to be tested. In the main Section 3, we give first some necessary background on functional hypotheses testing and then derive the corresponding asymptotically minimax non-adaptive and adaptive tests. In Section 4, we discuss the applications of the tests for finite samples and illustrate their performance on a small simulation study and on a real-life data from physiology. The concluding remarks and some possible extensions are given in Section 5. The proof of the main Theorem 1 is given in the appendix.

2. Testing in mixed-effects FANOVA models

2.1. Mixed-effects FANOVA model and hypotheses to be tested

Consider the following functional mixed-effects FANOVA model

$$\begin{aligned} &\mathrm{d}Y_{i,l}(t) = m_i(t)\,\mathrm{d}t + V_l(t)\,\mathrm{d}t + \varepsilon\,\mathrm{d}W_{i,l}(t),\\ &\quad i = 1, \ldots, r; \quad l = 1, \ldots, m; \qquad t \in [0, 1], \end{aligned} \tag{1}$$

where $m_i(t)$ are fixed effect functions, $V_l(t)$ are random effect functions modelled as independent realizations of a zero mean stochastic process $V(t)$ and $W_{i,l}(t)$ are independent realizations of a standard Wiener process. In addition, $V_l(t)$ and $W_{i,l}(t)$ are mutually independent.

Following Antoniadis (1984), each $m_i(t)$, $i = 1, \ldots, r$ in (1) admits the following unique decomposition

$$m_i(t) = m_0 + \mu(t) + a_i + \gamma_i(t), \quad i = 1, \ldots, r; \quad t \in [0, 1], \tag{2}$$

where m_0 is a constant (the *overall mean*), $\mu(t)$ is either zero or a non-constant function of t (the *main fixed effect* of t), a_i is either zero or a non-constant function of i (the *main effect* of i) and $\gamma_i(t)$ is either zero or a non-constant function which cannot be decomposed as a sum of a function of i and a function of t (the *fixed interaction* component). The components of the decomposition (2) satisfy the following identifiability conditions:

$$\begin{aligned} &\int_0^1 \mu(t)\,\mathrm{d}t = 0; \quad \sum_{i=1}^r a_i = 0; \quad \sum_{i=1}^r \gamma_i(t) \equiv 0; \quad \int_0^1 \gamma_i(t)\,\mathrm{d}t = 0,\\ &\quad \forall i = 1, \ldots, r; \quad t \in [0, 1]. \end{aligned} \tag{3}$$

As in traditional mixed-effects ANOVA models, one is naturally interested in testing the significance of the fixed-effects components of (1)–(3). In this paper, we study the corresponding asymptotically (as $\varepsilon \to 0$) optimal (in the minimax sense) functional hypotheses testing procedures for mixed-effects FANOVA models.

Testing the significance of the main effects and the interactions is equivalent to testing the following hypotheses:

$$\mathrm{H}_0 : \mu(t) \equiv 0, \quad t \in [0, 1] \quad \text{(no global trend)}, \tag{4}$$

$$\mathrm{H}_0 : a_i = 0 \quad \forall i = 1, \ldots, r \quad \text{(no differences in level)}, \tag{5}$$

$$\mathrm{H}_0 : \gamma_i(t) \equiv 0, \quad \forall i = 1, \ldots, r, \quad t \in [0, 1] \quad \text{(no differences in shape)}. \tag{6}$$

Integrating (1) with respect to t and using the identifiability conditions (3), we have

$$Y^*_{i,l} = m_0 + a_i + \tilde{V}_l + \varepsilon\xi_{i,l}, \quad l = 1, \ldots, m, \quad i = 1, \ldots, r, \quad \sum_{i=1}^{r} a_i = 0,$$

where $Y^*_{i,l} = \int_0^1 \mathrm{d}Y_{i,l}(t)$, $\tilde{V}_l = \int_0^1 V_l(t)\,\mathrm{d}t$ and $\xi_{i,l}$ are independent $\mathrm{N}(0, 1)$ random variables. This is the classical mixed-effects ANOVA model and testing (5) can be performed by standard techniques.

Consider now testing the *functional* hypotheses (4) and (6). Averaging (1)–(2) with respect to i and l and exploiting again the identifiability conditions (3) yield the following *random*-effects FANOVA model:

$$\mathrm{d}\bar{Y}(t) = (m_0 + \mu(t))\,\mathrm{d}t + \bar{V}(t)\,\mathrm{d}t + \varepsilon\,\mathrm{d}\bar{W}(t), \tag{7}$$

where $\bar{V}(t)$ is the average process of $V_1(t), \ldots, V_m(t)$ and $\bar{W}(t)$ is the average of $r \times m$ independent standard Wiener processes. Averaging the data is naturally justified by sufficiency arguments. Let $\bar{Y}_{i\cdot}(t) = (1/m)\sum_{l=1}^{m} Y_{i,l}(t)$ and $\bar{W}_{i\cdot}(t) = (1/m)\sum_{l=1}^{m} W_{i,l}(t)$, respectively. Then, from (7) we have

$$\mathrm{d}(\bar{Y}_{i\cdot}(t) - \bar{Y}(t)) = (a_i + \gamma_i(t))\,\mathrm{d}t + \varepsilon\mathrm{d}(\bar{W}_{i\cdot}(t) - \bar{W}(t)). \tag{8}$$

The last Eq. (8) does not involve random effects components and to test (6) one can use the corresponding procedures developed in Abramovich et al. (2004) for the *fixed*-effects FANOVA models. As in the traditional mixed-effects ANOVA, the random effect component in (1) essentially affects only testing the presence of the global trend (4) which will be the main focus of the paper.

2.2. The alternative set

Rewrite the random-effects FANOVA model (7) in the equivalent form:

$$\mathrm{d}\bar{Y}(t) = (m_0 + \mu(t))\,\mathrm{d}t + \bar{V}(t)\,\mathrm{d}t + \eta\,\mathrm{d}W(t), \quad t \in [0, 1], \tag{9}$$

where $\eta = \varepsilon/\sqrt{rm}$ and $W(t)$ is the standard Wiener process.

We want to test the null hypothesis (4) against a class of alternatives as large as possible and, hence, do not specify any parametric structure for the alternative set. Instead we only assume that $\mu(\cdot)$ possesses some smoothness properties. In particular, we assume that $\mu(\cdot)$ belongs to some Besov ball $B^s_{p,q}(M)$ of radius $M > 0$ on the unit interval, where $1 \leqslant p, q \leqslant \infty$, $sp > 1$. Roughly speaking, the (not necessarily integer) parameter s indicates the number of function's derivatives, where their existence is required in an L_p-sense, while the additional parameter q provides a further finer gradation. Besov classes have exceptional expressive power: for particular choices of the parameters s, p and q, they include, for example, the Hölder ($p = q = \infty$) and Sobolev ($p = q = 2$) classes of smooth functions, and the class of functions of bounded variation sandwiched between $B^1_{1,1}$ and $B^1_{1,\infty}$. We refer to Meyer (1992) for rigorous definitions and a detailed study of Besov spaces.

On the other hand, to be able to distinguish between the two hypotheses, $\mu(\cdot)$ should be also separated away from zero in the L_2-norm, $\|\mu\|_2 \geqslant \rho(\eta)$. This is a typical form of restrictions on the alternative set in the nonparametric testing (see Ingster and Suslina, 2003 for a comprehensive review). The smoothness assumptions bound the set of alternatives while the L_2-norm constraint cuts out the alternatives "too close" to the null.

Hence, given the data in (9) we wish to test

$$\mathrm{H}_0 : \mu(t) \equiv 0 \quad \text{versus} \quad \mathrm{H}_1 : \mu \in \mathscr{F}(\rho(\eta)), \tag{10}$$

where $\mathscr{F}(\rho(\eta)) = \{\mu : \mu \in B^s_{p,q}(M), \int \mu(t)\,\mathrm{d}t = 0, \|\mu\|_2 \geqslant \rho(\eta)\}$.

2.3. Model for the random effects

To complete (9) we need to specify the distribution of the stochastic process $\bar{V}(t)$ in (9) which is completely defined by the distribution of $V(t)$ in the original model (1). Instead of defining the distribution of $V(t)$ directly, we set the distribution on the coefficients of its wavelet expansion.

For simplicity of exposition we consider the orthonormal periodic wavelet bases in $L_2[0, 1]$ (see Daubechies, 1992, Section 9.3 for details) although in practice they might behave poorly near the boundaries for non-periodic functions. Choose a mother wavelet ψ of regularity $v > s$ and perform the periodic wavelet transform on (9):

$$\bar{Y}_{jk} = \mu_{jk} + \bar{V}_{jk} + \eta\xi_{jk}, \quad j \geqslant -1; \;\; k = 0, \ldots, 2^j - 1, \tag{11}$$

where $\bar{Y}_{jk} = \int_0^1 \psi_{jk}(t)\,\mathrm{d}\bar{Y}(t)$, $\mu_{jk} = \int_0^1 \mu(t)\psi_{jk}(t)\,\mathrm{d}t$, $\bar{V}_{jk} = \int_0^1 \bar{V}(t)\psi_{jk}(t)\,\mathrm{d}t$ and ξ_{jk} are independent N(0, 1) random variables. To simplify the notations we also denoted the corresponding scaling function $\phi(t)$ by $\psi_{-10}(t)$.

On the other hand, the process $\bar{V}(t)$ is an average of m independent realizations of $V(t)$ and in the wavelet domain $\bar{V}_{jk} = (1/m)\sum_{l=1}^m V_{jk,l}$, where $V_{jk,l} = \int_0^1 V_l(t)\psi_{jk}(t)\,\mathrm{d}t$, $l = 1, \ldots, m$.

It is natural to assume that unlike completely irregular white noise, the realizations of $V(t)$ posses some smoothness properties—for example, that they fall almost surely within some Besov ball (not necessarily the same as $\mu(t)$). As we have already mentioned, various functions from Besov spaces have a sparse representation in wavelet series and to capture this characteristic feature of wavelets, assume the following distribution on $V_{jk,l}$:

$$V_{jk,l} \sim \pi_j \mathrm{N}(0, \tau_j^2) + (1 - \pi_j)\delta(0) \quad j \geqslant 0; \;\; k = 0, \ldots, 2^j - 1 \tag{12}$$

and independent, where $0 \leqslant \pi_j \leqslant 1$, $\delta(0)$ is a point mass at zero. To complete the model place vague distributions on the scaling coefficients $V_{-10,l}$, $l = 1, \ldots, m$. In addition, assume that $V_{jk,l}$ and ξ_{jk} are independent.

According to (12), each $V_{jk,l}$ is either zero with probability $1 - \pi_j$ or with probability π_j is normally distributed with zero mean and variance τ_j^2. The probability π_j gives the proportion of non-zero wavelet coefficients at resolution level j while the variance τ_j^2 is a measure of their magnitudes. The parameters π_j and τ_j^2 are the same for all coefficients at a given resolution level j. Such type of distributions for the wavelet coefficients of a stochastic process has been proposed, for example, in Abramovich et al. (1998) and Clyde et al. (1998).

Let $\kappa_j^2 = r\tau_j^2/\varepsilon^2 = \tau_j^2/(m\eta^2)$ and also assume that $\limsup_j \kappa_j^2 \leqslant C < \infty$ to assure that the variances of both random components in (1) are of the same order.

The following proposition gives an insight on the corresponding distribution of the random effect component $V_l(t) = \sum_{j\geqslant 0}\sum_{k=0}^{2^j-1} V_{jk,l}\psi_{jk}(t)$ in the time domain:

Proposition 1. *Let the coefficients of wavelet expansions of* $V_l(t), l=1,\dots,m$ *have distribution* (12). *Then*, $V_l(t), l=1,\dots,m$ *are realizations of a* (*non-Gaussian*) *non-stationary zero mean stochastic process* $V(t)$ *with the covariance function*:

$$R(s,t) = \sum_{j\geqslant 0} \pi_j \tau_j^2 \sum_{k=0}^{2^j-1} \psi_{jk}(s)\psi_{jk}(t). \tag{13}$$

The proof follows directly from (12). In fact, (13) shows that wavelet series $\psi_{jk}(t)$ are the eigenfunctions of the covariance function $R(s,t)$ with the corresponding eigenvalues $\pi_j^{1/2}\tau_j$.

In particular, let τ_j^2 and π_j decrease exponentially, that is

$$\tau_j^2 = c_1 2^{-aj} \quad \text{and} \quad \pi_j = \min(1, c_2 2^{-bj}), \quad j\geqslant 0, \tag{14}$$

where $a, b \geqslant 0$ and $c_1, c_2 > 0$. The expected number of non-zero wavelet coefficients on the jth level then is $c_2 2^{j(1-b)}$. Applying the first Borel–Cantelli lemma, for $b>1$, the number of non-zero coefficients in the wavelet expansion is finite almost surely and, hence, with probability one, f will necessarily belong to the same Besov space as the mother wavelet ψ. A more interesting case is $0\leqslant b\leqslant 1$. For $b=1$ the expected number of non-zero wavelet coefficients is the same on each level which is typical for piecewise polynomial functions. The case $b=0$ assumes the same probability of being non-zero for all coefficients on all levels that characterizes self-similar processes such as white noise or Brownian motion. Abramovich et al. (1998, Theorem 1) showed that for $0\leqslant b\leqslant 1$ realizations $V_l(t)$ from (14) will fall (with probability one) within a Besov space $B^{\tilde{s}}_{\tilde{p},\tilde{q}}$ if and only if either

$$\tilde{s} + 1/2 - b/\tilde{p} - a/2 < 0$$

or

$$\tilde{s} + 1/2 - b/\tilde{p} - a/2 = 0 \quad \text{and} \quad 0\leqslant b<1, \quad 1\leqslant \tilde{p} < \infty, \quad \tilde{q} = \infty.$$

Fig. 1 shows an example of 15 realizations of length $n=1024$ of the random processes $V_l(t)$ for the specific choice $a=1, b=0.8, c_1=3000, c_2=25$.

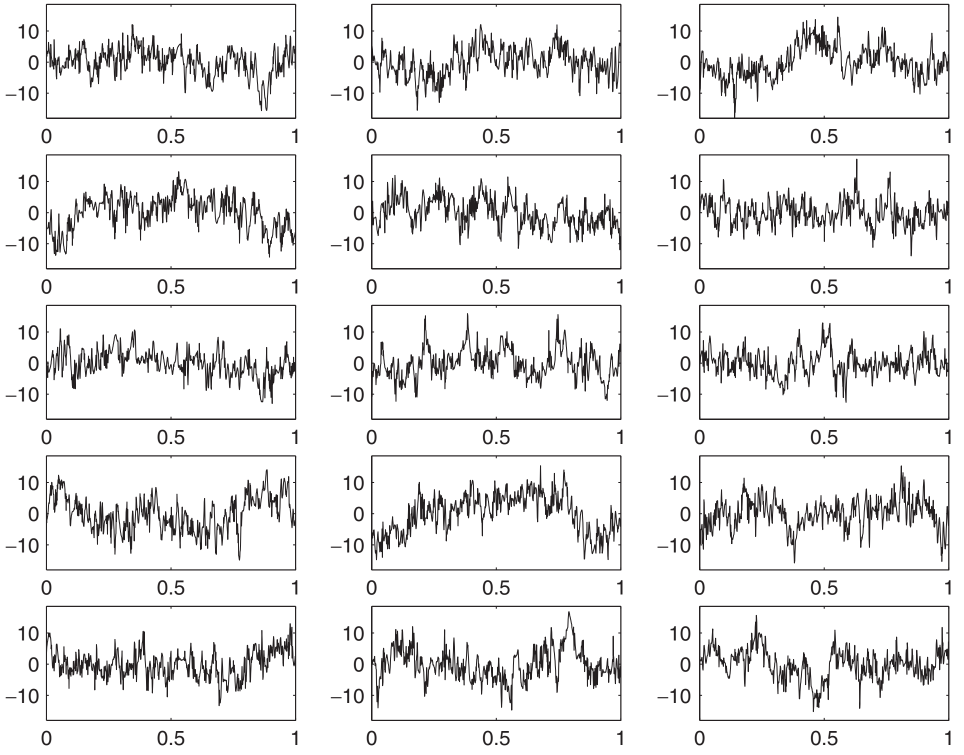

Fig. 1. An example of the $m = 15$ realizations of the random-effects $v_l(t)$ in (30). The realizations are generated sampling from (12) with $\tau_j^2 = c_1\sigma^2 2^{-aj}$ and $\pi_j = \min(1, c_2 2^{-bj})$ where $a = 1$, $b = 0.8$, $c_1 = 3000$, $c_2 = 25$ and σ^2 was taken in order to achieve a RSNR equal to 1.

3. Main results

3.1. Basic background in functional hypotheses testing

We remind first some basic definitions and results for the functional hypotheses testing. A (non-randomized) test ϕ is a measurable function of the data with the two values 0 and 1 that correspond to accepting and rejecting the null hypothesis, respectively. As usual, the quality of the test ϕ is measured by a Type I error (erroneous rejection of H_0) and a Type II error (erroneous acceptance of H_0). The probability of a Type I error is defined as

$$\alpha(\phi) = P_{\mu\equiv 0}(\phi = 1),$$

while the probability of a Type II error for the composite nonparametric alternative hypothesis H_1 is defined as

$$\beta(\phi, \rho(\eta)) = \sup_{\mu \in \mathscr{F}(\rho(\eta))} P_\mu(\phi = 0).$$

For the prescribed error probabilities of both types, the rate of decay of $\rho(\eta)$ as $\eta \to 0$ is a standard measure of asymptotical goodness of a test (e.g., Ingster, 1982, 1993; Ingster and Suslina, 2003). It is natural then to find the fastest rate for which such testing is still possible and to construct the rate-optimal test.

Definition 1. A sequence $\rho(\eta)$ is called the minimax rate of testing if $\rho(\eta) \to 0$ as $\eta \to 0$ and the following two conditions hold:

(i) for any $\rho'(\eta) = \mathrm{o}_\eta(\rho(\eta))$, one has

$$\inf_{\phi_\eta}[\alpha(\phi_\eta) + \beta(\rho'(\phi_\eta, \eta))] = 1 - \mathrm{o}_\eta(1),$$

where $\mathrm{o}_\eta(1)$ is a sequence tending to zero as $\eta \to 0$

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

$$\alpha(\phi_\eta^*) \leqslant \alpha + \mathrm{o}_\eta(1),$$

$$\beta(\phi_\eta^*, c\rho(\eta)) \leqslant \beta + \mathrm{o}_\eta(1).$$

The first condition states that testing with a rate faster than $\rho(\eta)$ is impossible while the second one guarantees that for the rate $\rho(\eta)$ there exists a rate-optimal test ϕ_η^*.

The random-effects FANOVA model (9) differs from the standard "signal + white noise" model by the presence of the additional random effect component. Ingster (1993) and Lepski and Spokoiny (1999) showed that the asymptotically minimax rate for of testing for the latter is

$$\rho(\eta) = \eta^{4s''/(4s''+1)}, \tag{15}$$

where $s'' = \min(s, s - \frac{1}{2p} + \frac{1}{4})$. Spokoiny (1996) derived the corresponding rate-optimal test based on the empirical wavelet coefficients of the data. Sparseness of wavelet bases over Besov spaces results in significant reduction of the dimensionality of the testing problem in the wavelet domain. We extend now the results of Spokoiny (1996) for the random-effects model (9).

3.2. Minimax test

Let $\mu \in B^s_{p,q}(M)$, $1 \leqslant p, q \leqslant \infty$, $sp > 1$ and $s > 1/4$ if $p \geqslant 2$. First assume that all the parameters $\theta = (s, p, q, M)$ of the Besov ball $B^s_{p,q}(M)$ are known.

From (12) a straightforward calculus yields

$$\bar{V}_{jk} \sim \sum_{l=0}^{m} \binom{m}{l} \pi_j^l (1 - \pi_j)^{m-l} \mathrm{N}\left(0, \frac{l\tau_j^2}{m^2}\right) \tag{16}$$

and, therefore,

$$\bar{Y}_{jk} \sim \sum_{l=0}^{m} \binom{m}{l} \pi_j^l (1 - \pi_j)^{m-l} \mathrm{N}\left(\mu_{jk}, \eta^2\left(1 + \frac{l}{m}\kappa_j^2\right)\right). \tag{17}$$

In the wavelet domain the null hypothesis $\mathrm{H}_0 : \mu(t) \equiv 0$ is equivalent to $\mathrm{H}_0 : \mu_{jk} = 0;\ j \geqslant 0, k = 0, \ldots, 2^j - 1$, or $\mathrm{H}_0 : \sum_{j=0}^{\infty}\sum_{k=0}^{2^j-1}\mu_{jk}^2 = 0$, where we drop the scaling coefficient μ_{-10} since due to the identifiability condition $\int \mu(t)\,\mathrm{d}t = 0$ (see (3)) it is zero under both hypotheses. The proposed test will be similar in spirit to those of Spokoiny (1996) and Abramovich et al. (2004) but will involve necessary changes due to the presence of the additional random effect component.

Let

$$J_\eta = \log_2(\eta^{-2}) \quad \text{and} \quad J_\theta = \frac{2}{4s'' + 1}\log_2(M\eta^{-2}). \tag{18}$$

Without loss of generality we can assume that J_η and J_θ are integers, otherwise, one can take the corresponding integer parts. Note that since $sp > 1$ and $s > 1/4$ for $p \geqslant 2$, $J_\theta < J_\eta$ for sufficiently small η.

Let ζ_j be independent random variables distributed as $\zeta_j \sim \sum_{l=0}^{m}\binom{m}{l}\pi_j^l(1-\pi_j)^{m-l}$ $\mathrm{N}(0, w_{lj}^2)$, where $w_{lj}^2 = 1 + (l/m)\kappa_j^2$. For each $j = 0, \ldots, J_\theta - 1$ define S_j to be

$$S_j = \sum_{k=0}^{2^j-1}(\bar{Y}_{jk}^2 - \eta^2(1 + \pi_j\kappa_j^2)) \tag{19}$$

while for each $j = J_\theta, \ldots, J_\eta - 1$ and for a given threshold λ define $S_j(\lambda)$ to be

$$S_j(\lambda) = \sum_{k=0}^{2^j-1}(\bar{Y}_{jk}^2\mathbf{1}\{|\bar{Y}_{jk}| > \eta\lambda\} - \eta^2 b_j(\lambda)), \tag{20}$$

where $\mathbf{1}(A)$ is the indicator function of the set A,

$$\begin{aligned} b_j(\lambda) &= \mathbb{E}[\zeta_j^2\mathbf{1}\{|\zeta_j| > \lambda\}] \\ &= 2\sum_{l=0}^{m}\binom{m}{l}\pi_j^l(1-\pi_j)^{m-l}w_{lj}^2\left(\Phi\left(-\frac{\lambda}{w_{lj}}\right) + \frac{\lambda}{w_{lj}}\phi\left(\frac{\lambda}{w_{lj}}\right)\right) \end{aligned}$$

and Φ and ϕ are the probability and density functions, respectively, of the standard Gaussian distribution.

With the above notation, introduce the following test statistics:

$$T(J_\theta) = \sum_{j=0}^{J_\theta-1} S_j \quad \text{and} \quad Q(J_\theta) = \sum_{j=J_\theta}^{J_\eta-1} S_j(\lambda_j), \tag{21}$$

where

$$\lambda_j = 4\sqrt{(1+\kappa_j^2)(j - J_\theta + 8)\ln 2}. \tag{22}$$

Let $v_0^2(J_\theta)$ and $\omega_0^2(J_\theta)$ be the variances of $T(J_\theta)$ and $Q(J_\theta)$, respectively, under the null hypothesis. A straightforward calculus implies

$$v_0^2(J_\theta) = 2\eta^4 \sum_{j=0}^{J_\theta - 1} 2^j \left(1 + \pi_j \kappa_j^4 \left(\frac{3}{2}\frac{1-\pi_j}{m} + \pi_j\right) + 2\pi_j \kappa_j^2\right) \quad \text{and}$$

$$\omega_0^2(J_\theta) = \eta^4 \sum_{j=J_\theta}^{J_\eta - 1} 2^j d_j(\lambda_j), \tag{23}$$

where $d_j(\lambda) = \mathbb{E}[\zeta_j^4 \mathbf{1}\{|\zeta_j| > \lambda\}] - b_j^2(\lambda)$ and

$$\mathbb{E}[\zeta_j^4 \mathbf{1}\{|\zeta_j| > \lambda\}] = \sum_{l=0}^{m} \binom{m}{l} \pi_j^l (1-\pi_j)^{m-l} w_{lj}^4 \times \left(6\Phi\left(-\frac{\lambda}{w_{lj}}\right) + 2\frac{\lambda}{w_{lj}}\left(3 + \frac{\lambda^2}{w_{lj}^2}\right)\phi\left(\frac{\lambda}{w_{lj}}\right)\right).$$

Finally, for any given significance level $\alpha \in (0, 1)$, let ϕ^* be the test defined by

$$\phi^* = \mathbf{1}\left\{\frac{T(J_\theta) + Q(J_\theta)}{\sqrt{v_0^2(J_\theta) + \omega_0^2(J_\theta)}} > z_{1-\alpha}\right\}, \tag{24}$$

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

The resulting test statistic is intuitively clear and is essentially the standardized sum of squares of the thresholded $\bar{Y}_{jk}$ with properly chosen level-dependent thresholds larger than those of Spokoiny (1996) for the "signal + white noise" model due to the additional random effects. The coefficients on the coarse levels $j < J_\theta$ are not thresholded. The resulting coefficients are then centered to imply $ES_j = 0$ and $ES_j(\lambda) = 0$ under H_0. The null hypothesis is rejected when the above sum of squares is "too large".

To establish the asymptotic optimality of the proposed test (24), note first that the minimax rate of testing (15) for the standard "signal + white noise" model is an obvious lower bound for the more general model (9) that involves an additional random effects component. The following theorem, whose proof is given in the appendix, shows that under assumption (12), test (24) achieves this lower bound and, hence, is a level α asymptotically (as $\eta \to 0$) rate-optimal test:

Theorem 1. *Let the mother wavelet $\psi(t)$ be of regularity $v > s$, and let the parameters $\theta = (s, p, q, M)$ of the ball $B_{p,q}^s(M)$ be known, where $1 \leqslant p, q \leqslant \infty$, $sp > 1$ and $s > 1/4$ for $p \geqslant 2$. Consider testing*

$$H_0 : \mu \equiv 0 \quad \text{versus}$$

$$H_1 : \mu \in \mathscr{F}(\rho(\eta)) = \{\mu : \mu \in B_{p,q}^s(M), \int \mu(t)\,dt = 0, \|\mu\|_2 \geqslant \rho(\eta)\}$$

in the mixed-effects model (9) *and* (12). *Then, for a fixed significance level* $\alpha \in (0, 1)$, *as* $\eta \to 0$, *the rate* $\rho(\eta)$ *of the test* ϕ^* *defined in* (24) *is*

$$\rho(\eta) = \eta^{4s''/(4s''+1)}. \tag{25}$$

Remark 1. For $p \geqslant 2$ corresponding to "spatially homogeneous" functions whose wavelet coefficients are concentrated on coarse resolution levels, the above optimal test can be simplified by truncating the wavelet series at level $J_\theta - 1$ (see also Abramovich et al., 2004). The resulting rate-optimal test ϕ^* becomes then

$$\phi^* = \mathbf{1}\left\{\frac{T(J_\theta)}{v_0(J_\theta)} > z_{1-\alpha}\right\}.$$

The proof of Remark 1 follows straightforwardly from the proof of Theorem 1 in the appendix.

3.3. Adaptive test

The rate-optimal test derived in the previous section relies on the knowledge of the parameters of the Besov ball $\theta = (s, p, q, M)$. In practice, however, they are typically unknown. In this section, we consider the *adaptive* local testing problem where the above parameters are not specified a priori but are assumed to lie within a given range.

Assume now that $\theta=(s, p, q, M)$ is unknown but $1/4 < s \leqslant s_{\max}$, $1 \leqslant p \leqslant p_{\max}$, $1 \leqslant q < \infty$, $sp > 1$ and $M_{\min} \leqslant M \leqslant M_{\max}$. Denote such a range of θ by $\mathscr{T}$. For each given set of parameters θ one may determine J_θ from (18). In fact, the range $\mathscr{T}$ derives essentially a range of admissible levels of the form $j_{\min} \leqslant J_\theta \leqslant j_{\max}$. The underlying idea of the adaptive test is analogous to that of Spokoiny (1996) and Abramovich et al. (2004) for the fixed-effects FANOVA: one performs a series of tests of type (24) for each admissible J_θ and rejects the null hypothesis if it is rejected at least for one of them.

More precisely, let $j_{\min} = 2\log_2 \eta^{-2}/(4s''_{\max} + 1)$, $j_{\max} = J_\eta - 1$, where $s''_{\max} = s_{\max} - 1/(2p_{\max}) + 1/4$ and $s' = s + \frac{1}{2} + \frac{1}{p'}$. Choose a mother wavelet of regularity $v > s_{\max}$. Since $\mathrm{card}(\mathscr{T}) = \mathcal{O}(\ln \eta^{-2})$, a Bonferroni type correction for multiple testing leads to the following asymptotic adaptive test:

$$\phi^{\mathrm{a}} = \mathbf{1}\left[\max_{j_{\min} \leqslant J_\theta \leqslant j_{\max}} \left\{\frac{T(J_\theta) + Q(J_\theta)}{\sqrt{v_0^2(J_\theta) + w_0^2(J_\theta)}}\right\} > \sqrt{2 \ln \ln \eta^{-2}}\right]. \tag{26}$$

The rate of ϕ^{a} is given in the following theorem:

Theorem 2. *As* $\eta \to 0$, *the rate* $\rho(\eta)$ *of the test* ϕ^{a} *defined in* (26) *for testing* (10) *is*

$$\rho(\eta) = \eta^{4s''/(4s''+1)} (\ln \ln \eta^{-2})^{s'/(4s''+1)}. \tag{27}$$

Moreover, there exists a constant c such that

$$\alpha(\phi^{\mathrm{a}}) = \mathrm{o}_\eta(1) \quad \text{and} \quad \sup_{\mathscr{T}} \beta(\phi^{\mathrm{a}}, c\rho(\eta)) = \mathrm{o}_\eta(1).$$

The proof of Theorem 2 essentially repeats the corresponding arguments in Spokoiny (1996) and is omitted.

Theorem 2 establishes that the adaptive test (26) is *nearly* rate-optimal (up to an additional ($\ln \ln \eta^{-2}$) factor). The results of Spokoiny (1996) show that there is no adaptive testing without loss of efficiency of the test and such an extra log–log factor is unavoidable (though not essential) price for adaptivity. In addition, the above theorem demonstrates the degenerate behavior of the error probabilities of ϕ^{a} which is also typical for adaptive testing (e.g., Ingster and Suslina, 2003).

4. Finite sample applications

In practice one always observes *discrete* data samples of size n at points t_h, $h=1,\ldots,n$ with the noise variance σ^2. The corresponding sampled versions of the proposed testing procedures are based then on the empirical wavelet coefficients obtained by the discrete periodic wavelet transforms with $\varepsilon_n=\sigma/\sqrt{n}$.

4.1. Estimation of parameters

To apply the tests developed in the previous section one needs to specify the random effect parameters π_j and τ_j^2(or, equivalently, κ_j^2) in (12). Ideally, π_j and τ_j^2 could be obtained from some prior information or assumptions about, for example, the regularity of realizations of $V_l(t)$ (see Section 2.3). In practice, however, it is usually difficult to elicit such a prior information about the regularity properties. Instead, we suggest to estimate π_j and τ_j^2 from the data somewhat similar in spirit to the empirical Bayes approach of estimating the hyperparameters of the prior on wavelet coefficients within the Bayesian framework (e.g., Clyde and George, 2000; Johnstone and Silverman, 2005).

Let the continuous wavelet coefficients $Y_{jk,l}=\int_0^1\psi_{jk}\,\mathrm{d}\bar{Y}_{\cdot l}$, where $\bar{Y}_{\cdot l}(t)=(1/r)\sum_{i=1}^r Y_{i,l}(t)$, and $\tilde{Y}_{jk,l}$ be their discrete counterparts for $\bar{Y}_{\cdot l}(t_h)$, $h=1,\ldots,n$. Then, $\tilde{Y}_{jk,l}=\sqrt{n}(Y_{jk,l}+\mathrm{O}(1/n))$, where the $\sqrt{n}$ factor arises due to the difference between the continuous and discrete orthonormality conditions. From the corresponding sampled versions of (1) and (12), $\tilde{Y}_{jk,l}$ are independent and

$$\tilde{Y}_{jk,l}\sim\pi_j\mathrm{N}\left(\tilde{\mu}_{jk},\frac{\sigma^2}{r}(1+\kappa_j^2)\right)+(1-\pi_j)\mathrm{N}\left(\tilde{\mu}_{jk},\frac{\sigma^2}{r}\right),\tag{28}$$

where $\tilde{\mu}_{jk}=\sqrt{n}(\mu_{jk,l}+\mathrm{O}(1/n))$ are the discrete wavelet coefficients of $\mu(t_h)$, $h=1,\ldots,n$.

The maximum likelihood estimators of π_j and κ_j cannot be obtained explicitly and numerical procedures should be adopted. However, given σ^2, it is possible to get their estimates in the closed form by the method of moments. Due to the symmetry of the distribution in (28), all its odd central moments are zeroes. A simple calculus yields

$$\mathrm{Var}[\tilde{Y}_{jk,l}]=\frac{\sigma^2}{r}(1+\pi_j\kappa_j^2)\quad\text{and}$$

$$\mathbb{E}[\tilde{Y}_{jk,l}-\tilde{\mu}_{jk}]^4=3\frac{\sigma^4}{r^2}(\pi_j(1+\kappa_j^2)^2+(1-\pi_j)).$$

Let $m_{jk}^{(u)} = \sum_{l=1}^{m}(\tilde{Y}_{jk,l} - \bar{\tilde{Y}}_{jk})^u/m$, $u = 2, 4$, be the corresponding sampled central moments of order u. Since π_j and κ_j are the same for all k, $k = 0, \ldots, 2^j - 1$ on the jth resolution level, by solving the corresponding equations, the method of moments provides 2^j independent estimates of π_j and κ_j for each j in the following closed form: $\hat{\pi}_{j(k)} = 3(m_{jk}^{(2)} - \sigma^2/r)^2/(m_{jk}^{(4)} - 6m_{jk}^{(2)}\sigma^2/r + 3\sigma^4/r^2)$ and $\hat{\kappa}_{j(k)}^2 = (m_{jk}^{(2)} - \sigma^2/r)/\hat{\pi}_{j(k)}$. To guarantee $0 \leqslant \hat{\pi}_{j(k)} \leqslant 1$ and $\hat{\kappa}_{j(k)}^2 \geqslant 0$ the above estimates should be corrected as

$$\hat{\pi}_{j(k)} = \min\left(\frac{3(m_{jk}^{(2)} - \sigma^2/r)^2}{(m_{jk}^{(4)} - 6m_{jk}^{(2)}\sigma^2/r + 3\sigma^4/r^2)_+}, 1\right) \quad \text{and}$$
$$\hat{\kappa}_{j(k)}^2 = \frac{(m_{jk}^{(2)} - \sigma^2/r)_+}{\hat{\pi}_{j(k)}}, \tag{29}$$

where $(x)_+ = \max(0, x)$. The estimates $\hat{\pi}_j$ and $\hat{\kappa}_j^2$ of the π_j and κ_j^2 are then obtained by averaging the resulting $\hat{\pi}_{j(k)}$ and $\hat{\kappa}_{j(k)}$ over $k = 0, \ldots, 2^j - 1$.

In most applications, the noise variance σ^2 is also unknown. In wavelet function estimation, the common practice is to robustly estimate σ by the median of the absolute deviation of the empirical wavelet coefficients of the data at the finest resolution level $J_n = \log_2(n) - 1$ divided by 0.6745 (Donoho and Johnstone, 1994). We estimate σ by averaging the m robust estimates obtained from each individual set of empirical wavelet coefficients and then substitute its estimate in (29).

4.2. Numerical examples

In this section, we investigate first the finite sample performance of the proposed functional hypotheses testing procedure on simulated data and then we apply it to a real-life data example arising from physiology. The simulation were performed using the MATLAB programming environment and the WaveLab toolbox (Ver. 8.0).

4.2.1. Simulation study

To investigate the finite sample properties of the proposed procedure for testing the significance of the global trend $\mu(t)$, we performed a simulation study based on the synthetic data generated according to the random-effects model

$$y_l(t_h) = \mu(t_h) + v_l(t_h) + \sigma z_{l,h}, \quad l = 1, \ldots, m;$$
$$t_h = (h-1)/n, \quad h = 1, \ldots, n. \tag{30}$$

The random-effects model is a particular case of a mixed-effects model and may be relevant in its own right, for example, in situations where one has measurements on the same object obtained from different sources or locations like it often happens for meteorological or seismic data.

We chose $\mu(t)=\sqrt{2}\sin(2\pi t)$ to satisfy the identifiability conditions (3) and to have $\|\mu\|_2=1$. Series of m random effects samples $v_l(t)$ $(l=1,\dots,m)$ were obtained by performing a discrete inverse wavelet transform of m sets of wavelet coefficients randomly sampled from (12). The parameters in (12) were of the form $\tau_j^2=c_1\sigma^2 2^{-aj}$ and $\pi_j=\min(1, c_2 2^{-bj})$ and several values for (c_1, c_2, a, b) were tried. In all simulations the compactly supported *Coiflet 18-tap filter* mother wavelet was used.

The adaptive test (26) was based on the asymptotic properties of the maxima of card$(\mathscr{T})$ weakly dependent Gaussian random variables where card$(\mathscr{T})=j_{\max}-j_{\min}+1=\mathcal{O}(\log(n))$ is sufficiently large. Obviously it holds only for very large samples and in practice to perform the adaptive test at the significance level α one can approximate the $(1-\alpha)100\%$-th percentile of the distribution of the maxima by the corresponding percentile $z_{(1-\alpha)^{1/\mathrm{card}(\mathscr{T})}}$ of the maxima of independent Gaussian variables:

$$\tilde{\phi}^{\mathrm{a}}=\mathbf{1}\left[\max_{j_{\min}\leqslant J_\theta\leqslant j_{\max}}\left\{\frac{T(J_\theta)+Q(J_\theta)}{\sqrt{v_0^2(J_\theta)+w_0^2(J_\theta)}}\right\}>z_{(1-\alpha)^{1/\mathrm{card}(\mathscr{T})}}\right]. \tag{31}$$

In what follows we present the simulation results for $n=1024$, $m=15$, $a=1$, $b=0.8$, $c_1=3000$, $c_2=25$. The value of σ was chosen to yield the ratio of the standard deviations of the signal and the noise (RSNR) to be one. Fig. 1 shows an example of the 15 realizations of the random-effect components $v_l(t)$ and Fig. 2 gives the corresponding realizations of the resulting process $y_l(t)$ in (30). The function $\mu(t)$ is superimposed to the noisy observations as reference. The above choice of the parameters implies a strong presence of random-effects component. Averaging (30) over $l=1,\dots,15$ reduces the variance by 15 but the average random effect is still quite strong (see Fig. 3). In order to make our procedure fully automatic and suited for applications with real data, τ_j^2, π_j and σ^2 were assumed unknown during the simulations and were estimated from the data by the methods described in Section 4.1.

Since $\mu(t)$ is smooth we have tried the discrete version of (24) with $J_\theta=4$ and $\eta^2=\sigma^2/15$ to test the null hypothesis $\mathrm{H}_0:\mu(t)\equiv 0$ at the significance level $\alpha=0.05$. The test statistic $(T(4)+Q(4))/\sqrt{v_0^2(4)+\omega_0^2(4)}$ was 3.505 (p-value $=2.28*10^{-4}$) while $z_{0.95}=1.645$ and the null hypothesis was therefore rejected.

We have also applied the adaptive test (31) within the range $j_{\min}=3$ and $j_{\max}=J_n-1=8$. The corresponding test statistic was 3.542 to be compared with $z_{0.95^{1/6}}=2.386$, and the null hypothesis was again rejected.

We have performed a power analysis of both non-adaptive (for $J_\theta=4$) and adaptive tests for $\mu(t)=\sqrt{2}\sin(2\pi t)$. Fig. 4 shows the empirical power functions based on 3000 replications for each RSNR as a function of RSNR where we used the same parameters in (12) as before. The values of σ were chosen according to the corresponding values of the RSNR. To investigate the effect of estimating τ_j^2 and π_j on the power we performed an analogous study with the true values of τ_j^2 and π_j (see also Fig. 4). In both cases for a fixed RSNR the power of the non-adaptive test is larger than that of the adaptive although the differences vanish as RSNR increases. For strong noise (small RSNR) the automatic procedures tend to yield somewhat larger values of the test statistic. The differences in the

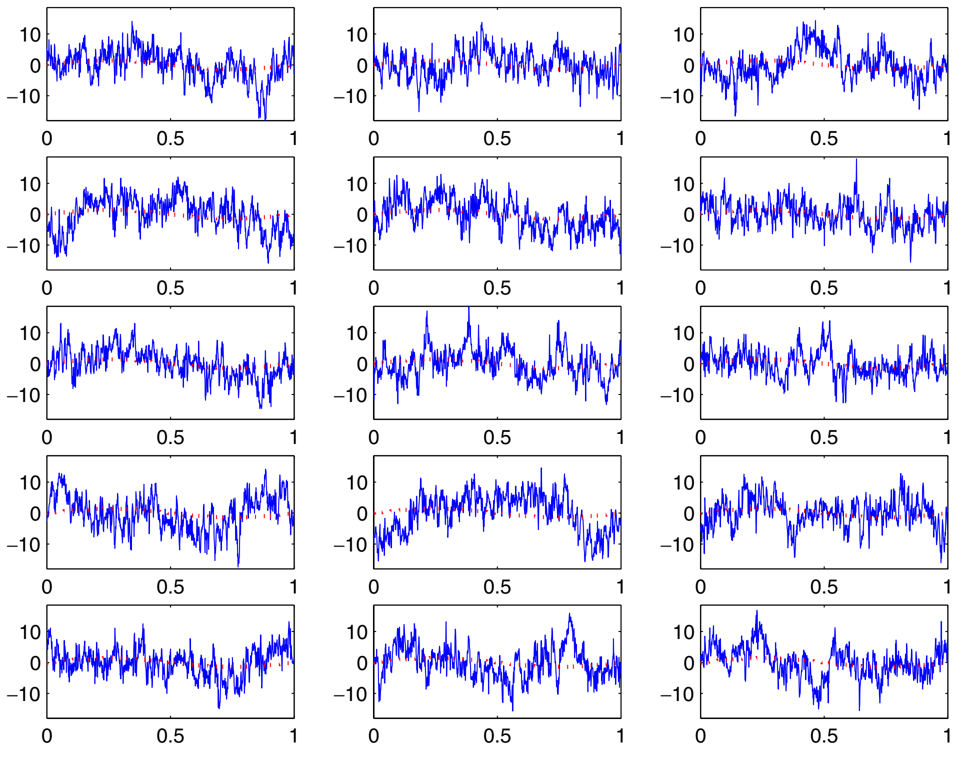

Fig. 2. $m = 15$ realizations of the observed processes $y_l(t)$ in (30) corresponding to the realizations of the random-effects components given in Fig. 1. The regression function $\mu(t)$ is superimposed in dotted line.

empirical power curves for true and estimated τ_j^2 and π_j disappear as RSNR increases due to the improving accuracy of estimation on low levels especially. Fig. 4 shows that the power of both non-adaptive and adaptive tests increases fast with RSNR and for RSNR $= 1.5$ (or, equivalently, $1.5\sqrt{15} = 5.8$ for the averaged signal $\bar{y}(t)$) is already about 0.95.

4.2.2. *Real-life data example*

We have applied the proposed mixed-effects FANOVA methodology to some interesting data on human movement. The data were acquired and computed by Dr. Amarantini and Dr. Martin (Laboratoire Sport et Performance Motrice, Grenoble University) to study the processes underlying movement generation under various levels of an externally applied moment to the knee. In this experiment, stepping-in-place was a relevant task to investigate how muscle redundancy could be appropriately used to cope with an external perturbation while complying with the mechanical requirements related either to balance control and/or minimum energy expenditure. For this purpose, 7 young male volunteers wore a spring-loaded orthosis of adjustable stiffness under 4 experimental conditions: a control condition (without orthosis), an orthosis condition (with the orthosis only), and two condi-

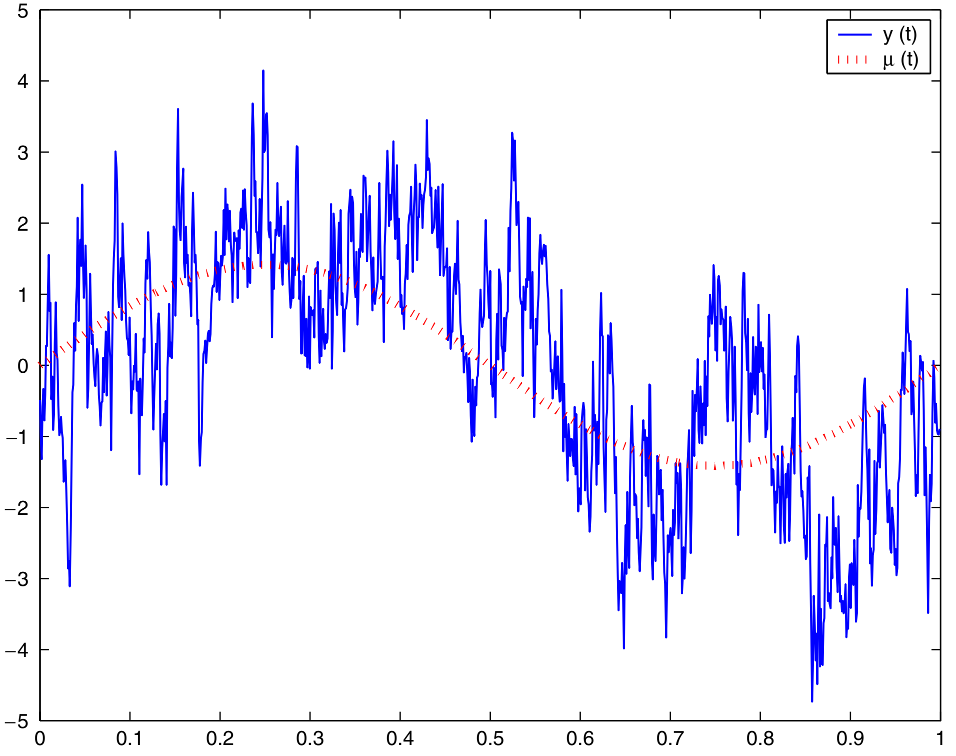

Fig. 3. Average model $\bar{y}(t)$ (continuous line) obtained from the 15 realizations in Fig. 2. Superimposed the regression function $\mu(t)$ (dotted line).

tions (spring1, spring2) in which stepping in place was perturbed by fitting a spring-loaded orthosis onto the right knee joint. The experimental session included 10 trials of 20 s under each experimental condition for each subject. Data sampling started 5 s after the onset of stepping, and lasted for 10 s for each trial. Anticipatory and joint movements induced by the initiation of the movement were, therefore, removed. For each of the 7 subjects, 10 stepping-cycles of data were analyzed under each experimental condition. The resultant moment at the knee is derived by means of body segment kinematics recorded with a sampling frequency of 200 Hz. We refer to Cahouët et al. (2002) for further details on how the data were recorded and how the resultant moment was computed.

For each stepping-in-place replication, the resultant moment was computed at 256 time points equally spaced and scaled so that a time interval corresponds to an individual gait cycle. A typical moment observation is therefore a one-dimensional function of normalized time t so that $t \in [0, 1]$. The data set consists in 280 separate runs and involves the 7 subjects over 4 described experimental conditions, replicated 10 times for each subject (see Fig. 5). Note that variability across subjects is much stronger than across treatments. Abramovich et al. (2004) analyzed this data using a two-way fixed-effects FANOVA model considering

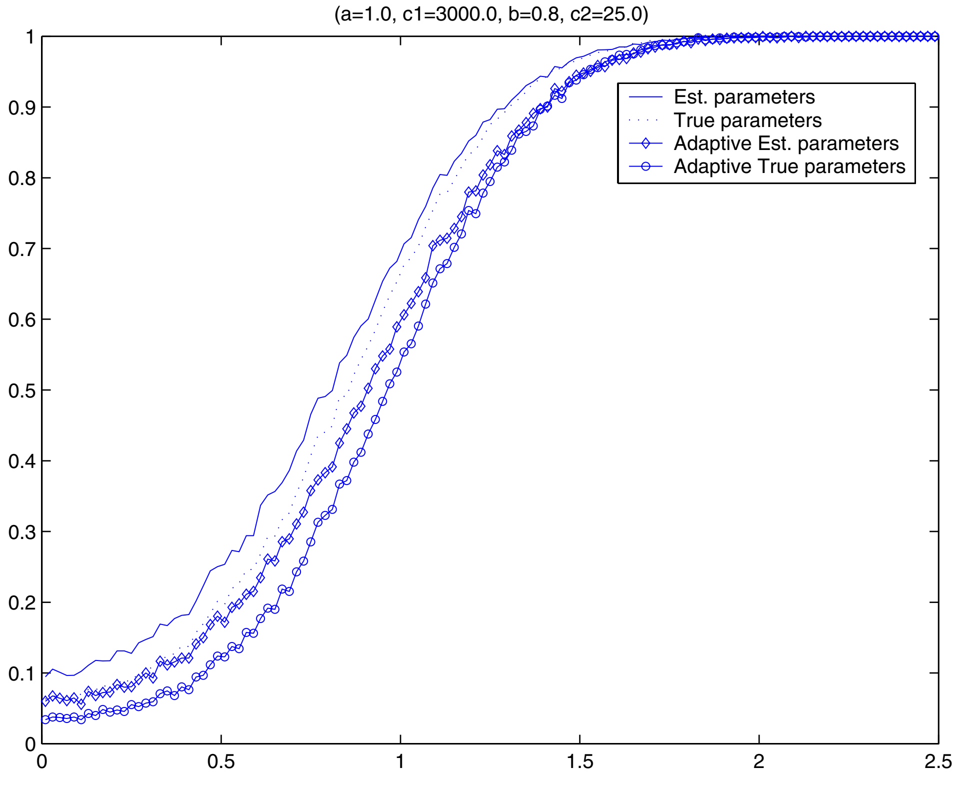

Fig. 4. Empirical power functions for testing $H_0\colon \mu(t) \equiv 0$ versus $H_1\colon \mu(t) = \sqrt{2}\sin(2\pi t)$ obtained from $m = 15$ synthetic signals generated according to model (30) for the non-adaptive automatic procedure (solid line) and for the non-adaptive procedure with true values of $\tau_j^2 = c_1\sigma^2 2^{-aj}$ and $\pi_j = \min(1, c_2 2^{-bj})$ where $a = 1, b = 0.8, c_1 = 3000, c_2 = 25$ (dotted line); for the adaptive full automatic procedure (circle line) and for the adaptive procedure with the same true value of τ^2 and π_j (diamond line). The sample size was $n = 1024$, the number of replication of each experiment was 3000.

both treatments and subjects as fixed effects. They found significant global differences between treatments although under the control and orthosis conditions the subjects behave similarly, the same being true under spring1 and spring2 conditions. They also found a highly significant global trend over time. However, it is probably more reasonable to treat subjects as random effects and to apply the corresponding mixed-effects FANOVA model. As we have discussed in Section 2.1, testing the differences between treatments will not be affected (see (8)) and the above results of Abramovich et al. (2004) remain valid. The differences will be in testing the presence of a global trend $H_0 : \mu(t) = 0$. Fig. 6 shows the averaged observed process $\bar{y}(t)$. It is now of natural interest to investigate whether the fluctuations observed in $\bar{y}(t)$ can be explained by the random variability between subjects and noise only or there is also a global trend over time. We applied the adaptive test (31) with $j_{\min} = 3$ and $j_{\max} = J_n - 1 = 6$. The resulting test statistic was 4.992 while $z_{0.95^{1/4}} = 2.234$

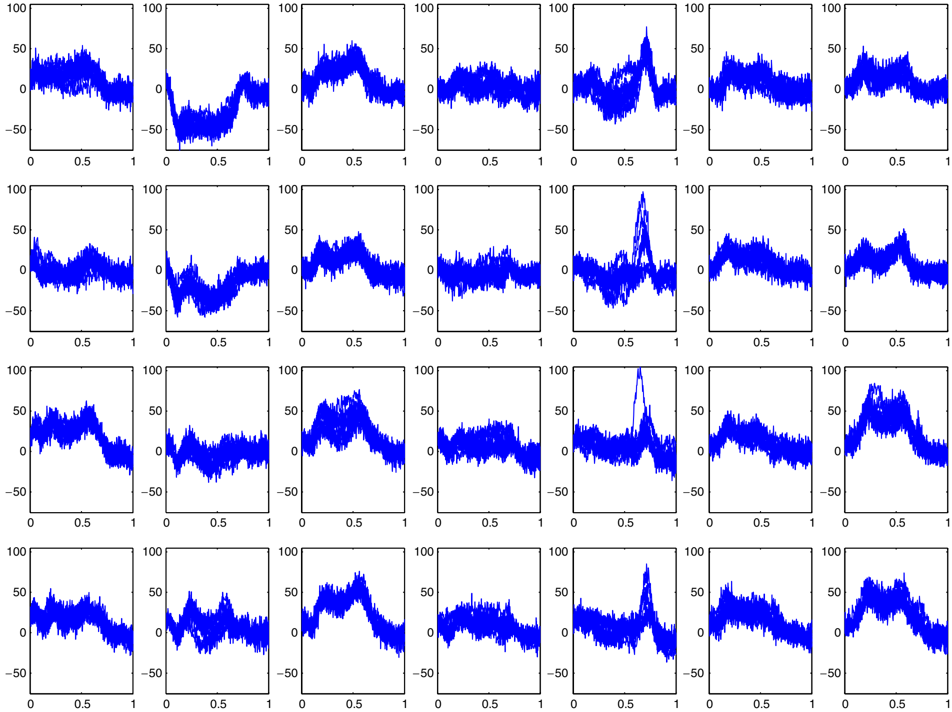

Fig. 5. Orthosis data set: the panels in rows correspond to *Treatments* while the panels in columns correspond to *Subjects*; there are ten repeated measurements in each panel.

and the null hypothesis is therefore rejected. It is interesting to compare this result with that of the fixed-effects FANOVA testing procedure of Abramovich et al. (2004) where the corresponding test statistic was 88,854.84 (although they used the non-adaptive test with $J_\theta = 4$). In both cases the results are significant but the difference in the two test statistics is tremendous due to the high variability among subjects relatively tothe noise level.

5. Concluding remarks

In this paper we considered the testing problem in the mixed-effects FANOVA model which arise in various applications involving longitudinal data. We extended the corresponding results of Abramovich et al. (2004) for the fixed-effects FANOVA and derived optimal (in the minimax sense) non-adaptive and adaptive testing procedures. The proposed tests are computationally fast and can be easily implemented.

Several possible extensions of the obtained results should be mentioned. Although for simplicity of exposition we considered the two-way mixed-effects FANOVA model, the similar techniques can be applied for the case of several fixed and/or random effects. The extensions from the analysis of one-dimensional curves to higher-dimensional data and

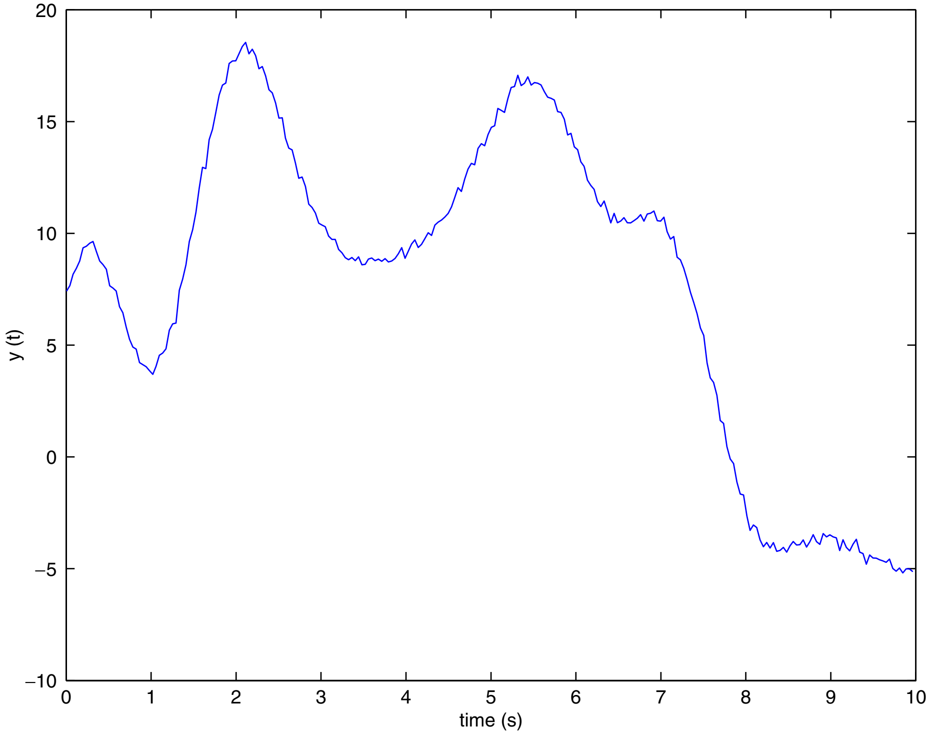

Fig. 6. Observed process $\bar{y}(t)$.

to the two-dimensional image analysis in particular is straightforward by using the corresponding higher-dimensional wavelet transforms.

Acknowledgements

Claudia Angelini would like to thank Felix Abramovich for the excellent hospitality while visiting Tel Aviv University to carry out this project. Both authors would like to thank Anestis Antoniadis for fruitful discussions. Constructive and helpful comments of the anonymous referees are gratefully acknowledged.

Appendix

Proof of Theorem 1. The statistics $T(J_\theta)$ and $Q(J_\theta)$ are the sums of J_θ and $J_\eta - J_\theta$ independent, squared integrable random variables that under the null hypothesis have zero means and variances $v_0^2(J_\theta)$ and $\omega_0^2(J_\theta)$, respectively. By the central limit theorem, the resulting standardized test statistic in (24) is then asymptotically (as $\eta \to 0$) standard normal and the significance level of ϕ^* is therefore asymptotically α.

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

$$\begin{aligned} &P_\mu(\phi^* = 0) \\ &\quad = \Phi\left(\left\{\frac{v_0^2(J_\theta) + \omega_0^2(J_\theta)}{\mathrm{Var}_\mu(T(J_\theta) + Q(J_\theta))}\right\}^{1/2} z_{1-\alpha} - \frac{E_\mu(T(J_\theta) + Q(J_\theta))}{\{\mathrm{Var}_\mu(T(J_\theta)) + Q(J_\theta))\}^{1/2}}\right) \\ &\qquad + \mathrm{o}_\eta(1). \end{aligned}$$

Since $(v_0^2(J_\theta) + \omega_0^2(J_\theta))/(\mathrm{Var}_\mu(T(J_\theta) + Q(J_\theta)))$ is bounded above by one, the asymptotic behavior of $P_\mu(\phi^* = 0)$ depends only on the ratio $E_\mu(T(J_\theta) + Q(J_\theta))/(\mathrm{Var}_\mu(T(J_\theta) + Q(J_\theta))^{1/2}$.

The following lemmas provide the necessary bounds for $E_\mu(T(J_\theta) + Q(J_\theta))$ and $\mathrm{Var}_\mu(T(J_\theta) + Q(J_\theta))^{1/2}$. Their proofs are given at the end of the section.

Define $p' = \min(p, 2)$ and $s' = s + 1/2 - 1/p'$. Then,

Lemma 1. *For any* $\mu \in \mathscr{F}(\rho(\eta))$,

$$E_\mu(T(J_\theta) + Q(J_\theta)) \geqslant \tfrac{1}{2}\|\mu\|_2^2 - M^2\eta^{4s'} - c_1\eta^{8s''/(4s''+1)}.$$

Lemma 2. *For any* $\mu \in \mathscr{F}(\rho(\eta))$,

$$\mathrm{Var}_\mu(T(J_\theta) + Q(J_\theta)) \leqslant c_2\eta^2\|\mu\|_2^2 + c_3\eta^{(16s''+2)/(4s''+1)} + c_4\eta^{(16s'')/(4s''+1)}.$$

Recall that $\|\mu\|_2 \geqslant \rho(\eta)$ for all $\mu \in \mathscr{F}(\rho(\eta))$. Then, substituting $\rho(\eta) = \eta^{4s''/(4s''+1)}$, Lemmas 1 and 2 imply that for any given β there exists a constant c_β such that

$$\inf_{\mu \in \mathscr{F}(c_\beta \rho(\eta))} \frac{E_\mu(T(J_\theta) + Q(J_\theta))}{(\mathrm{Var}_\mu(T(J_\theta) + Q(J_\theta)))^{1/2}} > \tilde{c}_\beta,$$

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

$$\beta(\phi^*, c_\beta \rho(\eta)) \leqslant \beta + \mathrm{o}_\eta(1).$$

This shows that the test ϕ^* achieves the lower bound (15) for the minimax rate and therefore is the rate-optimal. □

Proof of Lemma 1. Note first that

$$E_\mu(T(J_\theta)) = \sum_{j=0}^{J_\theta - 1} \sum_{k=0}^{2^j - 1} \mu_{jk}^2. \tag{32}$$

Apply Lemma 4.4 of Spokoiny (1996) that was originally stated for normally distributed random variables but, in fact, the same proof holds for any symmetric distribution:

$$\begin{aligned} E_\mu(Q(J_\theta)) &\geqslant \frac{1}{2}\sum_{j=J_\theta}^{J_\eta-1}\sum_{k=0}^{2^j-1}\mu_{jk}^2\mathbf{1}\{|\mu_{jk}|\geqslant\eta\lambda_j\} \\ &\geqslant \frac{1}{2}\left(\sum_{j=J_\theta}^{J_\eta-1}\sum_{k=0}^{2^j-1}\mu_{jk}^2 - \sum_{j=J_\theta}^{J_\eta-1}\sum_{k=0}^{2^j-1}\mu_{jk}^2\mathbf{1}\{|\mu_{jk}|\leqslant\eta\lambda_j\}\right). \end{aligned} \tag{33}$$

For any $\mu \in B_{p,q}^s(M)$, $\sum_{k=0}^{2^j-1}|\mu_{jk}|^{p'} \leqslant M^{p'}2^{-js'p'}$ for all $j \geqslant J_\theta$ (e.g., Meyer, 1992). Then, similar to Spokoiny (1996)

$$\begin{aligned} &\sum_{j=J_\theta}^{J_\eta-1}\sum_{k=0}^{2^j-1}\mu_{jk}^2\mathbf{1}\{|\mu_{jk}|\leqslant\eta\lambda_j\} \\ &\quad\leqslant \sum_{j=J_\theta}^{J_\eta-1}\sum_{k=0}^{2^j-1}(\eta\lambda_j)^{2-p'}|\mu_{jk}|^{p'}\mathbf{1}\{|\mu_{jk}|\leqslant\eta\lambda_j\} \\ &\quad\leqslant \eta^{2-p'}\sum_{j=J_\theta}^{J_\eta-1}\lambda_j^{2-p'}\sum_{k=0}^{2^j-1}|\mu_{jk}|^{p'} \leqslant M^{p'}\eta^{2-p'}\sum_{j=J_\theta}^{J_\eta-1}\lambda_j^{2-p'}2^{-js'p'} \\ &\quad\leqslant C\eta^{2-p'}2^{-s'p'J_\theta}\sum_{l\geqslant 0}2^{-ls'p'}\left(\sqrt{(1+\kappa_{l+J_\theta}^2)(l+8)\ln 2}\right)^{2-p'} \\ &\quad\leqslant C\eta^{2-p'}2^{-s'p'J_\theta}. \end{aligned} \tag{34}$$

On the other hand,

$$\sum_{j=J_\eta}^{\infty}\sum_{k=0}^{2^j-1}\mu_{jk}^2 \leqslant \sum_{j=J_\eta}^{\infty}\left(\sum_{k=0}^{2^j-1}\mu_{jk}^{p'}\right)^{2/p'} \leqslant M^2\sum_{j=J_\eta}^{\infty}2^{-2js'} \leqslant 2M^2\eta^{4s'}. \tag{35}$$

Combining (32)–(35) one finally has

$$\begin{aligned} E_\mu(T(J_\theta)+Q(J_\theta)) &\geqslant \frac{1}{2}\left(\sum_{j=0}^{J_\eta-1}\sum_{k=0}^{2^j-1}\mu_{jk}^2 - \sum_{j=J_\theta}^{J_\eta-1}\sum_{k=0}^{2^j-1}\mu_{jk}^2\mathbf{1}\left\{|\mu_{jk}|\leqslant\eta\lambda_j\right\}\right) \\ &\geqslant \frac{1}{2}\left(\|\mu\|_2^2 - \sum_{j=J_\eta}^{\infty}\sum_{k=0}^{2^j-1}\mu_{jk}^2 - \sum_{j=J_\theta}^{J_\eta-1}\sum_{k=0}^{2^j-1}\mu_{jk}^2\mathbf{1}\left\{|\mu_{jk}|\leqslant\eta\lambda_j\right\}\right) \\ &\geqslant \frac{1}{2}\|\mu\|_2^2 - M^2\eta^{4s'} - c_1\eta^{2-p'}2^{-J_\theta s'p'}. \end{aligned} \tag{36}$$

Substituting J_θ from (18) into (36) completes the proof of Lemma 1. □

Proof of Lemma 2. Clearly,

$$\mathrm{Var}_\mu(T(J_\theta)) = 2\eta^4 \sum_{j=0}^{J_\theta-1} 2^j \left(1 + \pi_j \kappa_j^4 \left(\frac{3}{2}\frac{1-\pi_j}{m} + \pi_j\right) + 2\pi_j\kappa_j^2\right) + 4\eta^2 \sum_{j=0}^{J_\theta-1} (1+\pi_j\kappa_j^2) \sum_{k=0}^{2^j-1} \mu_{jk}^2. \tag{37}$$

Consider now $\mathrm{Var}_\mu(Q(J_\theta))$. Repeating the proof of Lemma 4.5 of Spokoiny (1996) for our model with corresponding obvious changes we have

$$\begin{aligned}
&\mathrm{Var}_\mu(Y_{jk}^2 \mathbf{1}\{|Y_{jk}| > \eta\lambda_j\}) \\
&\quad \leqslant 4\eta^2(1+\pi_j\kappa_j^2)\mu_{jk}^2 + \eta^4\lambda_j^4 e^{-\lambda_j^2/(8(1+\kappa_j^2))} \\
&\qquad + 2\eta^4 \left(1 + \pi_j\kappa_j^4\left(\frac{3}{2}\frac{1-\pi_j}{m} + \pi_j\right) + 2\pi_j\kappa_j^2\right) \mathbf{1}\left\{|\mu_{jk}| > \frac{1}{2}\eta\lambda_j\right\}.
\end{aligned} \tag{38}$$

A straightforward calculus yields

$$\begin{aligned}
&\sum_{j=J_\theta}^{J_\eta-1} \sum_{k=0}^{2^j-1} \mathbf{1}\left\{|\mu_{jk}| > \frac{1}{2}\eta\lambda_j\right\} \\
&\quad \leqslant \sum_{j=J_\theta}^{J_\eta-1} \sum_{k=0}^{2^j-1} \left(\frac{1}{2}\lambda_j\eta\right)^{-p'} |\mu_{jk}|^{p'} \leqslant C \sum_{j=J_\theta}^{J_\eta-1} (\lambda_j\eta)^{-p'} 2^{-js'p'} \\
&\quad \leqslant C\eta^{-p'} 2^{-J_\theta s'p'} \sum_{l\geqslant 0} \left(\sqrt{(1+\kappa_{l+J_\theta}^2)(l+8)}\right)^{-p'} 2^{-ls'p'} \\
&\quad \leqslant C\eta^{-p'} 2^{-J_\theta s'p'}
\end{aligned} \tag{39}$$

and for the thresholds λ_j defined in (22)

$$\sum_{j=J_\theta}^{J_\eta-1} \sum_{k=0}^{2^j-1} \lambda_j^4 \mathrm{e}^{-\lambda_j^2/(8(1+\kappa_l^2))} \leqslant C2^{J_\theta}. \tag{40}$$

Since $\limsup_j \kappa_j^2$ is finite, (38)–(40) imply the following upper bound on $\mathrm{Var}_\mu(Q(J_\theta))$:

$$\mathrm{Var}_\mu(Q(J_\theta)) \leqslant c_2\eta^2\|\mu\|_2^2 + c_3\eta^{4-p'}2^{-J_\theta s'p'} + c_4\eta^4 2^{J_\theta}. \tag{41}$$

Combining (41) with (37) and substituting J_θ from (18) complete the proof of Lemma 2. □

References

Abramovich, F., Sapatinas, T., Silverman, B.W., 1998. Wavelet thresholding via a Bayesian approach. J.R. Statist. Soc. (Ser. B) 60, 725–749.

Abramovich, F., Antoniadis, A., Sapatinas, T., Vidakovic, B., 2004. Optimal testing in functional analysis of variance models. Int. J. Wavelets Multiresolution Inform. Process. 2, 323–349.

Angelini, C., De Canditiis, D., Leblanc, F., 2003. Wavelet regression estimation in nonparametric mixed effect models. J. Multivariate Anal. 85, 267–291.

Antoniadis, A., 1984. Analysis of variance on function spaces. Math. Operationsforsch. Statist. Ser. Statist. 15, 59–71.

Cahouët, V., Martin, L., Amarantini, D., 2002. Static optimal estimation of joint accelerations for inverse dynamic problem solution. J. Biomech. 35, 1507–1513.

Chen, J.C., 1994. Testing for no effect in nonparametric regression via spline smoothing techniques. Ann. Inst. Statist. Math. 45, 251–265.

Clyde, M., George, E.I., 2000. Flexible empirical Bayes estimation with wavelets. J.R. Statist. Soc. B 62, 681–698.

Clyde, M., Parmigiani, G., Vidakovic, B., 1998. Multiple shrinkageand subset selection in wavelets. Biometrika 85, 391–401.

Daubechies, I., 1992. Ten Lectures on Wavelets. SIAM, Philadelphia.

Dette, H., Derbort, S., 2001. Analysis of variance in nonparametric regression models. J. Multivariate Anal. 76, 110–137.

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

Eubank, R.L., 2000. Testing for no effect by cosine series methods. Scand. J. Statist. 27, 747–763.

Eubank, R.L., La Riccia, V.N., 1993. Testing for no effect in non-parametric regression. J. Statist. Planning Inference 36, 1–14.

Fan, J., 1996. Test of significance based on wavelet thresholding and Neyman's truncation. J. Am. Statist. Assoc. 91, 674–688.

Fan, J., Lin, S.-K., 1998. Test of significance when data are curves. J. Am. Statist. Assoc. 93, 1007–1021.

Faraway, J.J., 1997. Regression analysis for a functional response. Technometrics 39, 254–261.

Gu, C., 2002. Smoothing Spline ANOVA Models. Springer, New York.

Guo, W., 2002. Inference in smoothing spline analysis of variance. J. R. Statist. Soc. B 64, 887–898.

Huang, J.Z., 1998. Projection estimation in multiple regression with application to functional ANOVA models. Ann. Statist. 26, 242–272.

Ingster, Yu.I., 1982. Minimax non parametric detection of signals in white Gaussian noise. Problems Inform. Transmission 18, 130–140.

Ingster, Yu.I., 1993. Asymptotically minimax hypothesis testing for nonparametric alternatives: I, II, III. Math. Methods Statist. 2, 85–114 3, 171–189; 4, 249–268.

Ingster, Yu.I., Suslina, I.A., 2003. Nonparametric Goodness-of-Fit Testing under Gaussian Models. Lecture Notes in Statistics, vol. 169. Springer, New York.

Johnstone, I.M., Silverman, B.W., 2005. Empirical Bayes selection of wavelet thresholds. Ann. Statist. 33, to appear.

Lepski, O.V., Spokoiny, V.G., 1999. Minimax nonprametric hypothesis testing: the case of an inhomogeneous alternative. Bernulli 5, 333–358.

Lin, Y., 2000. Tensor product space ANOVA models. Ann. Statist. 28, 734–755.

Meyer, Y., 1992. Wavelets and Operators. Cambridge University Press, Cambridge.

Ramsay, J.O., Silverman, B.W., 1997. Functional Data Analysis. Springer, NewYork.

Ramsay, J.O., Silverman, B.W., 2002. Applied Functional Data Analysis. Springer, New York.

Raz, J., 1990. Testing for no effect when estimating a smooth function by nonparametric regression: a randomization approach. J. Am. Statist. Assoc. 85, 132–138.

Spitzner, D.J., Marron, J.S., Essick, G.K., 2003. Mixed-model functional ANOVA for studying human tactile perception. J. Am. Statist. Assoc. 98, 263–272.

Spokoiny, V.G., 1996. Adaptive hypothesis testing using wavelets. Ann. Statist. 24, 2477–2498.

Stone, C.J., Hansen, M., Kooperberg, C., Truong, Y., 1997. Polynomial splines and their tensor products in extended linear modeling (with discussion). Ann. Statist. 25, 1371–1470.

Wahba, G., Wang, Y., Gu, C., Klein, R., Klein, B., 1995. Smoothing spline ANOVA for exponential families, with application to the Wisconsin epidemiological study of diabetic retinopathy. Ann. Statist. 23, 1865–1895.