

OPTIMAL TESTING IN FUNCTIONAL ANALYSIS OF VARIANCE

MODELS

Felix Abramovich,

Department of Statistics and Operations Research, Tel Aviv University,
Tel Aviv 69978, Israel.

Anestis Antoniadis,

Laboratoire IMAG-LMC, University Joseph Fourier,
BP 53, 38041 Grenoble Cedex 9, France.

Theofanis Sapatinas,

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

and

Brani Vidakovic,

Industrial and Systems Engineering, Georgia Institute of Technology,
Atlanta, Georgia 30332-0205, USA.

Authors' Footnote:

Felix Abramovich is Senior Lecturer at Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv 69978, Israel. email: felix@math.tau.ac.il

Anestis Antoniadis is Professor at Laboratoire IMAG-LMC, University Joseph Fourier, BP 53, 38041 Grenoble Cedex 9, France. email: Anestis.Antoniadis@imag.fr

Theofanis Sapatinas is Assistant Professor in Department of Mathematics and Statistics, University of Cyprus, P.O.B. 20537, CY 1678 Nicosia, Cyprus. email: T.Sapatinas@ucy.ac.cy

Brani Vidakovic is Associate Professor in the School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA. email: brani@isye.gatech.edu

Abstract

We consider the testing problem in a fixed-effects functional analysis of variance model. We test the null hypotheses that the main effects and the interactions are zeros against the composite nonparametric alternative hypotheses that they are separated away from zero in L^2 -norm and also possess some smoothness properties. We adapt the minimax functional hypothesis testing procedures for testing a zero signal in a Gaussian “signal plus noise” model to derive asymptotically optimal (minimax) non-adaptive and adaptive functional hypothesis testing procedures for the main effects and the interactions based on the empirical wavelet coefficients of the data. Wavelet decompositions allow one to characterize different types of smoothness conditions assumed on the response function by means of its wavelet coefficients for a wide range of various function classes.

In order to shade some light on the theoretical results obtained, we carry out a small simulation study to examine the finite sample performance of the proposed functional hypothesis testing procedures. As an illustration, we also apply these tests to a real-life data example arising from physiology. Concluding remarks and hints for possible extensions of the proposed methodology are also given.

KEYWORDS: Analysis of Variance; Besov Spaces; Functional Analysis of Variance; Nonparametric Hypothesis Testing; Wavelets.

1. INTRODUCTION

Analysis of variance (ANOVA) is one of the most widely used tools in applied statistics. While very useful for handling low dimensional data, it has its limitations in analyzing *functional* responses. Such responses are encountered, for example, when units are observed over time or when, although a whole function itself is not observed, a sufficiently large number of evaluations over individual is available – a common feature of modern recording equipments. Sophisticated on-line sensing and monitoring equipments are now routinely used in research in medicine, seismology, meteorology, physiology, and many other fields. For instance, in the traditional analysis of electro-encephalogram (EEG) data, experts record EEG measurements of healthy men and women of different ages starting from young adults through middle ages to get a refined understanding of the variation in EEG due to age and gender. The general question to be answered is then the influence of age and gender on the shape of the EEG measurements.

In such cases, *functional analysis of variance* (FANOVA) methods provide alternatives to classical ANOVA methods while still allowing a simple interpretation. Due to a large set of applications involving functional data, FANOVA models have recently gained popularity and related literature has been steadily growing. For comprehensive reviews we refer, for example, to Ramsay & Silverman (1997) and Stone *et al.* (1997). Although there is an impressive literature on *fitting* FANOVA models and *estimating* their components (e.g., Wahba *et al.*, 1995; Stone *et al.*, 1997; Huang, 1998; Lin, 2000; Gu, 2002), there is no much work on developing *hypothesis testing* procedures in FANOVA models. This paper considers the testing problem in a fixed-effects FANOVA model and derives asymptotically optimal (minimax) non-adaptive and adaptive tests for the main effects and the interactions.

First we describe the following white noise (or diffusion) version of the FANOVA model we are going to consider hereafter. Suppose that one observes a series of sample paths of a stochastic process driven by

$$dY_i(\mathbf{t}) = m_i(\mathbf{t}) d\mathbf{t} + \epsilon dW_i(\mathbf{t}), \quad i = 1, \dots, r; \quad \mathbf{t} \in [0, 1]^d, \quad (1)$$

where $\epsilon > 0$ is the diffusion coefficient, r and d are finite integers, m_i are (unknown) d -dimensional response functions and W_i are independent d -dimensional standard Wiener processes. In most

applications, we are interested in the cases $d = 1$ (a set of *signals*) and $d = 2$ (a set of *images*).

In practice, obviously, one always observes *discrete* data samples of size n with a noise variance σ^2 , but from the well-known results of Brown & Low (1996), under some general conditions, the corresponding discrete model is asymptotically equivalent to the white noise model (1) with $\epsilon = \sigma/\sqrt{n}$.

Following Antoniadis (1984), each of the r response functions in model (1) admits the following unique decomposition

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

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

$$\int_{[0,1]^d} \mu(\mathbf{t}) \, d\mathbf{t} = 0, \quad \sum_{i=1}^r a_i = 0, \quad (3)$$

$$\sum_{i=1}^r \gamma_i(\mathbf{t}) = 0, \quad \int_{[0,1]^d} \gamma_i(\mathbf{t}) \, d\mathbf{t} = 0, \quad \forall i = 1, \dots, r; \quad \mathbf{t} \in [0, 1]^d. \quad (4)$$

As in the traditional fixed-effects ANOVA models, one is naturally interested in testing the significance of the main effects and the interactions in the fixed-effects FANOVA model (1)-(4). Since these testing problems involve functional data, we call them *functional hypothesis testing* problems. A first, somewhat naive, approach to the functional hypothesis testing is to consider the FANOVA problem as a standard *univariate* ANOVA problem for each specific t (e.g., Ramsay & Silverman, 1997) and perform a series of, say, corresponding pointwise F -tests. A crucial drawback of this approach is that an enormous number of hypotheses (the number of data points per curve can be hundreds or thousands) has to be tested simultaneously that causes a serious multiplicity problem. Ignoring multiplicity leads to an uncontrolled overall Type I error while, for example, Bonferroni type procedures are known to yield an extremely low power of the test. Another approach to fixed-effects FANOVA testing considered in the literature is to treat functional data as multivariate vectors and to apply traditional multivariate fixed-effects ANOVA techniques combined

sometimes with various initial dimensionality-reduction procedures (e.g., Raz, 1990; Eubank & La Riccia, 1993; Chen, 1994). However, the “curse of dimensionality” makes these attempts also problematic. Faraway (1997) discussed the difficulties of generalizing the ideas of multivariate testing procedures to the functional data analysis context. Fan & Lin (1998) proposed a powerful overall test for functional hypothesis testing based on the decomposition of the original functional data into Fourier and wavelet series expansions, and applied the adaptive Neyman and wavelet thresholding procedures of Fan (1996) to the resulting empirical Fourier and wavelet coefficients respectively. The underlying idea is based on the *sparsity* of the underlying signal’s representation in the Fourier or wavelet domains that allows a significant reduction of dimensionality. Somewhat similar approaches were considered in Eubank (2000) and Dette & Derbort (2001). Recently, Guo (2002) suggested a maximum likelihood ratio based test for functional variance components in mixed-effects FANOVA models. However, none of the above works investigates the optimality of the proposed tests.

In this paper we derive asymptotically (as $\epsilon \rightarrow 0$ or, equivalently, as $n \rightarrow \infty$ in the discrete version) optimal (minimax) *non-adaptive* and *adaptive* functional hypothesis testing procedures for testing the significance of the main effects and the interactions in the FANOVA model (1)-(4) against the composite nonparametric alternatives that they are separated away from zero in $L^2([0, 1]^d)$ -norm and also possess some smoothness properties. To derive the tests, we adapt the asymptotically optimal (minimax) functional hypothesis testing procedures for testing a zero signal in a Gaussian “signal plus noise” model originated by Ingster (1982) and further developed in Ermakov (1990), Ingster (1993), Spokoiny (1996, 1998), Lepski & Spokoiny (1999), Ingster & Suslina (2000) and Horowitz & Spokoiny (2001) for various separation distances between the two hypotheses and different function classes under the alternative. The proposed functional hypothesis testing procedures are based on the empirical wavelet coefficients of the data. Wavelet decompositions allow one to characterize different types of smoothness conditions assumed on the response function by means of its wavelet coefficients for a wider range of function classes than the ones obtained by, for example, their Fourier counterparts (e.g., Meyer, 1992).

The paper is organized as follows. In Section 2 we formulate the hypotheses to be tested and provide definitions and background on functional hypothesis testing and wavelet analysis necessary for

the proposed methodology. In Section 3 we derive asymptotically optimal (minimax) non-adaptive and adaptive functional hypothesis testing procedures for the main effects and the interactions. In Section 4 we carry out a small simulation study to examine the finite sample performance of the proposed functional hypothesis testing procedures. As an illustration, we also apply these tests to a real-life data example arising from physiology. Finally, in Section 5, we provide concluding remarks and provide some hints for possible extensions of the proposed methodology.

2. FORMULATIONS AND DEFINITIONS

2.1 FORMULATION OF THE HYPOTHESES TO BE TESTED

In Section 1 we defined unique orthogonal decompositions (2) of the response functions m_i in model (1). Due to (2), testing the significance of the main effects and the interactions is equivalent to testing the following hypotheses

$$H_0 : \mu(\mathbf{t}) \equiv 0, \quad \mathbf{t} \in [0, 1]^d, \quad (\text{no global trend}) \quad (5)$$

$$H_0 : a_i = 0, \quad \forall i = 1, \dots, r, \quad (\text{no differences in level}) \quad (6)$$

$$H_0 : \gamma_i(\mathbf{t}) \equiv 0, \quad \forall i = 1, \dots, r, \quad \mathbf{t} \in [0, 1]^d \quad (\text{no differences in shape}). \quad (7)$$

Integrating (1) with respect to \mathbf{t} and using the identifiability conditions (3)-(4), we have

$$Y_i^* = m_0 + a_i + \epsilon \xi_i, \quad i = 1, \dots, r, \quad \sum_{i=1}^r a_i = 0,$$

where $Y_i^* = \int_{[0,1]^d} dY_i(\mathbf{t})$ and ξ_i are independent $N(0, 1)$ random variables. This is the classical one-way fixed-effects ANOVA model, so testing (6) (no differences in level) can be performed by standard techniques.

We focus then on *functional* hypothesis testing for the null hypotheses (5) and (7). We do not specify any parametric forms on μ and γ_i under the alternative hypotheses and wish to test the corresponding null hypotheses against as large class of alternatives as possible. In particular, assume that m_i (and, hence, μ and γ_i as well) belong to some Besov ball of radius $C > 0$ on $[0, 1]^d$, $B_{p,q}^s(C)$, where $s > 0$ and $1 \leq p, q \leq \infty$. The parameter p can be viewed as a degree of function's inhomogeneity while s is a measure of its smoothness. Roughly speaking, the (not necessarily integer) parameter s indicates the number of function's 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, among others, the Hölder and Sobolev classes of smooth functions, and the class of functions of bounded variation. We refer to Meyer (1992) for rigorous definitions and a detailed study of Besov spaces.

In addition, since our concern will be the rate at which the distance between the null and alternative hypotheses can decrease to zero while still permitting consistent testing, we also need to assume that the set of alternative hypotheses is separated away from the set of null hypotheses (5) and (7) in the $L^2([0, 1]^d)$ -distance ρ . Hence, denoting hereafter the $L^2([0, 1]^d)$ -norm by $\|\cdot\|_2$, we consider the alternative hypotheses to be, respectively, of the form

$$H_1 : \mu \in \mathcal{F}(\rho), \tag{8}$$

$$H_1 : \gamma_i \in \mathcal{F}(\rho), \quad \text{at least for one } i = 1, \dots, r, \tag{9}$$

where $\mathcal{F}(\rho) = \{f \in B_{p,q}^s(C) : \|f\|_2 \geq \rho\}$.

In Section 3 we derive non-adaptive and adaptive functional hypothesis testing procedures for the null hypotheses (5) and (7) that are asymptotically optimal (minimax) uniformly over the whole range of Besov balls under the alternatives (8) and (9) respectively.

2.2 BASIC DEFINITIONS

We start from basic definitions of the functional hypothesis testing. Suppose we are given data

$$dZ(\mathbf{t}) = f(\mathbf{t}) d\mathbf{t} + \epsilon dW(\mathbf{t}), \quad \mathbf{t} \in [0, 1]^d, \tag{10}$$

where W is a d -dimensional standard Wiener process. We wish to test

$$H_0 : f \equiv 0 \quad \text{versus} \quad H_1 : f \in \mathcal{F}(\rho), \tag{11}$$

where $\mathcal{F}(\rho) = \{f \in B_{p,q}^s(C) : \|f\|_2 \geq \rho\}$.

A (non-randomized) test ϕ is a measurable function of the observations with two values $\{0, 1\}$, where $\phi = 0$ and $\phi = 1$ correspond to accepting and rejecting the null hypothesis H_0 respectively. As usual, the quality of test ϕ is measured by *Type I* (erroneous rejection of H_0) and *Type II* (erroneous acceptance of H_0) errors. The probability of Type I error is defined as

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

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

$$\beta(\phi, \rho) = \sup_{f \in \mathcal{F}(\rho)} P_f(\phi = 0).$$

In this paper we focus on the asymptotic behavior of functional hypothesis testing procedures as $\epsilon \rightarrow 0$ (that will correspond to $n \rightarrow \infty$ in the discrete version of (10) – see Section 1). It is clear that the smaller the ϵ , the less separated from zero alternatives can be detected without losing accuracy. For prescribed α and β , the rate of decay to zero of the “indifference threshold” $\rho = \rho(\epsilon)$, as $\epsilon \rightarrow 0$, can be viewed as a measure of goodness of a test. Thus, it is natural to seek the test with the optimal (fastest) rate. The optimal (minimax) rate of testing $\rho(\epsilon)$ is defined as follows (see Ingster, 1993; Spokoiny, 1996; Ingster & Suslina, 2000).

Definition 2.1 *A sequence $\rho(\epsilon)$ is the minimax rate of testing if $\rho(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$ and the following two conditions hold*

(i) *for any $\rho'(\epsilon)$ satisfying*

$$\rho'(\epsilon)/\rho(\epsilon) = o_\epsilon(1),$$

one has

$$\inf_{\phi_\epsilon} [\alpha(\phi_\epsilon) + \beta(\phi_\epsilon, \rho'(\epsilon))] = 1 - o_\epsilon(1),$$

where $o_\epsilon(1)$ is a sequence tending to zero as $\epsilon \rightarrow 0$.

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

$$\alpha(\phi_\epsilon^*) \leq \alpha + o_\epsilon(1)$$

$$\beta(\phi_\epsilon^*, c\rho(\epsilon)) \leq \beta + o_\epsilon(1).$$

The corresponding test ϕ_ϵ^* is called an asymptotically optimal (minimax) test. Ingster (1993) and Lepski & Spokoiny (1999) showed that for $sp > d$ the asymptotically optimal (minimax) rate for such a testing problem is

$$\rho(\epsilon) = \epsilon^{4s''/(4s''+d)}, \tag{12}$$

where $s'' = \min(s, s - \frac{d}{2p} + \frac{d}{4})$. They also derived tests that achieve this optimal rate. However, although the proposed asymptotically optimal (minimax) tests were consistent, they are *non-adaptive*

in the sense that they involve the smoothness parameters s and p of the corresponding Besov ball which are usually unknown in practice. Spokoiny (1996) and Horowitz & Spokoiny (2001) considered the problem of *adaptive* minimax testing where s and p are unknown *a priori* but assumed to lie within a given range. They showed that no adaptive test can achieve the exact optimal rate (12) uniformly over all s and p within this range –“lack of adaptivity” property of the functional hypothesis testing problem, and there is always a price to pay for adaptivity. However, it turned out that the price is remarkably low. If one allows to increase $\rho(\epsilon)$ by an additional log-log adaptive factor $t_\epsilon = (\ln \ln \epsilon^{-2})^{1/4}$, i.e. to consider $\rho(\epsilon t_\epsilon)$ instead of $\rho(\epsilon)$, then Horowitz & Spokoiny (2001) showed that the optimal rate of adaptive testing is

$$\rho(\epsilon t_\epsilon) = (\epsilon t_\epsilon)^{4s''/(4s''+d)}, \quad (13)$$

which is only within a log-log factor of (12). Moreover, they showed that the resulting asymptotically optimal (minimax) adaptive test is uniformly consistent. The adaptive factor t_ϵ is unavoidable and cannot be reduced. For more details we refer to Horowitz & Spokoiny (2001).

2.3 WAVELET BACKGROUND

Since the test statistics we shall develop for the main effects and the interactions in the FANOVA model (1)-(2) will be based on appropriate wavelet decompositions, we briefly recall some relevant facts about wavelets. For detailed expositions of the mathematical aspects of wavelets we refer, for example, to Meyer (1992) and Daubechies (1992), while comprehensive expositions and reviews on wavelets applications in statistical settings are given, for example, in Antoniadis (1997), Vidakovic (1999) and Abramovich *et al.* (2000).

To simplify the notation, we consider the case $d = 1$ and work with orthonormal periodic wavelet bases in $L^2([0, 1])$ generated by dilations of a compactly supported scaling function ϕ and dilations and translations of a corresponding compactly supported mother wavelet ψ , i.e.

$$\phi^P(t) = \sum_{l \in \mathbb{Z}} \phi(t - l) \quad \text{and} \quad \psi_{jk}^P(t) = \sum_{l \in \mathbb{Z}} \psi_{jk}(t - l), \quad j \geq 0, \quad k = 0, \dots, 2^j - 1$$

where

$$\psi_{jk}(t) = 2^{j/2} \psi(2^j t - k).$$

The collection

$$\{\phi^p; \psi_{jk}^p, j \geq 0, k = 0, 1, \dots, 2^j - 1\}$$

generates an orthonormal periodic wavelet basis in $L^2([0, 1])$. Despite the poor behavior of periodic wavelets near the boundaries, where they yield large coefficients for non-periodic functions, they are commonly used because their numerical implementation is particularly simple. To unify the notations, the scaling function ϕ^p will be denoted by $\psi_{-1,0}^p$ and the superscript ‘‘p’’ will be suppressed for convenience. For any $f \in L^2([0, 1])$ we denote by $\theta_{-10} = \langle f, \psi_{-10} \rangle$ the scaling coefficient and by $\theta_{jk} = \langle f, \psi_{jk} \rangle, j \geq 0, k = 0, 1, \dots, 2^j - 1$ the wavelet coefficients of f for the orthonormal periodic wavelet basis defined above.

We end this section by noting that if the scaling function ϕ (and, thus, the mother wavelet ψ) is of regularity $r > 0$, the corresponding wavelet basis is an unconditional basis in the Besov spaces $B_{p,q}^s([0, 1])$ for $0 < s < r, 1 \leq p, q \leq \infty$. This allows one to characterize Besov balls in terms of wavelet coefficients (e.g., Meyer, 1992).

3. MAIN RESULTS

In this section we apply the results of Spokoiny (1996) and Horowitz & Spokoiny (2001) to derive asymptotically optimal (minimax) non-adaptive and adaptive functional hypothesis testing procedures for the main effects and the interactions in the FANOVA model (1)-(4). To simplify the exposition and to emphasize the main idea we consider in detail the case $d = 1$. This restriction will be relaxed in Remark 3.3 below wherein we briefly explain straightforward extensions to the case $d > 1$.

3.1 TESTING IN FIXED-EFFECTS FUNCTIONAL ANALYSIS OF VARIANCE MODELS

Averaging over the r observed paths in the FANOVA model (1)-(2) yields

$$d\bar{Y}(t) = \frac{1}{r} \sum_{i=1}^r dY_i(t) = \left(\frac{1}{r} \sum_{i=1}^r m_i(t) \right) dt + \epsilon d\bar{W}(t), \quad t \in [0, 1],$$

where \bar{W} is the average of r independent standard Wiener processes on $[0, 1]$. Given the identifiability conditions (3)-(4) for the components of m_i , the latter equation can be rewritten as

$$d\bar{Y}(t) = (m_0 + \mu(t)) dt + \epsilon d\bar{W}(t), \quad t \in [0, 1] \tag{14}$$

which, in view of (1)-(2), implies

$$d(Y_i - \bar{Y})(t) = a_i + \gamma_i(t)dt + \epsilon d(W_i - \bar{W})(t), \quad i = 1, \dots, r; \quad t \in [0, 1]. \quad (15)$$

By the basic properties of the increments of a standard Wiener process on $[0, 1]$, the stochastic processes $\{W_i - \bar{W}; i = 1, \dots, r\}$ are Wiener processes with the same covariance kernel $C(s, t) = \frac{r-1}{r} \min(s, t)$, though they are no longer independent. Hence, the model (14) and each of the r models stated in the equation (15) can be written in the following general form

$$dZ(t) = f(t) dt + \eta dW(t), \quad t \in [0, 1], \quad (16)$$

where $Z(t) = \bar{Y}(t)$, $f(t) = m_0 + \mu(t)$, $\eta = \epsilon/\sqrt{r}$ and $Z(t) = (Y_i - \bar{Y})(t)$, $f(t) = a_i + \gamma_i(t)$, $\eta = \epsilon\sqrt{(r-1)/r}$ for the models (14) and (15), respectively.

In both cases, under the null hypotheses (5) and (7), f is a constant function though for the latter case, the composite null hypothesis contains r constraints of this type. Thus, our goal is to derive an optimal test for testing

$$H_0 : f \equiv \text{constant} \quad \text{versus} \quad H_1 : \left(f - \int_0^1 f(t)dt \right) \in \mathcal{F}(\rho),$$

in the general model (16), where $\mathcal{F}(\rho) = \{f \in B_{p,q}^s(C) : \|f\|_2 \geq \rho\}$ and, obviously, $\text{constant} = \int_0^1 f(t)dt$.

Choose a mother wavelet ψ of regularity $r > s$. Performing the periodic wavelet transform (see Section 2.3) on (16), one has

$$Y_{jk} = \theta_{jk} + \eta \xi_{jk}, \quad j \geq -1; \quad k = 0, 1, \dots, 2^j - 1, \quad (17)$$

where $Y_{jk} = \int_0^1 \psi_{jk}(t)dZ(t)$, $\theta_{jk} = \int_0^1 \psi_{jk}(t)f(t)dt$ and ξ_{jk} are independent $N(0, 1)$ random variables. Note that since $\int \psi_{jk}(t)dt = 0$ for all $j \geq 0$ and all $k = 0, 1, \dots, 2^j - 1$, under the null hypothesis $H_0 : f \equiv \text{constant}$, the only possibly nonzero coefficient of the wavelet decomposition of f is the scaling coefficient $\theta_{-10} = \int_0^1 f(t)dt$. Therefore, testing

$$H_0 : f \equiv \text{constant}$$

is equivalent to testing

$$H_0 : \theta_{jk} = 0 \quad \forall j \geq 0; \quad k = 0, 1, \dots, 2^j - 1.$$

The above testing problem differs from $H_0 : f \equiv 0$ in (11) studied by Spokoiny (1996) in a Gaussian “signal plus noise” model only by removing the requirement $\theta_{-10} = 0$ under the null hypothesis. Obviously this only difference does not affect the asymptotic properties of the resulting functional hypothesis testing procedures and we can, therefore, apply the results obtained by Spokoiny (1996) to develop asymptotically optimal (minimax) tests in the FANOVA model (1)-(2).

To apply the results of Spokoiny (1996), we assume that the parameters of the corresponding Besov ball $B_{p,q}^s(C)$ satisfy $1 \leq p, q \leq \infty$, $sp > 1$ and $s - \frac{1}{2p} + \frac{1}{4} > 0$. Such assumptions are common in wavelet function estimation (e.g., Donoho *et al.*, 1995; Donoho & Johnstone, 1998). Hence, summarizing, we consider the general white noise model (16) and want to test

$$H_0 : f \equiv \text{constant} \left(= \int_0^1 f(t)dt \right) \quad \text{versus} \quad H_1 : \left(f - \int_0^1 f(t)dt \right) \in \mathcal{F}(\rho),$$

where $\mathcal{F}(\rho) = \{f \in B_{p,q}^s(C) : \|f\|_2 \geq \rho\}$, $1 \leq p, q \leq \infty$, $sp > 1$ and $s - \frac{1}{2p} + \frac{1}{4} > 0$.

3.2 MINIMAX TESTS

NON-ADAPTIVE MINIMAX TEST. Consider first the case when all the parameters s , p , q and the radius C of the corresponding Besov ball $B_{p,q}^s(C)$ are known. Let $s'' = \min\{s, s - \frac{1}{2p} + \frac{1}{4}\}$. Note that, when $p \geq 2$, the condition $sp > 1$ leads to $s''p > 1$ while, for $1 \leq p < 2$, the condition $sp > 1$ implies $s''p > \frac{3}{4}$. Under such conditions the asymptotically optimal (minimax) rate of testing in (3.1) is given by

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

(see Spokoiny, 1996, Theorem 2.1).

Now we construct the test that achieves this optimal rate. The resulting test statistic has a clear intuitive meaning. Note that in terms of the wavelet coefficients θ_{jk} , the null hypothesis $H_0 : f \equiv \text{constant}$ can be re-formulated as $H_0 : \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} \theta_{jk}^2 = 0$ and the test statistic will be essentially based on the sum of squares of the thresholded empirical wavelet coefficients Y_{jk} with properly chosen level-dependent thresholds. The null hypothesis will then be rejected when this sum of squares is greater than some critical value.

Let j_η be the largest possible integer such that $j_\eta \leq \log_2 \eta^{-2}$. In fact, asymptotically we can assume that $j_\eta = \log_2 \eta^{-2}$. Let also $j(s)$ be the resolution level given by

$$j(s) = \frac{2}{4s''+1} \log_2 (C\eta^{-2}).$$

Here we assume that the right-hand sides of the above expressions are integers; otherwise we take the integer parts of these quantities. Note that, for any admissible value of s , $j(s) < j_\eta$ and that $j(s), j_\eta \rightarrow \infty$ as $\eta \rightarrow 0$. Let $\mathcal{J} = \mathcal{J}_- \cup \mathcal{J}_+$ where \mathcal{J}_- is the set of resolution levels below $j(s)$ and \mathcal{J}_+ is the set of resolution levels between $j(s)$ and $j_\eta - 1$, i.e.

$$\mathcal{J}_- = \{0, \dots, j(s) - 1\}, \quad \mathcal{J}_+ = \{j(s), \dots, j_\eta - 1\}.$$

For each $j \in \mathcal{J}_-$, define S_j as

$$S_j = \sum_{k=0}^{2^j-1} (Y_{jk}^2 - \eta^2) \quad (18)$$

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

$$S_j(\lambda) = \sum_{k=0}^{2^j-1} [(Y_{jk}^2 \mathbf{1}(|Y_{jk}| > \eta\lambda) - \eta^2 b(\lambda)], \quad (19)$$

where $\mathbf{1}(A)$ is the indicator function of the set A , $b(\lambda) = \mathbb{E} [\xi^2 \mathbf{1}(|\xi| > \lambda)]$ and ξ is a $N(0, 1)$ random variable. Note that the terms in $S_j(\lambda)$ for $j \in \mathcal{J}_+$ are defined by applying hard thresholding on the empirical wavelet coefficients, which is a standard procedure for minimax wavelet estimation in nonparametric regression settings (e.g., Donoho *et al.*, 1995). The thresholded coefficients are then centered to imply $ES_j(\lambda) = 0$ under H_0 ; note that no thresholding is needed on coarse levels $j \in \mathcal{J}_-$ and that $ES_j = 0$ under H_0 .

With the above notation, introduce the following test statistics

$$T(j(s)) = \sum_{j=0}^{j(s)-1} S_j, \quad (20)$$

and

$$Q(j(s)) = \sum_{j=j(s)}^{j_\eta-1} S_j(\lambda_j), \quad (21)$$

where $\lambda_j = 4\sqrt{(j - j(s) + 8) \ln 2}$. Let also $v_0^2(j(s))$ and $w_0^2(j(s))$ be the variances of $T(j(s))$ and $Q(j(s))$, respectively, under H_0 . It is easy to see that

$$v_0^2(j(s)) = 2\eta^4 2^{j(s)} \quad \text{and} \quad w_0^2(j(s)) = \eta^4 \sum_{j=j(s)}^{j_\eta-1} 2^j d(\lambda_j),$$

where $d(\lambda_j) = \mathbb{E} [\xi^4 \mathbf{1}(|\xi| > \lambda_j)]$. Using the results of Fan (1996), one gets the following asymptotic expressions

$$\mathbb{E}(\xi^{2k} \mathbf{1}(|\xi| > \lambda_j)) = \sqrt{2/\pi} \lambda_j^{2k-1} 2^{-8(j-j(s)+8)} + O\left(\lambda_j^{2k-3} 2^{-8(j-j(s)+8)}\right), \quad k = 1, 2, \dots \quad (22)$$

However, such an approximation might be poor in finite sample examples and result in too low values of $d(\lambda_j)$ and $b(\lambda_j)$. A better approximation for $d(\lambda_j)$ is obtained by using the Taylor expansions for $\lambda_j \leq 1$ and the Laurent expansions for $\lambda_j > 1$ respectively

$$d(\lambda_j) = 3 - \sqrt{2/\pi}\Lambda_j^5/5 + \Lambda_j^7/(7\sqrt{2\pi}) + o(\Lambda_j^8), \quad (23)$$

where $\Lambda_j = \min(\lambda_j, 1/\lambda_j)$. Similar approximations can be derived for $b(\lambda_j) = \mathbb{E}(\xi^2 \mathbf{1}(|\xi| > \lambda_j))$.

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

$$\phi^* = \begin{cases} \mathbf{1} \{T(j(s)) > v_0(j(s))z_{1-\alpha}\} & \text{if } p \geq 2 \\ \mathbf{1} \left\{ T(j(s)) + Q(j(s)) > \sqrt{v_0^2(j(s)) + w_0^2(j(s))}z_{1-\alpha} \right\} & \text{if } 1 \leq p < 2, \end{cases} \quad (24)$$

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

The following proposition, whose proof is given in the Appendix, establishes the asymptotic optimality of ϕ^* .

Proposition 3.1 *Let the mother wavelet ψ be of regularity $r > s$, and let the parameters s, p, q and the radius C of the Besov ball $B_{p,q}^s(C)$ be known, where $1 \leq p, q \leq \infty$, $sp > 1$, $s - \frac{1}{2p} + \frac{1}{4} > 0$ and $C > 0$. Then, for a fixed significance level $\alpha \in (0, 1)$, the test ϕ^* , defined in (24), for testing*

$$H_0 : f \equiv \text{constant} \left(= \int_0^1 f(t)dt \right) \quad \text{versus} \quad H_1 : \left(f - \int_0^1 f(t)dt \right) \in \mathcal{F}(\rho),$$

where $\mathcal{F}(\rho) = \{f \in B_{p,q}^s(C) : \|f\|_2 \geq \rho\}$, is level- α asymptotically optimal (minimax) test, as $\eta \rightarrow 0$. That is, for any $\beta \in (0, 1)$, it attains the optimal rate of testing

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

where $s'' = \min\{s, s - \frac{1}{2p} + \frac{1}{4}\}$.

Remark 3.1 For $p \geq 2$, the test defined in (24) differs from that developed by Spokoiny (1996), who proposed to perform *nonlinear* thresholding for all $j \geq j(s)$ regardless of p . Instead, since $p \geq 2$ corresponds to ‘‘spatially homogeneous’’ functions whose wavelet coefficients are concentrated on coarse resolution levels, we suggest simply to truncate wavelet series at level $j(s) - 1$ in this case. These results are similar to those known in nonparametric quadratic functional estimation where *linear* estimators are still optimal for $p \geq 2$ and $sp > 1$ (e.g., Efremovich & Low, 1996).

ADAPTIVE MINIMAX TEST. The structure of the rate-optimal tests developed above essentially relies on the smoothness of the underlying function via the parameters s and p of the corresponding Besov ball while such kind of prior information is typically lacking in practical applications. Our aim now is to develop an *adaptive* functional hypothesis testing procedure that does not require the knowledge of s and p and, moreover, achieves an optimal testing rate up to an unavoidable log-log factor (see Section 2.2).

We consider now the case when the parameters s , p , q and the radius C of the corresponding Besov ball $B_{p,q}^s(C)$ are unknown but assume that $0 < s \leq s_{\max}$, $1 \leq p, q \leq \infty$, $sp > 1$, $s - \frac{1}{2p} + \frac{1}{4} > 0$ and $0 < C \leq C_{\max}$. The corresponding range of these parameters will be denoted by \mathcal{T} . Let $t_\eta = (\ln \ln \eta^{-2})^{1/4}$ and $j_{\min} = \frac{2}{4s_{\max}+1} \log_2 \eta^{-2}$. Suppose now that the mother wavelet ψ is of regularity $r > s_{\max}$.

The idea of adaptive test is to consider the whole possible range of $j(s) = j_{\min}, \dots, j_\eta - 1$ and reject H_0 if it is rejected at least for one level $j(s)$. Since $\text{card}(\{j_{\min}, \dots, j_\eta - 1\}) = O(\ln \eta^{-2})$, performing a Bonferroni type testing leads to the following asymptotic *adaptive* test

$$\phi_\eta^* = \mathbf{1} \left[\max_{j_{\min} \leq j(s) \leq j_\eta - 1} \left\{ \frac{T(j(s)) + Q(j(s))}{\sqrt{v_0^2(j(s)) + w_0^2(j(s))}} \right\} > \sqrt{2 \ln \ln \eta^{-2}} \right]. \quad (25)$$

Spokoiny (1996) showed that the test ϕ_η^* defined in (25) is an adaptive optimal test, i.e. such that

$$\alpha(\phi_\eta^*) = o_\eta(1)$$

and

$$\sup_{\mathcal{T}} \beta(\phi_\eta^*, c\rho(\eta t_\eta)) = o_\eta(1),$$

where $\rho(\eta t_\eta) = (\eta t_\eta)^{4s''/(4s''+1)}$, $o_\eta(1)$ is a sequence tending to zero as $\eta \rightarrow 0$ and c is a constant. If, in addition, it is known that $p \geq 2$ then, similar to (24), the above adaptive test defined in (25) can be simplified as

$$\phi_\eta^* = \mathbf{1} \left[\max_{j_{\min} \leq j(s) \leq j_\eta - 1} \left\{ \frac{T(j(s))}{\sqrt{v_0^2(j(s))}} \right\} > \sqrt{2 \ln \ln \eta^{-2}} \right]. \quad (26)$$

The proof of this assertion is given in the Appendix.

In fact, the test statistic of ϕ_η^* in (25) is similar in spirit to that used in Fan (1996) and Fan & Lin (2000) though they apply a certain *global* threshold and do not discuss the optimality of their testing procedure.

We finish this section by the following remarks.

Remark 3.2 The results obtained in this section remain true if *different* (sufficiently regular) mother wavelets are used for $\mu(t)$ and different $\gamma_i(t)$ ($i = 1, \dots, r$).

Remark 3.3 The extension of the above non-adaptive and adaptive tests to the case $d > 1$ is straightforward using the d -dimensional periodic wavelet transform. Note that, in this case, the optimal rate of testing is $\rho(\eta) = \eta^{4s''/(4s''+d)}$ and the additional factor t_η for the adaptive test remains the same (e.g., Horowitz & Spokoiny, 2001). It is easy to show that in order to achieve this optimal rate, one should perform essentially the same procedures as for $d = 1$, but based on the empirical wavelet coefficients of a d -dimensional periodic wavelet transform with similar test statistics within an appropriate resolution range for $j(s)$. For example, in the non-adaptive case one should consider the resolution level $j(s) = 2(4s'' + d)^{-1} \log_2(C\eta^{-2})$ while, in the adaptive case one should consider the resolution range $j(s) = 2(4s_{\max} + d)^{-1} \log_2(\eta^{-2}), \dots, d^{-1} \log_2(\eta^{-2}) - 1$, where $s'' = \min\{s, s - \frac{d}{2p} + \frac{d}{4}\}$, $sp > d$ and $s - \frac{d}{2p} + \frac{d}{4} > 0$.

APPLICATIONS TO FUNCTIONAL ANALYSIS OF VARIANCE MODELS. Here we apply the derived functional hypothesis testing procedures (24), (25) and (26) to the FANOVA model (1)-(4).

To test the main effect $H_0 : \mu \equiv 0$ (no global trend), we apply (24), (25) and (26) directly, using the average process \bar{Y} defined in (14) as Z in (16), and setting $\eta = \epsilon/\sqrt{r}$.

To test the interactions (no differences in shape), note first that testing

$$H_0 : \gamma_i \equiv 0, \quad \forall i = 1, \dots, r,$$

is equivalent to testing

$$H_0 : \theta_{jk}^{(i)} = 0, \quad \forall i = 1, \dots, r, \quad j \geq 0; \quad k = 0, \dots, 2^j - 1,$$

where $\theta_{jk}^{(i)}$ are the wavelet coefficients of γ_i . To apply (24), (25) and (26), define

$$S_j^{(i)} = \sum_{k=0}^{2^j-1} [(Y_{jk}^{(i)})^2 - \eta^2]$$

for each $j \in \mathcal{J}_-$, and, for each $j \in \mathcal{J}_+$,

$$S_j^{(i)}(\lambda_j) = \sum_{k=0}^{2^j-1} [(\{Y_{jk}^{(i)}\})^2 \mathbf{1}(|Y_{j,k}^{(i)}| > \eta\lambda_j) - \eta^2 b(\lambda_j)],$$

where $Y_{jk}^{(i)}$ are the empirical wavelet coefficients of $d(Y_i - \bar{Y})(t)$ in (15), $\eta = \epsilon\sqrt{(r-1)/r}$ and $\lambda_j = 4\sqrt{(j - j(s) + 8) \ln 2}$.

Define now

$$T(j(s)) = \sum_{i=1}^r T^{(i)}(j(s)) = \sum_{i=1}^r \sum_{j=0}^{j(s)-1} S_j^{(i)}$$

and

$$Q(j(s)) = \sum_{i=1}^r Q^{(i)}(j(s)) = \sum_{i=1}^r \sum_{j=0}^{j(s)-1} S_j^{(i)}(\lambda_j).$$

Although both $T^{(i)}(j(s))$ and $Q^{(i)}(j(s))$ are correlated for different i (see comments after (15)), we have

$$\text{Var}(T(j(s))) \leq r^2 \text{Var}(T^{(i)}(j(s))) = 2r^2\eta^4 2^{j(s)}$$

and

$$\text{Var}(Q(j(s))) \leq r^2 \text{Var}(Q^{(i)}(j(s))) = r^2\eta^4 \sum_{j=j(s)}^{j_\eta-1} 2^j d(\lambda_j).$$

We can therefore apply (24), (25) and (26) with

$$v_0^2(j(s)) = 2r^2\eta^4 2^{j(s)} \quad \text{and} \quad w_0^2(j(s)) = r^2\eta^4 \sum_{j=j(s)}^{j_\eta-1} 2^j d(\lambda_j).$$

The additional r^2 factor does not depend on η and does not affect the asymptotic properties of the proposed tests (see Appendix).

Finally, it is important to point out again that in practice one always deals with *discrete* data and, therefore, should apply the *sampled* versions of the tests proposed in Section 3.2 to the empirical wavelet coefficients obtained by the *discrete* periodic wavelet transforms (e.g., Vidakovic, 1999, Chapter 5.6). To justify the optimality of the sampled versions of the tests we can exploit the general asymptotic equivalence results between the Gaussian white noise model and the corresponding Gaussian nonparametric regression setting with equispaced design and variance $\sigma^2 = n\epsilon^2$ (see Brown & Low, 1996). Although the proposed methodology is simple and powerful, the practical use of discrete wavelet transforms requires the sample size to be a power of two and the equispaced design. The first limitation is usually avoided in practice by appending the data by periodic extension or symmetric reflection. Another possible approach is to use alternatives to the fast discrete wavelet transform for vectors of arbitrary length (see Antoniadis & Pham, 1998). The requirement of the

equispaced design is more difficult to handle though, fortunately, in a wide set of signal processing (e.g., EEG, seismic or acoustic signals) or image processing (pixels) applications involving functional data, one usually has an equispaced design. Nevertheless, this requirement can be also relaxed by mapping first the original data to equispaced points by either binning or interpolation methods. Such a preprocessing yields optimal wavelet estimators for the nonparametric regression setting (see Cai & Brown, 1998), while the asymptotic equivalence with the corresponding white noise model is established in Brown *et al.* (2002). We believe that similar results can be obtained within the testing framework as well but a rigorous analysis is beyond the scope of the present paper.

Furthermore, in most applications, σ is unknown and its estimation is crucial for the success of the functional hypothesis testing procedures described above. In wavelet function estimation, the common practice is to robustly estimate σ by the median of absolute deviation of the empirical wavelet coefficients of the data at the highest resolution level divided by 0.6745 (see Donoho & Johnstone, 1995). In model (1), unless there are replications, we estimate σ by averaging its r robust estimates obtained from each individual data curve. Note that in this case the estimate of σ is *independent* from the test statistics (24)-(26) that do not involve empirical wavelet coefficients from the finest level. In the one dimensional case and for smooth alternatives one could also use the nonparametric estimators described in Hall *et al.* (1990, 1991) and Dette *et al.* (1998), but their methods are not easy to extend to higher dimensional settings.

4. NUMERICAL EXAMPLES

The purpose of this section is to shade some light on the theoretical results discussed in Section 3. First, we carry out a small simulation study to investigate the finite sample performance of the proposed functional hypothesis testing procedures. Then, as an illustration, we apply these tests to a real-life data example arising from physiology.

The computational algorithms related to wavelet analysis were performed using Version 8 of the WaveLab toolbox for MATLAB. The entire study was carried out using the MATLAB programming environment.

4.1 SIMULATION STUDY

The simulation study is based on synthetic data constructed by using the battery of standard test functions of Donoho & Johnstone (1995) frequently used in wavelet benchmarking, namely the BLOCKS, BUMPS, DOPPLER and HEAVISINE functions. We added the additional test function MISHMASH, defined as

$$\text{MISHMASH} = -(\text{BLOCKS} + \text{BUMPS} + \text{DOPPLER} + \text{HEAVISINE}),$$

to satisfy the first part of the identifiability condition (4), i.e., to ensure that the sum of all functions is zero at each point.

The observations are simulated as discretized versions of equations (1)-(2), satisfying discretized identifiability conditions (3)-(4). At n equispaced time points $t_j = j/n$, $j = 1, \dots, n$, the data are generated as multivariate vectors $y_i(t_j)$, $i = 1, \dots, 5$ defined as the sum of a constant $m_0 = 1$, the mean function $\mu(t) = 2 \sin(2\pi t)$ (the main effect of t), a corresponding test function specified by the treatment i , and a Gaussian noise of a given size. The test functions actually represent the *group effects* and can be decomposed as $a_i + \gamma_i(t)$, where the *main group effects* a_i are the integrals of the original test functions and the *interaction components* $\gamma_i(t)$ are their *centered* versions so that $\int_0^1 \gamma_i(t) dt = 0$. The mean function $\mu(t)$ and the functions $\gamma_i(t)$ are depicted in Figure 4.1.

The size of noise is selected in accordance with the energy (or variance) of the test functions, i.e. their squared L^2 -norm. The five test functions are in addition rescaled so that for all of them a noise of size 1 gives the prescribed signal-to-noise ratio (SNR), defined as the ratio of standard deviations of the signal and of the noise. Five simulated observations (one for each test function shown in Figure 4.1) of a specific length ($n = 1024$), with two different SNRs (SNR = 3 and 7), are shown superimposed and separately in Figure 4.2.

The standard deviation σ of the error terms is assumed unknown and it was robustly estimated by the median of absolute deviation of the empirical wavelet coefficients of the data at the highest resolution level, as discussed in Section 3.2. We applied the non-adaptive functional hypothesis testing procedure discussed in Section 3.2 for the main effect $\mu(t)$ and the interactions $\gamma_i(t)$ ($i = 1, \dots, 5$) separately at the significance level $\alpha = 0.05$ for both SNRs.

To test the hypothesis $H_0 : \mu(t) = 0$, we applied the truncation version ($p \geq 2$) of the non-

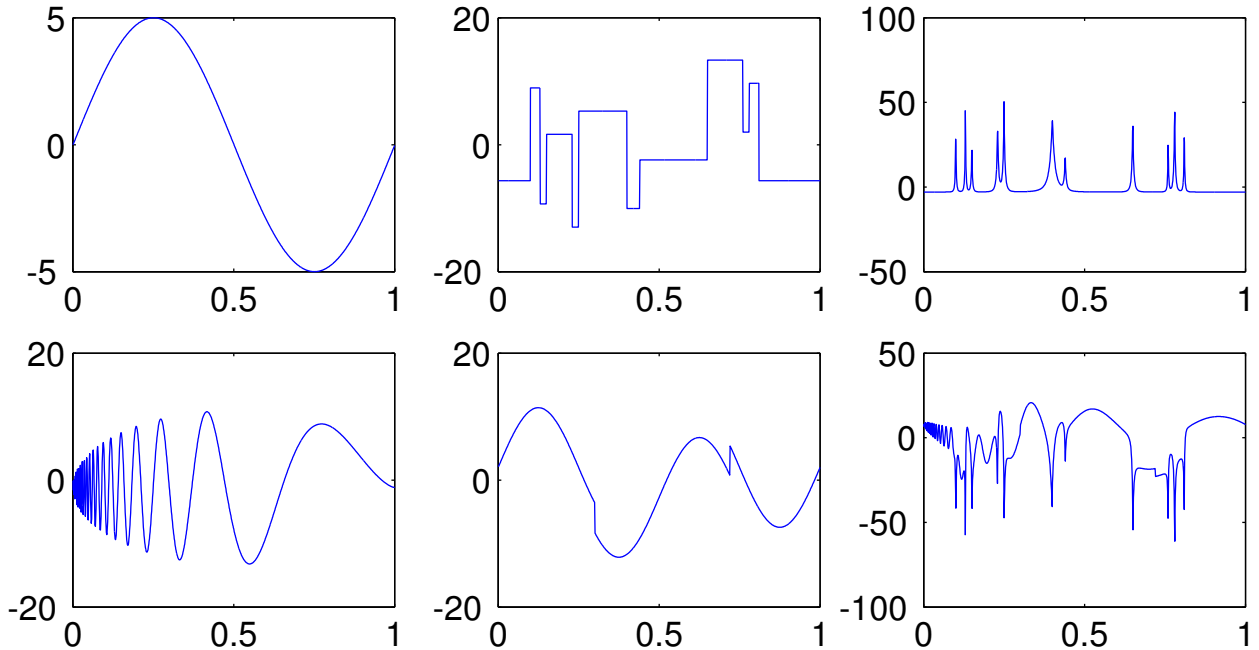


Figure 4.1: The mean function $\mu(t) = 2 \sin(2\pi t)$ and the centered treatment effect functions $\gamma_i(t)$, $i = 1, \dots, 5$ (i.e., centered BLOCKS, BUMPS, DOPPLER, HEAVISINE, and MISHMASH), sampled at $n = 1024$ data points.

adaptive functional hypothesis testing procedure defined in (24). Since μ is smooth, the compactly supported *Symmlet 8-tap filter* mother wavelet was used with $j(s) = 3$. For the case SNR=3, the value of the test statistic $T(3)$ was equal to 15.2851 to be compared with the critical value 1.5949, while for the case SNR=7, the corresponding values of the test statistic and of the critical value were 97.5164 and 1.6316 respectively.

To test the hypothesis $H_0 : \gamma_i(t) = 0$ ($i = 1, \dots, 5$), we applied the general (thresholding) version of the non-adaptive functional hypothesis testing procedure defined in (24). The compactly supported *Daubechies 6-tap filter* mother wavelet was used with $j(s) = 3$. Although more resolution levels are needed for less smooth functions γ_i , the wavelet coefficients on the seventh and higher levels could be hardly distinguished from noise, which lead us to set $j_\eta = 7$. For SNR=3, the value of the test statistic $T(3) + Q(3)$ was equal to 275.3326 to be compared with the critical value of 154.6294 while, for the case SNR=7, the corresponding values of the test statistic and of the critical

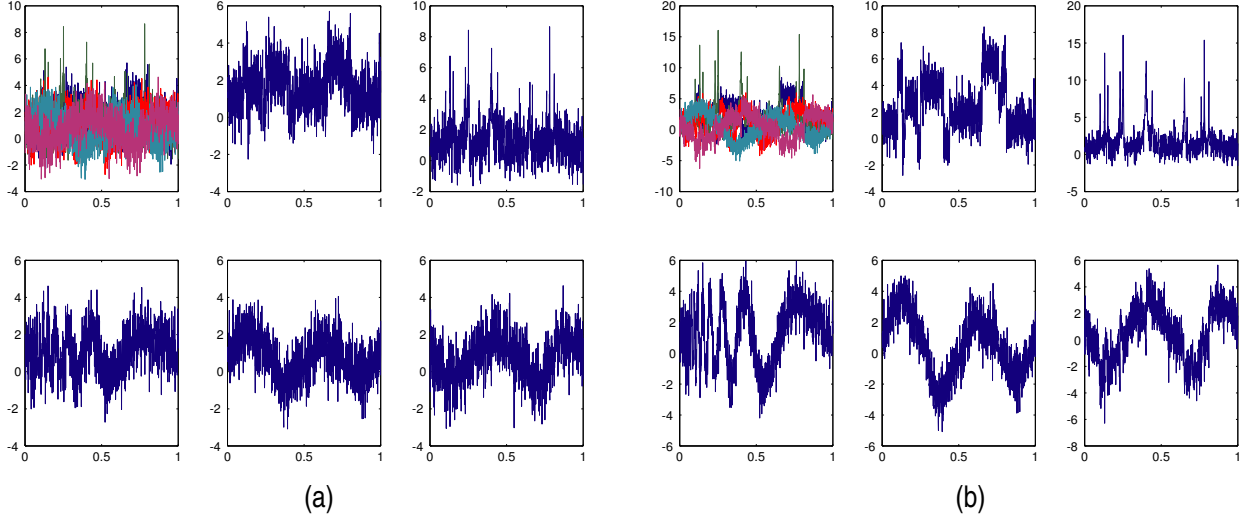


Figure 4.2: Five simulated observations (one for each test function shown in Figure 4.1) sampled at $n = 1024$ data points are shown superimposed (first plot) and separately (remaining five plots) for (a) $\text{SNR} = 3$ and (b) $\text{SNR} = 7$.

value were 5941.0998 and 156.4943 respectively.

The simulations demonstrated that some asymptotic specifications required adjustments at moderate sample sizes. For instance, the thresholds $\lambda_j = C\sqrt{(j - j(s) + 8) \ln 2}$ with $C = 4$ used in defining the test statistics in expression (21) were unreasonably high, and finite sample situations called for $C = 1$. Moreover, using the asymptotic expression (22) derived in Fan (1996), one may get extremely small values of the $Q(j(s))$. In such cases we have used the approximation (23) for $d(\lambda_j)$.

We have also performed an extensive power analysis for the above tests against the composite alternatives

$$H_1 : \mu \in \mathcal{F}(\rho) \quad \text{and} \quad H_1 : \frac{1}{5} \sum_{i=1}^5 \gamma_i \in \mathcal{F}(\rho). \quad (27)$$

We provide a brief summary and figure that illustrates the simulation. The significance levels had been fixed at $\alpha = 5\%$. For fixed ρ (set to 1), the magnitudes of the signals have been modified to achieve the prescribed SNR. The graphs of *empirical* power functions (computed on 500 replications for each test and each SNR) against the SNR are given in Figure 4.3 and demonstrate how fast the power of the proposed procedure increases with increasing of SNR or, equivalently, of the L^2 -

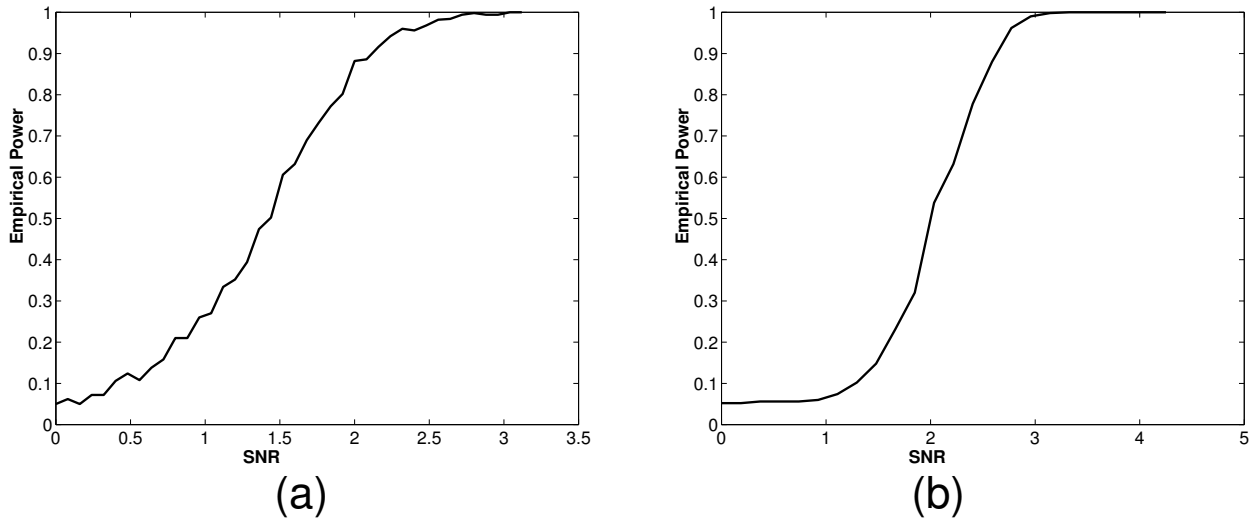


Figure 4.3: Empirical power functions for testing (a) $H_0 : \mu(t) = 0$ versus $H_1 : \|\mu\|_2 = \rho$ and (b) $H_0 : \gamma_i(t) = 0$ ($i = 1, \dots, 5$) versus $H_1 : \|\sum_i \gamma_i/5\|_2 = \rho$. In both panels, the sample size was $n = 512$ and the number of trials at a fixed discretized SNR was 500.

distance ρ between the null hypothesis and the composite alternative. The sample size of $n = 512$ was taken.

When the null hypotheses are true, the power was around the significance level of 5%. We see that the tests perform quite well. The difference in power for each type of the tested hypotheses relates to smoothness of the underlying functions. Thus, the simulation results illustrate that the proposed functional hypothesis testing procedures possess satisfactory power in the presence of specified, relatively extreme, alternatives.

4.2 ORTHOSIS DATA ANALYSIS

We have applied the proposed FANOVA methodology to some interesting data on human movement. The data were acquired and computed by Dr. Amarantini David and Dr. Martin Luc (Laboratoire Sport et Performance Motrice, EA 597, UFRAPS, Grenoble University, France). The purpose of recording such data was the interest to better understand 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. Seven young male volunteers wore a spring-loaded orthosis of adjustable stiffness under four experimental conditions: a control condition (without orthosis), an orthosis condition (with the orthosis only), and two conditions (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 seconds under each experimental condition for each subject. Data sampling started 5 seconds after the onset of stepping, and lasted for 10 seconds for each trial. So, anticipatory and joint movements induced by the initiation of the movement were not sampled. For each of the seven subjects, ten 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 Cahouet *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 time t , where $t \in [0, 1]$. The data set consists in 280 separate runs and involves the seven subjects over four described experimental conditions, replicated ten times for each subject. Figure 4.4 shows the available data set; typical moment plots over gait cycles. Since the purpose of the experiment was to understand how a subject can cope with the external perturbation, we need to quantify the ways in which the individual mean cross-sectional functions differ over the various conditions. We model the data as arising from a fixed-effects FANOVA model with 2 qualitative factors and 10 replications for each level combination, and the purpose is now to show that substantial insight can be gained by applying directly the proposed functional hypothesis testing procedures. The appropriate model for the available data set is a block design, written as

$$dY_{ijk}(t) = m_{ij}(t) dt + \epsilon dW_{ijk}(t), \quad i = 1, \dots, I; \quad j = 1, \dots, J; \quad k = 1, \dots, K; \quad t \in [0, 1],$$

with

$$m_{ij}(t) = m_0 + \mu(t) + \alpha_i + \gamma_i(t) + \beta_j + \delta_j(t), \quad i = 1, \dots, I; \quad j = 1, \dots, J; \quad t \in [0, 1],$$

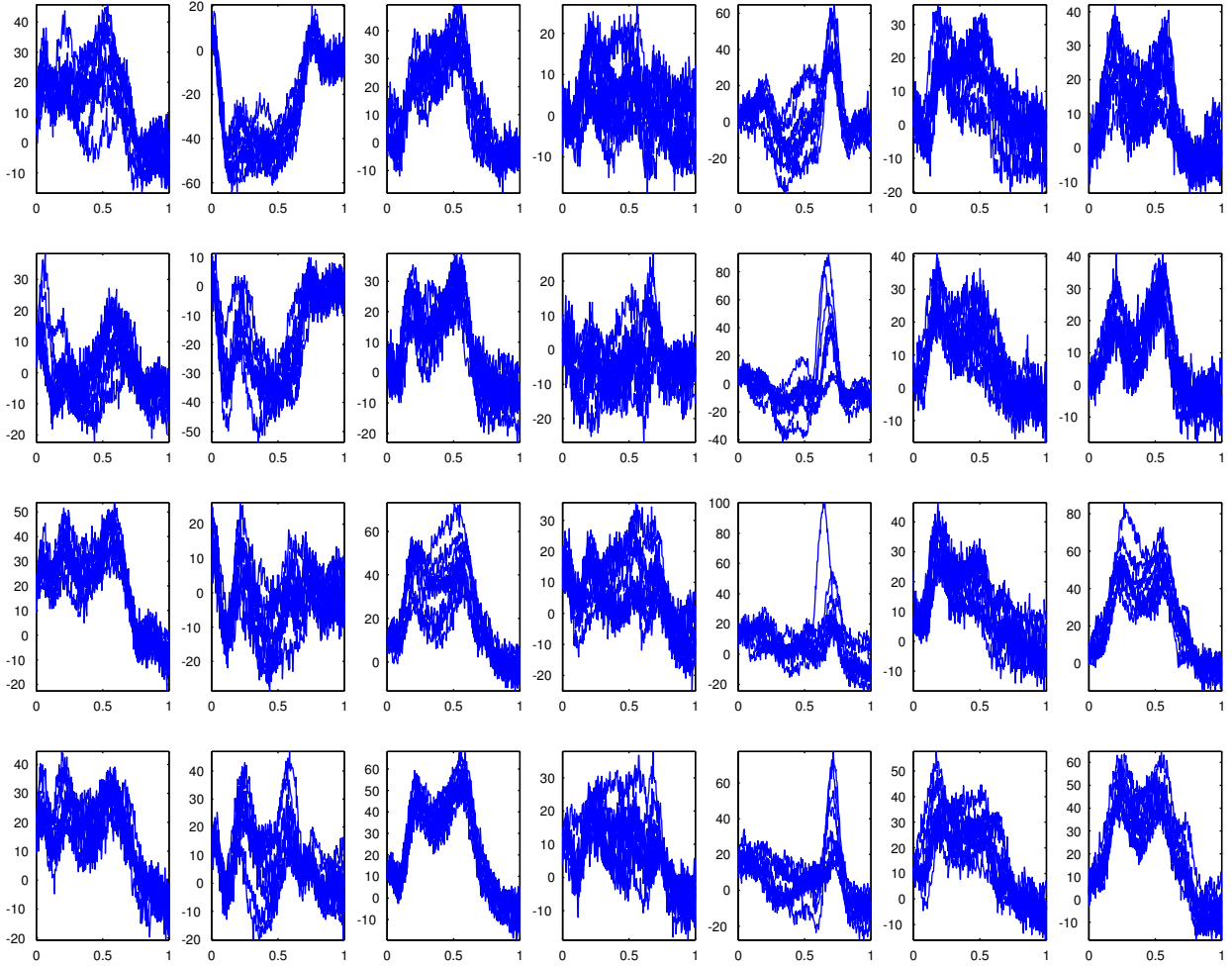


Figure 4.4: Orthosis data set: panels in rows correspond to *Treatments* while the panels in columns correspond to *Subjects*.

where i is the condition index, j is the subject index, k is the replication index and t is the time variable. Subjects in the above model are naturally considered as block effects; subjects obviously differ but the researchers are not interested in their differences. Treating subjects as blocks allows us to make inference about the treatments of interest more precise since variability due to block effects is excluded (in this example, we have $I = 4$ treatments). Therefore, given the standard identifiability conditions, equivalent to those in (3)-(4), we can now apply the proposed FANOVA methodology on the averaged model

$$d\bar{Y}_{i..}(t) = m_i(t) dt + \eta dW_{i..}(t), \quad i = 1, \dots, I; \quad t \in [0, 1],$$

with

$$m_i(t) = m_0 + \mu(t) + \alpha_i + \gamma_i(t), \quad i = 1, \dots, I; \quad t \in [0, 1],$$

where $\eta = \epsilon/\sqrt{JK}$.

Since the number of available detail levels in the wavelet decomposition is seven ($\log_2(256) - 1$), we adopted $j(s) = 4$ and $j_\eta = 6$. That is, we expect that a smooth function can be well described within the first four coarsest levels, the treatment effect functions require up to six levels and the seventh level contains the noise. For each treatment, we used the median of absolute deviation of the empirical wavelet coefficients of the empirical estimators of the corresponding treatment effect at the highest resolution level divided by 0.6745 as an estimate of ϵ/\sqrt{JK} . The final estimate of ϵ/\sqrt{JK} is the average over all treatments and found to be 0.3586. This is a sensible procedure since one can assume that the functional components $\mu(t)$ and $\gamma_i(t)$ of the averaged model are sufficiently smooth and that their presence at the finest levels of detail in wavelet decompositions is minimal. We also found that our tests are robust with respect to the selection of wavelets used. In this analysis we have used the compactly supported *Coiflet 18-tap filter* mother wavelet, motivated by their excellent compromise in smoothness, compactness, almost-symmetry, and good approximation characteristics.

We first test the hypothesis $H_0 : \mu(t) = 0$ versus $H_1 : \mu(t) \in \mathcal{F}(\rho)$ (for some $\rho > 0$). An empirical estimator of overall mean $m_0 + \mu(t)$ is shown in Figure 4.5a. It is therefore reasonable to test $H_0 : \mu(t) = 0$ with the version of the proposed test given in (24) for $p \geq 2$. The null hypothesis is rejected, the value of the statistic $T(j(s))$ being 12602.54 with p -value essentially equal to 0 (the 5%-level critical value is 0.3686). The test statistic $T(j(s)) + Q(j(s))$ for testing the null hypothesis of equality of treatment effects $H_0 : \gamma_i = 0$ ($i = 1, \dots, 4$) takes the value 13531.30 and also leads to reject H_0 with a p -value essentially equal to 0 (the 5%-level critical value is 309.62).

The researchers were also interested in testing the contrasts: (i) Control and Orthosis functional treatment effects are equal ($H_0 : \gamma_1(t) = \gamma_2(t)$), and (ii) Spring 1 and Spring 2 functional treatment effects are equal ($H_0 : \gamma_3(t) = \gamma_4(t)$). By inspecting the corresponding empirical estimators, shown in Figure 4.5b, it is again reasonable to test these hypotheses using $p \geq 2$. Both of these hypotheses are not rejected, with $T(j(s))$ equal to 158.95 and 134.03 respectively. The 5%-level critical values are 309.62 for both tests and the p -values are 0.157 and 0.198, respectively. One can therefore

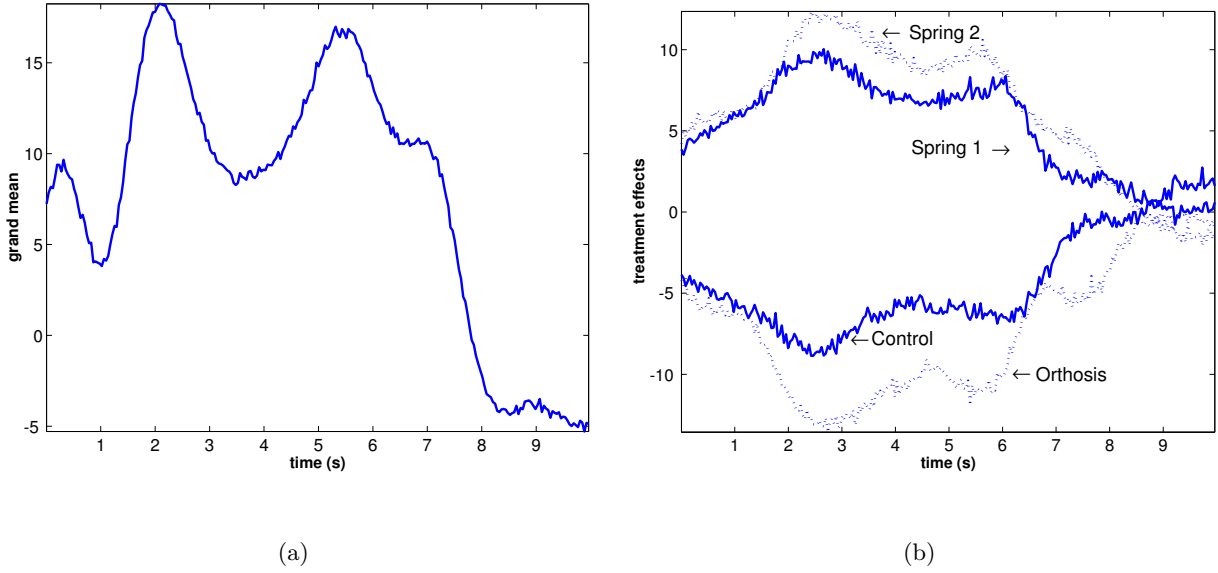


Figure 4.5: Panel (a) depicts the time domain empirical estimator of the grand mean, $m_0 + \mu(t)$. The noise is suppressed by averaging, but still of size ϵ/\sqrt{IJK} ; Panel (b) depicts empirical estimators of the treatment effects of interest. Constant and functional components α_i and $\gamma_i(t)$ ($i = 1, \dots, 4$) are not separated.

say that under Control and Orthosis conditions the subjects behave similarly, the same being true under Spring 1 and Spring 2 conditions.

By inspecting the empirical estimators of treatments, shown in Figure 4.5b, one could hypothesize that the contrast $(\gamma_1(t) + \gamma_2(t)) - (\gamma_3(t) + \gamma_4(t))$ is significant. In terms of the application, this means that the sum of Control and Orthosis differ from the sum of Spring 1 and Spring 2. The corresponding null hypothesis is highly rejected; the p -value is 0.0103, with a test statistics $T(j(s))$ equal to 46276.42 compared to a 5%-level critical value equal to 619.25.

In conclusion, we have found that all null hypotheses in this study had been rejected except the contrasts $\gamma_1 - \gamma_2 \equiv 0$ and $\gamma_3 - \gamma_4 \equiv 0$, which were not rejected at level $\alpha = 5\%$. This supports the fact that individuals adjust their posture similarly under perturbations of similar nature.

5. CONCLUDING REMARKS AND POSSIBLE EXTENSIONS

We considered the testing problem in a fixed-effects functional analysis of variance model and derived asymptotically optimal (minimax) non-adaptive and adaptive functional hypothesis testing procedures for the main effect and the interactions. The decomposition of Gaussian “signal plus noise” processes allowed one to present a wide variety of models in the same format, which facilitates the application of general nonparametric testing procedures to assess the nature of the underlying mean function. An important characteristic of the developed functional hypothesis testing methodology is that it allows one to perform a similar analysis for various types of hypotheses in fixed-effects functional analysis of variance models. Moreover, the resulting procedures are computationally inexpensive and can be easily implemented.

The proposed approach differs from other recent functional hypothesis testing procedures that treat curved data as multivariate vectors and adapt the traditional analysis of variance with various initial dimensionality-reduction techniques (e.g., Faraway, 1997; Dette & Derbort, 2001). It also differs from the spline smoothing functional analysis of variance testing in mixed-effects models developed in Guo (2002). In fact, it is much closer in spirit to the overall tests of Fan (1996) and Fan & Lin (2000). Unlike these approaches, we establish the asymptotic optimality of the proposed functional hypothesis testing procedures.

We would like to finish the paper by pointing at several possible extensions. An interesting and practically important extension of model (1) for $d = 1$ is a model of the form

$$dY(s, t) = (m_0 + a(s) + \mu(t) + \gamma(s, t)) dt ds + \epsilon dW(s, t), \quad (s, t) \in [0, 1]^2,$$

where W is a two-dimensional standard Wiener process, i.e. when both predictors are continuous. We are interested in testing the null-hypotheses of the form: $H_0 : a \equiv 0$, $H_0 : \mu \equiv 0$, $H_0 : \gamma \equiv 0$. The latter essentially corresponds to testing the additivity of the response function $m(s, t)$. Applying the two-dimensional periodic wavelet transform to the data, a specific structure of the matrix V of the resulting empirical wavelet coefficients implies that testing $H_0 : a \equiv 0$ or $H_0 : \mu \equiv 0$ will be based only on the first row or the first column of the matrix V respectively, while the remaining (major) part of the matrix of coefficients should be used to test the interaction γ (additivity of the response). The extension of the proposed functional hypothesis testing procedures

to this case is quite straightforward.

Thresholding in (19) can be possibly performed by other methods than that of Spokoiny (1996), which was adopted in this paper. In particular, in view of recent results in quadratic functional estimation (e.g., Gayraud & Tribouley, 1999; Laurent & Massart, 2000; Johnstone, 2001) we believe it can be performed by grouping empirical wavelet coefficients within each resolution level in a block and using thresholding blockwise rather than individually. It will be also interesting to investigate how *data-driven* thresholding procedures like SURE (see Donoho & Johnstone, 1995) or FDR (see Abramovich & Benjamini, 1996), developed in the context of function estimation can be adapted within the functional hypothesis testing framework. This could improve the finite sample properties of the tests.

All the above are avenues for further research that hope will be addressed in the future.

ACKNOWLEDGEMENTS

Felix Abramovich and Anestis Antoniadis were supported by ‘Project IDOPT, INRIA-CNRS-IMAG’ and ‘Project AMOA, IMAG’. Theofanis Sapatinas was supported by a ‘University of Cyprus–2001 Research Grant’. Brani Vidakovic was supported by the NSF grant DMS 0004131 at Georgia Institute of Technology. Felix Abramovich and Theofanis Sapatinas would like to thank Anestis Antoniadis for excellent hospitality while visiting Grenoble to carry out this work.

We are very grateful to the Editor, Associate Editor and the two anonymous referees for their constructive comments that led to a significant improvement of the paper.

APPENDIX

Here we prove Proposition 3.1 and the assertion (26).

PROOF OF PROPOSITION 3.1 For $1 \leq p < 2$ the proof follows directly from Spokoiny (1996). Thus, we only consider here the case $p \geq 2$. Under the assumption $sp > 1$, $s'' = s$ and the condition $s - \frac{1}{2p} + \frac{1}{4} > 0$ holds automatically. Regardless of the true hypothesis, one always has

$$e(j(s)) = \mathbb{E}(T(j(s))) = \|P_{V_{j(s)}}(f - \int_0^1 f(t)dt)\|_2^2,$$

where $P_{V_{j(s)}}$ denotes the orthogonal projector onto the approximation space $V_{j(s)}$ of the multiresolution analysis (e.g., Mallat, 1989). Using standard results for non-central chi-squared distributions (e.g., Johnstone, 2001) we have

$$v^2(j(s)) = \text{Var}(T(j(s))) = 2\eta^4 2^{j(s)} + 4\eta^2 e(j(s)). \quad (28)$$

The test statistic $T(j(s))$ is a sum of $j(s)$ independent, squared-integrable random variables and, since $j(s) \rightarrow \infty$ as $\eta \rightarrow 0$, by the central limit theorem, $T(j(s))$ is asymptotically normal. Moreover, note that when the null hypothesis is true, $e(j(s)) = 0$ and $v^2(j(s)) = v_0^2(j(s))$, and therefore, the test ϕ^* given in (24) is asymptotically of significance level α .

Let $\beta > 0$, denote by $\beta(\phi^*, f) = P_f(\phi^* = 0)$ the Type II error, and let

$$\beta(\phi^*, \rho) = \sup_{(f - \int_0^1 f(t)dt) \in \mathcal{F}(\rho)} \beta(\phi^*, f)$$

be the probability of Type II error for the composite alternative $H_1 : (f - \int_0^1 f(t)dt) \in \mathcal{F}(\rho)$. It is straightforward to see that, for any specific f within the alternative, one has

$$\beta(\phi^*, f) = \Phi \left(\frac{v_0(j(s))}{v(j(s))} z_{1-\alpha} - \frac{e(j(s))}{v(j(s))} \right),$$

where Φ is the cumulative distribution function of a $N(0, 1)$ random variable. Set $\kappa(j(s)) = \frac{v_0(j(s))}{v(j(s))}$. Since $v(j(s)) \geq v_0(j(s))$, $\kappa(j(s))$ is bounded above by 1. Hence, as $\eta \rightarrow 0$, the asymptotic behavior of $\beta(\phi^*, f)$ depends only on the ratio of squared bias to standard deviation, $\frac{e(j(s))}{v(j(s))}$. Since $\mathcal{F}(\rho)$ belongs to the Besov ball $B_{p,q}^s(C)$ with $p \geq 2$ and $sp > 1$,

$$\sum_{j=j(s)}^{\infty} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \leq c_0 2^{-2sj(s)},$$

for some constant c_0 (e.g., Meyer, 1992) and, therefore, for any f within the alternative set

$$e(j(s)) \geq (\|f\|_2^2 - c_0 2^{-2sj(s)}) \geq (\rho^2 - c_0 2^{-2sj(s)}).$$

From (28) one has

$$v^2(j(s)) \geq 2^{j(s)+1} \eta^4 + 4\eta^2 (\rho^2 - c_0 2^{-2sj(s)}).$$

Thus, for $j(s) = \frac{2}{4s+1} \log_2(C\eta^{-2})$ and the optimal rate of testing $\rho(\eta) = \eta^{4s/(4s+1)}$, one can verify that there exists a constant c_β such that

$$\lim_{\eta \rightarrow 0} \inf_{(f - \int_0^1 f(t)dt) \in \mathcal{F}(c_\beta \rho(\eta))} \frac{e(j(s))}{v(j(s))} > \tilde{c}_\beta,$$

where $\tilde{c}_\beta > 0$ satisfies $\Phi(z_{1-\alpha} - \tilde{c}_\beta) = \beta$ and, hence, $\tilde{c}_\beta = z_{1-\alpha} + z_{1-\beta}$. This shows that the test ϕ^* is indeed asymptotically minimax.

PROOF OF THE ASSERTION (26) To prove that the test (26) is asymptotically adaptive minimax and uniformly consistent, we need to show that

$$\alpha(\phi_\eta^*) = o_\eta(1),$$

and

$$\sup_{\mathcal{T}} \beta(\phi_\eta^*, c\rho(\eta t_\eta)) = o_\eta(1),$$

for some constant $c > 0$. As we have mentioned in the proof of Proposition 3.1, under the null hypothesis, for every $j(s) = j_{\min}, \dots, j_\eta - 1$, $T(j(s))/v_0(j(s))$ are asymptotically $N(0, 1)$ random variables (though dependent) and applying the well known extreme value results for Gaussian random variables (e.g., Leadbetter, *et al.*, 1986) we have

$$\alpha(\phi_\eta^*) = P_{H_0} \left\{ \max_{j_{\min} \leq j(s) \leq j_\eta - 1} \left\{ \frac{T(j(s))}{\sqrt{v_0^2(j(s))}} \right\} > \sqrt{2 \ln \ln \eta^{-2}} \right\} \rightarrow 0, \quad \text{as } \eta \rightarrow 0.$$

Choose now any set of parameters $(s, p, q, C) \in \mathcal{T}$. Note that $\frac{1}{p} < s < s_{\max}$. For the chosen set define $j^*(s)$ by

$$2^{-j^*(s)} = (\eta t_\eta)^{4/(4s+1)}.$$

Then for any f within the alternative set we have

$$\begin{aligned} P_f \left\{ \max_{j_{\min} \leq j(s) \leq j_\eta - 1} \left\{ \frac{T(j(s))}{\sqrt{v_0^2(j(s))}} \right\} \leq \sqrt{2 \ln \ln \eta^{-2}} \right\} &\leq P_f \left\{ \frac{T(j^*(s))}{\sqrt{v_0^2(j^*(s))}} \leq \sqrt{2 \ln \ln \eta^{-2}} \right\} \\ &\leq \Phi \left(\sqrt{2 \ln \ln \eta^{-2}} - \frac{e(j^*(s))}{v(j^*(s))} \right) \end{aligned} \quad (29)$$

Repeating the arguments of the proof of Proposition 3.1 and substituting $c\rho(\eta t_\eta)$ and $j^*(s)$ in (29), the straightforward calculus yields

$$\frac{e(j^*(s))}{v(j^*(s))} = O(t_\eta^2) = O\left(\sqrt{2 \ln \ln \eta^{-2}}\right), \quad (30)$$

where one can always find a constant c such that the ratio of squared bias to standard deviation in (30) is larger than $\sqrt{2 \ln \ln \eta^{-2}}$. Thus, for this c , the probability of Type II error in (29) will tend to zero for any f and any specific set of parameters within \mathcal{T} .

Finally, note that the above proofs still hold for $v_0^2(j(s))$ and $v^2(j(s))$ multiplied by r^2 that appears in testing the interaction component (see Section 3.2.3).

REFERENCES

- Abramovich, F., Bailey, T.C. & Sapatinas, T. (2000). Wavelet analysis and its statistical applications. *The Statistician*, **49**, 1–29.
- Abramovich, F. & Benjamini, Y. (1996). Adaptive thresholding of wavelet coefficients. *Comput. Statist. Data Anal.*, **22**, 351–361.
- Antoniadis, A. (1984). Analysis of variance on function spaces. *Math. Operationsforsch. u. Statist., Ser. Statist.*, **15**, 59–71.
- Antoniadis, A. (1997). Wavelets in statistics: a review (with discussion). *J. Ital. Statist. Soc.*, **6**, 97–144.
- Antoniadis, A. & Pham, D.T. (1998). Wavelet regression for random or irregular design. *Computat. Statist. Data Anal.*, **28**, 353–369.
- Brown, L.D. & Low, M.G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.*, **24**, 2384–2398.
- Brown, L.D., Cai, T.T., Low, M.G & Zhang, C.H. (2002). Asymptotic equivalence theory for nonparametric regression with random design. *Ann. Statist.*, **30**, 688–707.
- Cahouet, V., Martin, L. & Amarantini, D. (2002). Static optimal estimation of joint accelerations for inverse dynamic problem solution. *Journal of Biomechanics*, **35**, 1507–1513.
- Cai, T.T. & Brown, L.D. (1998). Wavelet shrinkage for nonequispaced samples. *Ann. Statist.*, **26**, 1783–1799.
- Chen, J.C. (1994). Testing for no effect in nonparametric regression via spline smoothing techniques. *Ann. Inst. Statist. Math.*, **45**, 251–265.
- Daubechies, I. (1992). *Ten Lectures on Wavelets*. Philadelphia: SIAM.
- Dette, H. & Derbort, S. (2001). Analysis of variance in nonparametric regression models. *J. Multivariate Anal.*, **76**, 110–137.
- Dette, H., Munk, A. & Wagner, T. (1998). Estimating the variance in nonparametric regression – what is a reasonable choice? *J. R. Statist. Soc. B*, **60**, 751–764.

- Donoho, D.L. & Johnstone, I.M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Statist. Assoc.*, **90**, 1200–1224.
- Donoho, D.L. & Johnstone, I.M. (1998). Minimax estimation via wavelet shrinkage. *Ann. Statist.*, **26**, 879–921.
- Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. & Picard, D. (1995). Wavelet shrinkage: asymptopia? (with discussion). *J. R. Statist. Soc. B*, **57**, 301–337.
- Efromovich, S. & Low, M. (1996). On optimal adaptive estimation of a quadratic functional. *Ann. Statist.*, **24**, 1106–1125.
- Ermakov, M.S. (1990). Minimax detection of a signal in a Gaussian white noise. *Theory Probab. Appl.*, **35**, 667–679.
- Eubank, R.L. (2000). Testing for no effect by cosine series methods. *Scand. J. Statist.*, **27**, 747–763.
- Eubank, R.L. & LaRiccia, 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.
- Gayraud, G. & Tribouley, K. (1999). Wavelet methods to estimate an integrated quadratic functional: adaptivity and asymptotic law. *Statist. Probab. Lett.*, **44**, 109–122.
- Gu, C. (2002). *Smoothing spline ANOVA models*. New York: Springer-Verlag.
- Guo, W. (2002). Inference in smoothing spline analysis of variance. *J. R. Statist. Soc. B*, **64**, 887–898.
- Hall, P., Kay, J.W. & Titterton, D.M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika*, **77**, 521–528.
- Hall, P., Kay, J.W. & Titterton, D.M. (1991). On estimation of noise variance in two-dimensional processing. *Adv. Appl. Probab.*, **23**, 476–495.
- Horowitz, J.L. & Spokoiny, V.G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, **69**, 599–631.

- 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 nonparametric 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; 171–189; 249–268.
- Ingster, Yu.I. & Suslina, I.A. (2000). Minimax nonparametric hypothesis testing for ellipsoids and Besov bodies. *ESAIM: Probab. Statist.*, **4**, 53–135.
- Johnstone, I. (2001). Thresholding for weighted χ^2 . *Statist. Sinica*, **11**, 691–704.
- Laurent, B. & Massart, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.*, **28**, 1302–1338.
- Leadbetter, M.R, Lindgren, G. & Rootzén, H. (1986). *Extremes and Related Properties of Random Sequences and Processes*. New York: Springer-Verlag.
- Lepski, O.V. & Spokoiny, V.G. (1999). Minimax nonparametric hypothesis testing: the case of an inhomogeneous alternative. *Bernoulli*, **5**, 333–358.
- Lin, Y. (2000). Tensor product space ANOVA models. *Ann. Statist.*, **28**, 734–755.
- Mallat, S.G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pat. Anal. Mach. Intel.*, **11**, 674–693.
- Meyer, Y. (1992). *Wavelets and Operators*. Cambridge: Cambridge University Press.
- Ramsay, J.O. & Silverman, B.W. (1997). *Functional Data Analysis*. New York: Springer-Verlag.
- 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.
- Spokoiny, V.G. (1996). Adaptive hypothesis testing using wavelets. *Ann. Statist.*, **24**, 2477–2498.
- Spokoiny, V.G. (1998). Adaptive and spatially adaptive testing of nonparametric hypothesis hypothesis. *Math. Methods Statist.*, **7**, 245–273.
- 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.
- Vidakovic, B. (1999). *Statistical Modeling by Wavelets*. New York: John Wiley & Sons.
- Wahba, G., Wang, Y., Gu, C., Klein, R. & Klein, B. (1995). Smoothing spline ANOVA for exponen-

tial families, with application to the Wisconsin epidemiological study of diabetic retinopathy.
Ann. Statist., **23**, 1865–1895.