# Testing for additivity and joint effects in multivariate nonparametric regression using Fourier and wavelet methods

# DANIELA DE CANDITIIS\* and THEOFANIS SAPATINAS<sup>†</sup>

\*Istituto per le Applicazioni del Calcolo "Mauro Picone", Sezione di Roma, Consiglio Nazionale delle Ricerche, Italy <sup>†</sup>Department of Mathematics and Statistics, University of Cyprus, P.O. Box 20537, CY 1678, Nicosia, Cyprus T.Sapatinas@ucy.ac.cy

Received February 2003 and accepted November 2003

We consider the problem of testing for additivity and joint effects in multivariate nonparametric regression when the data are modelled as observations of an unknown response function observed on a *d*-dimensional ( $d \ge 2$ ) lattice and contaminated with additive Gaussian noise. We propose tests for additivity and joint effects, appropriate for both homogeneous and inhomogeneous response functions, using the particular structure of the data expanded in tensor product Fourier or wavelet bases studied recently by Amato and Antoniadis (2001) and Amato, Antoniadis and De Feis (2002). The corresponding tests are constructed by applying the adaptive Neyman truncation and wavelet thresholding procedures of Fan (1996), for testing a high-dimensional Gaussian mean, to the resulting empirical Fourier and wavelet coefficients. As a consequence, asymptotic normality of the proposed test statistics under the null hypothesis and lower bounds of the corresponding powers under a specific alternative are derived. We use several simulated examples to illustrate the performance of the proposed tests, and we make comparisons with other tests available in the literature.

*Keywords:* additive models, Fourier transform, hypothesis testing, joint effects, nonparametric regression, tensor product Hilbert spaces, wavelet transform

# 1. Introduction

The estimation of a multivariate response function contaminated with additive noise suffers from the so-called "curse of dimensionality". This notion reflects the fact that statistical methods in multivariate nonparametric regression estimation lose much of their power if the dimension of the response function is large, making these methods less attractive in practical applications. Additive models are one means of circumventing the "curse of dimensionality" in multivariate nonparametric regression problems by approximating the response function by a sum of univariate functions, one for each dimension. A theoretical justification for such models is that under the assumption of additivity the response function can be estimated with the same rate of estimation error as in the univariate case (see Stone 1985). In other words, multivariate nonparametric regression models remain tractable efficiently and allow for a simple interpretation when the additive structure is justified. Nowadays there is a plethora of research work on fitting additive models and estimating their components. We refer to, for example, Buja, Hastie and Tibshirani (1989), Hastie and Tibshirani (1990), Lindon and Nielsen (1995), Lindon (1997), Opsomer and Ruppert (1997, 1998), Sperlich, Linton and Härdle (1999), Amato and Antoniadis (2001), Amato, Antoniadis and De Feis (2002), Sperlich, Tjøstheim and Yang (2002), Zhang and Wong (2003) and Sardy and Tseng (2004).

Because the additive structure is important in terms of interpretability and its ability to deliver fast rates of convergence in multivariate nonparametric regression estimation, the application of additive models should be accompanied by a proper "model check of additivity". Although early work dates back to Tukey (1949), it is only recent that the problem of testing for additivity (which corresponds to the hypothesis of checking vanishing interaction terms) has been of real interest. We refer to, for example, Barry (1993), Eubank *et al.* (1995), Dette and Derbort (2001), Dette and Wilkau (2001), Derbort, Dette and Munk (2002) and Sperlich, Tjøstheim and Yang (2002). Furthermore, the problem of testing for joint effects (which corresponds to the hypothesis of checking specific vanishing main effects and interaction terms) has also been considered in Dette and Derbort (2001). However, asymptotic results and finite-sample properties of these methods have been studied under the assumption that the response function obeys a homogeneous behaviour.

In this paper we consider the problem of testing for additivity and joint effects in multivariate nonparametric regression when the data are modelled as observations of an unknown response function observed on a *d*-dimensional ( $d \ge 2$ ) lattice and contaminated with additive Gaussian noise. We propose tests for additivity and joint effects, appropriate for both homogeneous and inhomogeneous response functions, using the particular structure of the data expanded in tensor product Fourier or wavelet bases recently studied by Amato and Antoniadis (2001) and Amato, Antoniadis and De Feis (2002).

The paper is organized as follows. In Section 2 we first formulate the problem of testing for additivity and joint effects in the bivariate nonparametric regression case, for which the basic ideas of the proposed methodology are most transparent. In particular, we transform the data matrix in the Fourier or wavelet domain by a corresponding tensor product algorithm. Under the null hypothesis of additivity or joint effects, the empirical Fourier or wavelet coefficient matrices have some particular structure which separates the empirical coefficients corresponding to the response function and to the noise. Due to the orthogonality of the Fourier or wavelet bases, the empirical coefficients corresponding to the noise preserve the same structure and then the part of the matrices corresponding to the noise is a random sample from a Gaussian distribution with mean 0 and finite variance  $\sigma^2$ . Therefore, testing for additivity or joint effects is equivalent to testing the hypothesis that the part of the empirical coefficient matrices corresponding to the noise come from an r-dimensional Gaussian distribution with mean 0 and variance  $\sigma^2 I_r$ , for appropriately defined r. This problem is then handled in Section 3 by applying the adaptive Neyman truncation and wavelet thresholding procedures of Fan (1996), for testing a high-dimensional Gaussian mean, to the resulting empirical Fourier and wavelet coefficient matrices. As a consequence, asymptotic normality of the proposed test statistics under the null hypothesis and lower bounds of the corresponding powers under a specific alternative are derived. In Section 4 we use several simulated examples to illustrate the performance of the proposed tests, and we make comparisons with other tests available in the literature. Extension of the proposed methodology to high-dimensional (i.e., d > 3) predictors is discussed in Section 5. Finally, some concluding remarks are made in Section 6.

#### 2. Bivariate nonparametric regression models

We consider the bivariate nonparametric regression model

$$Y_t = m(t) + \epsilon_t, \tag{1}$$

where  $Y_t$  is the response variable,  $t = (t_1, t_2) \in [0, 1]^2$  is a twodimensional predictor, and  $\epsilon_t$  is a Gaussian random variable with mean 0 and variance  $0 < \sigma^2 < \infty$ . Borrowing ideas from the theory of analysis of variance, an interaction model writes m(t)as a constant term plus a sum of two functions of one variable (the "main effects") plus a function of two variables (the "two factor interactions"). In other words, the underlying response function satisfies the following decomposition

$$m(t_1, t_2) = m_0 + m_1(t_1) + m_2(t_2) + m_{12}(t_1, t_2), \qquad (2)$$

where  $m_0$  is a constant term,  $m_1(t_1)$  is either zero or a nonconstant function of  $t_1$  (the main effect of  $t_1$ ),  $m_2(t_2)$  is either zero or a non-constant function of  $t_2$  (the main effect of  $t_2$ ), and  $m_{12}(t_1, t_2)$  is either zero or a non-zero function which cannot be decomposed as a sum of a function of  $t_1$  and a function of  $t_2$  (the interaction term). In order to make the decomposition (2) unique, the functions  $m_1(t_1), m_2(t_2)$  and  $m_{12}(t_1, t_2)$  satisfy the following identifiability conditions

$$\int_0^1 m_1(t_1) dt_1 = \int_0^1 m_2(t_2) dt_2 = 0,$$
  
$$\int_0^1 m_{12}(t_1, t_2) dt_1 = \int_0^1 m_{12}(t_1, t_2) dt_2 = 0,$$
  
$$\int_0^1 \int_0^1 m_{12}(t_1, t_2) dt_1 dt_2 = 0.$$

We discuss the problem of testing for additivity, i.e.

$$H_0: \ m_{12}(t_1, t_2) \equiv 0 \tag{3}$$

which corresponds to a vanishing interaction term, and the problem of testing for joint effects, i.e.

$$H_0: m_1(t_1) \equiv 0 \text{ and } m_{12}(t_1, t_2) \equiv 0$$
 (4)

which corresponds to a vanishing first main effect plus a vanishing interaction term. (Treatment for the problem of testing the hypothesis  $H_0$ :  $m_2(t_2) \equiv 0$  and  $m_{12}(t_1, t_2) \equiv 0$  is analogous and it is omitted for brevity.)

In practical applications, however, the experimenter always deals with discrete data. We assume that the response variable  $Y_t$  in model (1)–(2) is observed on the two-dimensional lattice  $\{(t_{1i}, t_{2j}) : i = 0, 1, ..., n_1 - 1; j = 0, 1, ..., n_2 - 1\}$  in  $[0, 1]^2$ . We will first deal with the regular design; however, the non-regular design will also be investigated in later sections via a bivariate spline interpolation algorithm. Using the corresponding discrete identifiability conditions, the

nonparametric regression model is then uniquely decomposed as

$$Y_{ij} = m(t_{1i}, t_{2j}) + \epsilon_{ij}, \qquad (5)$$

where

$$m(t_{1i}, t_{2j}) = m_0 + m_1(t_{1i}) + m_2(t_{2j}) + m_{12}(t_{1i}, t_{2j}).$$
(6)

We also assume that the  $\epsilon_{ij}$  are independent and identically distributed Gaussian random variables with mean 0 and variance  $0 < \sigma^2 < \infty$ . The noise level  $\sigma$  may, or may not, be known. If it is unknown, it will be estimated from the data (see Section 3). The goal is to develop testing procedures for the hypotheses (3)– (4) from the observations  $Y_{ij}$  in (5)–(6) without assuming any particular parametric structure for  $m_1(t_1), m_2(t_2)$  and  $m_{12}(t_1, t_2)$ .

## 2.1. Tensor product Hilbert spaces

The representation of model (2) in terms of tensor product Hilbert spaces plays an important role in the analysis that follows and it is discussed below. Hereafter, we assume that the main effect functions  $m_1(t_1)$  and  $m_2(t_2)$  belong to the Sobolev spaces  $H_1 = W^{\beta_1}[0, 1]$  and  $H_2 = W^{\beta_2}[0, 1]$  respectively, where the regularity indices  $\beta_1$  and  $\beta_2$  take non-integer values greater than 1/2.

By following Antoniadis (1984), if  $H_1$  and  $H_2$  are two real separable Hilbert spaces, the space  $W = H_1 \hat{\otimes}_2 H_2$  denotes the tensor product Hilbert space obtained by completing the prehilbertian space  $H_1 \otimes H_2$  endowed with the following scalar product

$$\langle h_1 \otimes h_2, k_1 \otimes k_2 \rangle_W = \langle h_1, k_1 \rangle_{H_1} \langle h_2, k_2 \rangle_{H_2}, h_1, k_1 \in H_1,$$
  
$$h_2, k_2 \in H_2.$$

Let  $K^1$  and  $K^2$  be the linear closed subspaces of  $H_1$  and  $H_2$ spanned respectively by the constant functions  $\mathbf{1}^1$  and  $\mathbf{1}^2$  on [0, 1], defined by  $\mathbf{1}^1(t) = \mathbf{1}^2(t) = 1$  for all  $t \in [0, 1]$ . Then, from standard literature on tensor product Hilbert spaces, we have that  $W^{\beta_1} = H_1 \otimes_2 K^2$  and  $W^{\beta_2} = K^1 \otimes_2 H_2$  are closed linear subspaces of W, and that  $W^{\beta_1}$  and  $W^{\beta_2}$  are isomorphic to  $H_1$  and  $H_2$  respectively.

According to the above notation, the constant function  $m_0$  of the model can be represented in W by the corresponding function  $M^0$  in  $W^0 = K^1 \hat{\otimes}_2 K^2$ , and the main effects  $m_1(t_1)$  and  $m_2(t_2)$ of the model can be represented in W by the corresponding functions  $M^1(t_1, t_2)$  in  $W^{\beta_1}$  and  $M^2(t_1, t_2)$  in  $W^{\beta_2}$ , defined by

$$M^0 = m_0 (\mathbf{1}^1(t_1) \otimes \mathbf{1}^2(t_2)), \quad M^1(t_1, t_2) = m_1(t_1) \otimes \mathbf{1}^2(t_2)$$

and 
$$M^2(t_1, t_2) = \mathbf{1}^1(t_1) \otimes m_2(t_2).$$
 (7)

With these notations the mean function  $m(t_1, t_2)$  of model (2), under the additivity assumption (3), can be explicitly written as

$$m(t_1, t_2) = M^0 + M^1(t_1, t_2) + M^2(t_1, t_2)$$
(8)

while, under the joint effects assumption (4), can be explicitly written as

$$n(t_1, t_2) = M^0 + M^2(t_1, t_2).$$
(9)

Similarly, the discretized values  $m(t_{1i}, t_{2j})$  of the mean function  $m(t_1, t_2)$  of models (8) and (9) can be also explicitly written as

$$m(t_{1i}, t_{2j}) = M^0 + M^1(t_{1i}, t_{2j}) + M^2(t_{1i}, t_{2j}), \quad (10)$$

$$m(t_{1i}, t_{2j}) = M^0 + M^2(t_{1i}, t_{2j}).$$
(11)

#### 2.2. Fourier and wavelet direct separation

Let us consider two orthogonal and periodic bases for  $L^2([0, 1])$ ,  $B^i = \{\varphi_k^i(t) : k \in \mathcal{I}_i\}, i = 1, 2$ . They can be Fourier bases, i.e.  $\varphi_k^i(t) = \exp(i2\pi kt)$  and  $\mathcal{I}_i = \mathbb{Z}$  for i = 1, 2, or wavelet bases obtained from a multiresolution analysis of  $L^2(\mathbb{R})$  adapted to the interval [0, 1] by periodic boundary handling (see, for example, Mallat 1999, Section 7.5.1), i.e.  $\mathcal{I}_i = \{(-1, 0), (j, l) : j \ge 0, 0 \le l \le 2^j - 1\}$  and  $\varphi_{(-1,0)}^i(t) = \phi_{0,0}^i(t), \varphi_{(j,l)}^i(t) = \psi_{j,l}^i(t)$ , for i = 1, 2. It is easy to prove that  $B^1 \otimes B^2 = \{\varphi_{k_1}^1 \otimes \varphi_{k_2}^2 : k_1 \in \mathcal{I}_1, k_2 \in \mathcal{I}_2\}$  is a basis for  $L^2([0, 1]^2) = L^2([0, 1]) \otimes_2 L^2([0, 1])$ . Note that the wavelet bases  $\{\phi_{0,0}^1(t), \psi_{j,l}^1(t)\}$  and  $\{\phi_{0,0}^2(t), \psi_{j,l}^2(t)\}$ are not necessary the same on the two coordinates. Therefore, similar to Amato, Antoniadis and De Feis (2002) in the Fourier case and similar to Amato and Antoniadis (2001) in the wavelet case,  $M^1(t_1, t_2)$  and  $M^2(t_1, t_2)$  given in (7) can be decomposed as

$$M^{1}(t_{1}, t_{2}) = \sum_{k_{1} \in \mathcal{I}_{1}, k_{2} \in \mathcal{I}_{2}} \mu^{1}_{k_{1}, k_{2}} \varphi^{1}_{k_{1}}(t_{1}) \varphi^{2}_{k_{2}}(t_{2}),$$
$$M^{2}(t_{1}, t_{2}) = \sum_{k_{1} \in \mathcal{I}_{1}, k_{2} \in \mathcal{I}_{2}} \mu^{2}_{k_{1}, k_{2}} \varphi^{1}_{k_{1}}(t_{1}) \varphi^{2}_{k_{2}}(t_{2}),$$

where

$$\mu_{k_1,k_2}^1 = \langle M^1(t_1, t_2), \varphi_{k_1}^1 \otimes \varphi_{k_2}^2 \rangle = \langle m_1, \varphi_{k_1}^1 \rangle_{H_1} \langle \mathbf{1}^2, \varphi_{k_2}^2 \rangle_{H_2},$$
  
$$\mu_{k_1,k_2}^2 = \langle M^2(t_1, t_2), \varphi_{k_1}^1 \otimes \varphi_{k_2}^2 \rangle = \langle \mathbf{1}^1, \varphi_{k_1}^1 \rangle_{H_1} \langle m_2, \varphi_{k_2}^2 \rangle_{H_2};$$

Furthermore,  $m(t_1, t_2)$  can be decomposed as

$$m(t_1, t_2) = \sum_{(k_1, k_2)} \mu_{k_1, k_2} \varphi_{k_1}^1(t_1) \varphi_{k_2}^2(t_2)$$

where

$$\mu_{k_1,k_2} = \langle m(t_1, t_2), \varphi_{k_1}^1 \otimes \varphi_{k_2}^2 \rangle$$
  
=  $m_0 \langle \mathbf{1}^1, \varphi_{k_1}^1 \rangle_{H_1} \langle \mathbf{1}^2, \varphi_{k_2}^2 \rangle_{H_2} + \mu_{k_1,k_2}^1 + \mu_{k_1,k_2}^2$ 

In the Fourier case, since  $\varphi_k^i$  has zero first moment for  $k \neq 0$  and i = 1, 2, we then have

$$\begin{cases} \mu_{k_1,k_2}^1 = 0, & \text{if } k_2 \neq 0, \\ \mu_{k_1,k_2}^2 = 0, & \text{if } k_1 \neq 0, \end{cases}$$

implying that  $\mu_{0,0} = m_0$  and

$$\begin{cases} \mu_{k_1,0} = \mu_{k_1,0}^1, & \text{if } k_1 \neq 0, \\ \mu_{0,k_2} = \mu_{0,k_2}^2, & \text{if } k_2 \neq 0. \end{cases}$$

In the wavelet case, since the mother wavelets  $\psi^1$  and  $\psi^2$  have null moments, we then have

$$\mu_{k_1,k_2}^1 = 0, \quad \text{if } k_2 \neq (-1,0), \\ \mu_{k_1,k_2}^2 = 0, \quad \text{if } k_1 \neq (-1,0),$$

implying that  $\mu_{(-1,0),(-1,0)} = m_0$  and

$$\mu_{k_1,(-1,0)} = \mu_{k_1,(-1,0)}^1, \quad \text{if } k_1 \neq (-1,0), \\ \mu_{(-1,0),k_2} = \mu_{(-1,0),k_2}^2, \quad \text{if } k_2 \neq (-1,0).$$

In other words, in view of the above, the two-dimensional continuous Fourier transform (wavelet transform) of the additive model (8) is simply given by the constant  $m_0$ , the continuous Fourier coefficients  $\mu_{k_1,0}^1$  (wavelet coefficients  $\mu_{k_1,(-1,0)}^1$ ) of the component  $m_1(t_1)$  and the continuous Fourier coefficients  $\mu_{0,k_2}^2$ (wavelet coefficients  $\mu_{(-1,0),k_2}^2$ ) of the component  $m_2(t_2)$ , all the other continuous Fourier coefficients (wavelet coefficients) being zero. Similarly, the two-dimensional continuous Fourier transform (wavelet transform) of the joint effects model (9) is simply given by the constant  $m_0$  and the continuous Fourier coefficients  $\mu_{0,k_2}^2$  (wavelet coefficients  $\mu_{(-1,0),k_2}^2$ ) of the component  $m_2(t_2)$ , all the other continuous Fourier coefficients (wavelet coefficients) being zero.

Assume now that the response variable  $Y_t$  in model (1)–(2) is observed on the two-dimensional lattice  $\{(t_{1i}, t_{2j}) : i = 0, 1, ..., n_1 - 1; j = 0, 1, ..., n_2 - 1\}$  in  $[0, 1]^2$ , and consider the equivalent discrete model (5)–(6). Assuming further that the lattice is equispaced with  $n_1 = 2^{J_1}$  and  $n_2 = 2^{J_2}$  (for some integers  $J_1 > 0$  and  $J_2 > 0$ ), it is then possible to evaluate the two-dimensional discrete Fourier or wavelet transform of the sampled function  $m(t_{1i}, t_{2j})$ .

Let us consider first the two-dimensional discrete Fourier transform of the sampled function  $m(t_{1i}, t_{2j})$ . Let  $\hat{m}_{k_1,k_2}^1$  and  $\hat{m}_{k_1,k_2}^2$   $(k_1 = -n_1/2, \ldots, n_1/2 - 1; k_2 = -n_2/2, \ldots, n_2/2 - 1)$  be the two-dimensional discrete Fourier coefficients of  $M^1(t_{1i}, t_{2j})$  and  $M^2(t_{1i}, t_{2j})$  respectively. Similarly to the two-dimensional continuous Fourier transform, it is also true for the two-dimensional discrete Fourier transform that

$$\begin{cases} \hat{m}_{k_1,k_2}^1 = 0, & \text{if } -n_2/2 \le k_2 \ne 0 \le n_2/2 - 1, \\ \hat{m}_{k_1,k_2}^2 = 0, & \text{if } -n_1/2 \le k_1 \ne 0 \le n_1/2 - 1. \end{cases}$$

This means that the matrix  $(\hat{m}_{k_1,k_2})_{k_1,k_2}$  of the two-dimensional discrete Fourier coefficients of the additive model (10) has the

following form

$$\begin{pmatrix} 0 & \cdots & 0 & \hat{m}_{-n_{1}/2,0}^{1} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \hat{m}_{-1,0}^{1} & 0 & \cdots & 0 \\ \hat{m}_{0,-n_{2}/2}^{2} & \cdots & \hat{m}_{0,-1}^{2} & \hat{m}_{0,0} & \hat{m}_{0,1}^{2} & \cdots & \hat{m}_{0,n_{2}/2-1}^{2} \\ 0 & \cdots & 0 & \hat{m}_{1,0}^{1} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \hat{m}_{n_{1}/2-1,0}^{1} & 0 & \cdots & 0 \end{pmatrix}.$$

$$(12)$$

Similarly, the matrix  $(\hat{m}_{k_1,k_2})_{k_1,k_2}$  of the two-dimensional discrete Fourier coefficients of the joint effects model (11) has the following form

$$\begin{pmatrix} 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \hat{m}_{0,-n_2/2}^2 & \cdots & \hat{m}_{0,-1}^2 & \hat{m}_{0,0} & \hat{m}_{0,1}^2 & \cdots & \hat{m}_{0,n_2/2-1}^2 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$
(13)

Let us consider now the two-dimensional (standard form) discrete wavelet transform of the sampled function  $m(t_{1i}, t_{2j})$ . It can be fastly computed provided that an approximation of the scaling coefficients  $\alpha_{(J_1,l_1),(J_2,l_2)}^n = \langle m, \phi_{J_1,l_1}^1 \otimes \phi_{J_2,l_2}^2 \rangle_{L^2} (0 \le l_1 \le 2^{J_1} - 1; 0 \le l_2 \le 2^{J_2} - 1)$  is given. A simple approximation consists of considering the sampled values as scaling coefficients

$$\alpha_{(J_1, l_1), (J_2, l_2)}^n \approx \frac{m(t_{1l_1}, t_{2l_2})}{\sqrt{2^{J_1} 2^{J_2}}}, \quad l_1 = 0, 1, \dots, 2^{J_1} - 1;$$
  
 $l_2 = 0, 1, \dots, 2^{J_2} - 1. \quad (14)$ 

Although the approximation (14) is made usually in applications, we point out that it is more accurate if Coiflets are chosen as wavelet bases (see Beylkin, Coifman and Rokhlin 1991, Antoniadis 1994).

Let  $\hat{m}_{k_1,k_2}^1$  and  $\hat{m}_{k_1,k_2}^2$   $(k_1 = (-1, 0), (0, 0), \dots, (J_1 - 1, 2^{J_1 - 1} - 1), k_2 = (-1, 0), (0, 0), \dots, (J_2 - 1, 2^{J_2 - 1} - 1))$  be the twodimensional (standard form) discrete wavelet coefficients of  $M^1(t_{1i}, t_{2j})$  and  $M^2(t_{1i}, t_{2j})$  respectively. The two-dimensional

238

(standard form) discrete wavelet transform inherits the property of the continuous wavelet transform, then it holds that

$$\begin{cases} \hat{m}_{k_1,k_2}^1 = 0, & \text{if } k_2 \neq (-1,0), \\ \hat{m}_{k_1,k_2}^2 = 0, & \text{if } k_1 \neq (-1,0). \end{cases}$$

This means that the matrix  $(\hat{m}_{k_1,k_2})_{k_1;k_2}$  of the two-dimensional (standard form) discrete wavelet coefficients of the additive model (10) has the following form

that, in the case of additive structure, only the central row and central column of the empirical Fourier coefficient matrix (see (12)) or only the first row and the first column of the empirical wavelet coefficient matrix (see (15)) contribute to the model; all the remaining elements represent only noise.

Let  $A_1$  be the set of elements belonging to this significant part of the empirical Fourier or wavelet coefficient matrices, and let  $A_2$  be the set of elements belonging to the remaining (major) part of the corresponding matrices. The number of elements in  $A_2$  is  $r = n_1n_2 - n_1 - n_2 + 1$ . Letting U to be the r-dimensional

$\hat{m}_{(-1,0),(-1,0)}$	$\hat{m}^2_{(-1,0),(0,0)}$	$\hat{m}^2_{(-1,0),(1,0)}$				$\hat{m}^2_{(-1,0),(J_2-1,2^{J_2-1})}$
$\hat{m}^1_{(0,0),(-1,0)}$	0	0			•••	0
$\hat{m}^1_{(1,0),(-1,0)}$	0	·	·			0
÷	÷	·	·	·.		÷
÷	÷	·	·	·	·	÷
:	÷	·	·	۰.	·	÷
$\hat{m}^{1}_{(J_{1}-1,2^{J_{1}-1}),(-1,0)}$	0	0	•••	•••	•••	0

Similarly, the matrix  $(\hat{m}_{k_1,k_2})_{k_1,k_2}$  of the two-dimensional (standard form) discrete wavelet coefficients of the joint effects model (11) has the following form vector consisting of the elements of the set  $A_2$  then, by orthogonality of the Fourier and wavelet transforms, U is distributed as an *r*-dimensional Gaussian distribution,  $\mathcal{N}(a, \sigma^2 I_r)$ .

( <sup>n</sup>	$\hat{n}_{(-1,0),(-1,0)}$	$\hat{m}^2_{(-1,0),(0,0)}$	$\hat{m}^2_{(-1,0),(1,0)}$				$\hat{m}^2_{(-1,0),(J_2-1,2^{J_2-1})}$
	0	0	0				0
	0	0	·	·.	۰.		0
	:	÷	·	۰.	۰.	۰.	÷
	:	÷	·	۰.	۰.	۰.	÷
	÷	÷	·	·.	·.	·.	÷
	0	0	0	•••	•••	• • •	0

# 3. Testing procedures for additivity and joint effects

In this section we show how the particular structure of matrices (12), (13), (15) and (16) discussed in Section 2.2 can be combined with the adaptive procedures of Fan (1996) to derive powerful testing procedures for additivity and joint effects in bivariate nonparametric regression models.

## 3.1. Testing for additivity

We first consider the problem of testing for additivity through the data  $Y_{ij}$  discussed in (5)–(6). We perform a Fourier or wavelet transform of the data and then operate on the obtained matrices of corresponding empirical Fourier or wavelet coefficients. The Fourier and wavelet analysis discussed in Section 2.2 shows

Under the additivity assumption, U is distributed as  $\mathcal{N}(0, \sigma^2 I_r)$ and, therefore, testing the hypothesis (3) (additivity) is equivalent to testing the hypotheses

 $H_0: a = 0$  versus  $H_1:$  at least one component of a is not zero

for the r-dimensional vector U. The proposed tests for additivity are based on the adaptive Neyman truncation and wavelet thresholding procedures of Fan (1996) for testing the hypothesis (17).

3.1.1. Fourier and wavelet adaptive Neyman truncation tests

We define the tests *FAN* (Fourier Adaptive Neyman) and *WAN* (Wavelet Adaptive Neyman) whose statistic is

$$T_{AN} = \sqrt{2 \log \log r} \ T_{AN}^{\star} - \{2 \log \log r + 0.5 \log \log \log r - 0.5 \log(4\pi)\}, \quad (18)$$

with

$$T_{AN}^{\star} = \max_{1 \le k \le r} \left\{ \frac{1}{\sqrt{2k}} \sum_{j=1}^{k} \left( \frac{U_j^2}{\sigma^2} - 1 \right) \right\},$$

where  $U_j$  (j = 1, 2, ..., r) are the components of the *r*-dimensional vector *U* obtained from the appropriate part of the empirical Fourier matrix (for the *FAN* test) and the empirical wavelet matrix (for the *WAN* test), respectively.

By using Theorem 2.1 of Fan (1996), the asymptotic distribution of the *FAN* and *WAN* tests, under the additivity hypothesis (17), is given by

$$P(T_{AN} < x) \rightarrow \exp(-\exp(-x)), \text{ as } r \rightarrow \infty,$$

and its critical region, for the significance level  $\alpha$ , is given by

$$T_{AN} > -\log(-\log(1-\alpha)).$$

As noted by Fan (1996), the convergence of the  $T_{AN}$  test statistic given above is quite slow. However, the finite sample distribution based on one million simulations is given in Table 1 of Fan and Lin (1998).

By using Theorem 2.2 of Fan (1996), the following lower bound for the power of the *FAN* and *WAN* tests under the specific alternative  $a = \theta_{0r}$ , with  $\theta_{0r}$  an *r*-dimensional vector,

**Table 1.** Empirical powers ( $\alpha = 0.01$ ) for the  $V_d$ ,  $V_f$ ,  $M_{2,\alpha}$ ,  $\hat{T}_n^{(1,2)}$ , FAN, WAN, H and S tests, obtained for the functions  $m_1 - m_5$  (additive cases), for sample size (8,8) over an equispaced design in  $[0,1]^2$ , and for standard deviations  $\sigma = 0.1$ , 0.5, 1. The parameter  $j_0$  was ranged for both H and S in [0, 2]

	σ	$V_d$	$V_f$	Μ <sub>2,α</sub>	$\hat{T}_n^{(1,2)}$	FAN	WAN	Н	S
$m_1$	0.1	0.988	0.980	0.985	0.968	0.947	0.788	0.773	0.678
	0.5	0.990	0.974	0.990	0.974	0.959	0.795	0.778	0.635
	1	0.989	0.979	0.987	0.973	0.951	0.779	0.768	0.616
$m_2$	0.1	1.000	0.971	0.992	1.000	1.000	0.806	0.792	0.699
	0.5	0.989	0.975	0.991	0.993	0.980	0.783	0.773	0.630
	1	0.985	0.874	0.991	0.968	0.956	0.786	0.779	0.616
$m_3$	0.1	1.000	0.978	1.000	1.000	1.000	0.782	0.762	0.684
	0.5	0.995	0.971	0.995	0.998	0.992	0.805	0.794	0.648
	1	0.984	0.966	0.988	0.980	0.971	0.795	0.764	0.586
$m_4$	0.1	1.000	0.968	1.000	1.000	1.000	0.790	0.776	0.694
	0.5	0.989	0.972	0.994	0.999	0.961	0.788	0.780	0.635
	1	0.992	0.973	0.991	0.985	0.970	0.814	0.800	0.608
$m_5$	0.1	1.000	0.977	0.994	1.000	1.000	0.786	0.775	0.677
	0.5	0.992	0.972	0.990	0.998	0.989	0.789	0.775	0.622
	1	0.984	0.967	0.991	0.979	0.964	0.799	0.794	0.626

is obtained

$$\begin{split} P_{\theta_{0r}} \left( T_{AN} > -\log(-\log(1-\alpha)) \right) \\ &\geq P_{\theta_{0r}} \left( \frac{1}{\sqrt{2m_0}} \sum_{j=1}^{m_0} \left( \frac{U_j^2}{\sigma^2} - 1 - \theta_{0j}^2 \right) \geq \sqrt{2\log\log r} \\ &\times (1+o(1)) - \max_{1 \leq k \leq r} \frac{1}{\sqrt{2k}} \sum_{j=1}^k \theta_{0j}^2 \right), \end{split}$$

where  $m_0 = \operatorname{argmax}_{1 \le k \le r} \frac{1}{\sqrt{2k}} \sum_{j=1}^k \theta_{0j}^2$ . Moreover, by following the arguments given in Fan (1996, Section 2.1), we conclude that the *FAN* and *WAN* tests perform at least as well as the ideal Neyman truncation test within a factor of logarithmic logarithmic order.

The asymptotic properties of the  $T_{AN}$  test statistic given in (18) are based on  $\sigma$  which has to be known and fixed. When  $\sigma$  is unknown, we replace it by a consistent estimator. For the *FAN* test, we take the following estimate

$$\hat{\sigma}^2 = \frac{1}{r-2} \sum_{j=2}^{r-1} \left( \sum_{k=-1}^{1} \omega_k U_{j+k} \right)^2$$

with  $\omega_{-1} = \omega_1 = 1/\sqrt{6}$  and  $\omega_0 = -2/\sqrt{6}$  as proposed by Müller and Stadtmüller (1987). For the WAN test, as an estimate of  $\sigma$  we take the median absolute deviation of the empirical wavelet coefficients in  $A_2$  associated to the finest resolution level and divided by 0.6745, as proposed by Donoho and Johnstone (1994). Such choices may affect the small-sample performances of the tests, but investigation of the small-sample performances of alternative variance estimators is beyond the scope of this paper. Furthermore, the results of Horowitz and Spokoiny (2001, Section 2.5) could be used to investigate whether the asymptotic results of the FAN and WAN tests still remain true if  $\sigma$  is replaced by a consistent estimator, regardless that  $H_0$  is true or not, although again such an investigation is beyond the scope of this paper. Let us just mention that we have found the proposed variance estimators to work well in finite-sample situations and it is therefore safe to recommend their use in practical applications.

#### 3.1.2. Hard and soft wavelet thresholding tests

Since large values of  $|\theta_{0j}|$  appear at large indices, the *FAN* test considered in Section 3.1.1 will not perform well against the alternative hypothesis whose energy concentrates at very high frequencies; that is, for inhomogeneous functions. Although an improvement is possible in this case by the *WAN* test considered also in Section 3.1.1, a better performance is expected, however, by the *H* (Hard thresholding) and *S* (Soft thresholding) tests that we describe below. Hereafter, we assume that  $U_j$  (j = 1, 2, ..., r) are the components of the *r*-dimensional vector *U* obtained from the appropriate part of the empirical wavelet matrix.

240

We first define the Hard thresholding test statistic,  $T_H$ , as

$$T_H = \sigma_H^{-1} (T_H^{\star} - \mu_H), \tag{19}$$

where

$$T_{H}^{\star} = \sigma^{-2} \sum_{j=1}^{J_{0}} U_{j}^{2} + \sigma^{-2} \sum_{j=J_{0}+1}^{r} U_{j}^{2} I(|U_{j}| \ge \sigma \delta_{H}),$$

and

$$\mu_H = J_0 + \sqrt{2/\pi} a_r^{-1} \delta_H (1 + \delta_H^{-2}) \quad \text{and}$$
  
$$\sigma_H^2 = 2J_0 + \sqrt{2/\pi} a_r^{-1} \delta_H^3 (1 + 3\delta_H^{-2}).$$

The threshold value is given by  $\delta_H = \sqrt{2 \log ((r - J_0)a_r)}$ ,  $J_0$  is the number of the empirical wavelet coefficients left unchanged (see below) and  $a_r$  is given by

$$a_r = \begin{cases} \min\left(4\left(\max_{1 \le j \le r} \left|\frac{U_j}{\sigma}\right|\right)^{-4}, \log^{-2} r\right), & J_0 = 0\\ \log^{-2} (r - J_0), & J_0 \ne 0. \end{cases}$$

By Theorem 2.3 of Fan (1996), under the additivity hypothesis (17), we have that  $T_H \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$ , provided that  $a_r \log^{1/2} r \to 0$ . Finally, we get the following testing procedure: reject the null hypothesis (additivity) if

$$T_H > \Phi^{-1}(1-\alpha),$$

where  $\alpha$  is the significance level and  $\Phi$  is the standard Gaussian cumulative distribution function.

By using Theorem 2.4 of Fan (1996), the following lower bound for the power of the *H* test under the specific alternative  $a = \theta_{0r}$ , with  $\theta_{0r}$  an *r*-dimensional vector, is obtained

$$P_{\theta_{0r}}\{T_H > \Phi^{-1}(1-\alpha)\}$$

$$\geq P_{\theta_{0r}}\left\{\sum_{j\in S_0} \left(\frac{U_j^2}{\sigma^2} - 1\right) \geq m_0 \delta_H^2 + \sigma_H(\Phi^{-1}(1-\alpha) - Z_r)\right\},$$

provided that  $m_0 = o(r\sqrt{a_r})$ , where  $S_0$  is the "oracle" best subset of  $m_0$  elements containing the index of the first  $m_0$  largest empirical wavelet coefficients, and  $Z_r$  is a sequence of random variables converging to the standard Gaussian random variable and independent of  $U_j$ ,  $j \in S_0$ . Moreover, by following the arguments given in Fan (1996, Section 2.2), we conclude that the Htest mimics the performance of the oracle wavelet thresholding test within a factor of logarithmic order.

We now define the Soft test statistic,  $T_S$ , as

$$T_{S} = \sigma_{S}^{-1} (T_{S}^{\star} - \mu_{S}), \qquad (20)$$

where

$$T_{S}^{\star} = \sigma^{-2} \sum_{j=1}^{J_{0}} U_{j}^{2} + \sum_{j=J_{0}+1}^{r} \left( \operatorname{sgn}(U_{j}) \left( \frac{|U_{j}|}{\sigma} - \delta_{S} \right)_{+} \right)^{2},$$

and

$$\mu_{S} = J_{0} + 2\sqrt{2/\pi} a_{r}^{-1} \delta_{S}^{-3} (1 - 4\delta_{S}^{-2}) \text{ and}$$
  
$$\sigma_{S}^{2} = 2J_{0} + 2\sqrt{2/\pi} a_{r}^{-1} \delta_{S}^{-1}.$$

The threshold value  $\delta_S$  is given by  $\delta_S = \sqrt{2 \log((r - J_0)a_r)}$ , where  $J_0$  is defined as in the *H* test discussed above and  $a_r$  is given by

$$a_r = \begin{cases} \min\left(\left\{\log\left(\sum_{j=1}^r \left(\frac{U_j}{\sigma}\right)^2\right)\right\}^{-2}, \log^{-2} r\right), & J_0 = 0\\ \log^{-2} (r - J_0), & J_0 \neq 0. \end{cases}$$

By Theorem 2.3 of Fan (1996), under the additivity hypothesis (17), we have that  $T_S \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$ , provided that  $a_r \log^{5/2} r \to 0$ . Finally, we get the following testing procedure: reject the null hypothesis (additivity) if

$$T_S > \Phi^{-1}(1-\alpha),$$

where  $\alpha$  is the significance level and  $\Phi$  is the standard Gaussian cumulative distribution function. Since the convergence of  $T_S$  is slower than the one based on  $T_H$ , the test statistic  $T_S$  is not further pursued.

The asymptotic properties of the  $T_H$  and  $T_S$  test statistics given in (19) and (20) respectively are based on  $\sigma$  which has to be known and fixed. When  $\sigma$  is unknown, we replace it by a consistent estimator; it is again estimated by the median absolute deviation of the empirical wavelet coefficients associated to the finest resolution level, divided by 0.6745, as proposed by Donoho and Johnstone (1994). Again, as in Section 3.1.1, we have found the proposed variance estimators to work well in finite-sample situations and it is therefore safe to recommend their use in practical applications.

Finally we discuss the choice of  $J_0$  appearing in  $T_H$  and  $T_S$ . Since the wavelet coefficients at the same resolution level carry the same information, it is reasonable to leave unchanged the empirical wavelet coefficients belonging to the same resolution level. If  $J_0$  is not zero then its value can be, for example, the number of elements contained in the first row and first column of the matrix corresponding to  $A_2$ , i.e. the empirical wavelet coefficients of the coarsest level 0. If we know that the wavelet coefficient are reasonable large we can then decide to leave unchanged the empirical wavelet coefficient belonging to levels 0 and 1, i.e. to leave unchanged the elements of the first 3 rows and columns of the matrix corresponding to  $A_2$ . If we decide to leave unchanged the empirical wavelet coefficients belonging to levels 0, 1 and 2 then we leave unchanged the elements of the first 7 rows and columns of the matrix corresponding to  $A_2$ , and so on. The parameter  $J_0$  can be defined in terms of an option parameter  $j_0$  by the following formula

$$J_0 = (2^{j_0} - 1)((n_1 - 1) + (n_2 - 1) - (2^{j_0} - 1)),$$
  
$$j_0 = 0, 1, 2, \dots, \min\{J_1 - 1, J_2 - 1\}$$

Selecting  $j_0 = 0$  we threshold all the empirical wavelet coefficients, selecting  $j_0 = 1$  we threshold all the empirical wavelet

**Table 2.** Empirical powers ( $\alpha = 0.01$ ) for the  $V_d$ ,  $V_f$ ,  $M_{2,\alpha}$ ,  $\hat{T}_n^{(1,2)}$ , FAN, WAN, H and S tests, obtained for the functions  $m_6 - m_{13}$  (non-additive cases), for sample size (8,8) over an equispaced design in  $[0,1]^2$ , and for standard deviations  $\sigma = 0.1$ , 0.5, 1. The parameter  $j_0$  was ranged for both H and S in [0, 2]

	σ	$V_d$	$V_f$	$M_{2,\alpha}$	$\hat{T}_n^{(1,2)}$	FAN	WAN	Н	S
<i>m</i> <sub>6</sub>	0.1	1.000	1.000	0.572	0.520	0.997	0.631	0.624	0.836
	0.5	0.092	0.122	0.017	0.036	0.118	0.231	0.233	0.391
	1	0.030	0.044	0.016	0.037	0.054	0.232	0.230	0.391
$m_7$	0.1	0.425	0.454	0.066	0.148	0.496	0.414	0.317	0.501
	0.5	0.023	0.038	0.012	0.030	0.057	0.188	0.228	0.321
	1	0.017	0.028	0.016	0.032	0.054	0.198	0.227	0.316
$m_8$	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.5	0.344	0.545	0.228	0.490	0.879	0.866	0.542	0.783
	1	0.022	0.071	0.037	0.087	0.270	0.441	0.303	0.498
$m_9$	0.1	1.000	1.000	1.000	1.000	1.000	0.999	0.999	1.000
	0.5	0.073	0.187	0.064	0.118	0.299	0.404	0.345	0.531
	1	0.025	0.050	0.015	0.040	0.091	0.255	0.253	0.397
$m_{10}$	0.1	0.160	0.116	0.480	0.422	0.041	0.995	0.997	0.999
	0.5	0.029	0.047	0.109	0.224	0.049	0.315	0.320	0.562
	1	0.018	0.030	0.015	0.049	0.048	0.232	0.239	0.400
$m_{11}$	0.1	0.002	0.053	0.000	0.000	0.004	0.373	0.322	0.499
	0.5	0.009	0.027	0.005	0.013	0.040	0.199	0.239	0.386
	1	0.013	0.020	0.011	0.018	0.035	0.214	0.214	0.386
$m_{12}^{[24]}$	0.1	0.000	0.055	0.000	0.000	0.000	0.273	0.294	0.490
$\gamma = 1$	0.5	0.000	0.026	0.000	0.000	0.006	0.215	0.219	0.380
	1	0.007	0.021	0.008	0.008	0.031	0.223	0.238	0.388
$m_{12}^{[24]}$	0.1	0.000	0.149	0.000	0.000	0.000	0.497	0.494	0.703
$\gamma = 2$	0.5	0.000	0.027	0.000	0.000	0.003	0.214	0.226	0.386
-	1	0.005	0.025	0.006	0.003	0.024	0.223	0.239	0.390
$m_{12}^{[24]}$	0.1	0.000	0.023	0.000	0.000	0.001	0.231	0.236	0.417
$\delta = 1/2$	0.5	0.006	0.025	0.007	0.012	0.033	0.204	0.221	0.375
,	1	0.010	0.023	0.012	0.025	0.049	0.219	0.236	0.386
$m_{12}^{[24]}$	0.1	0.002	0.023	0.000	0.000	0.023	0.221	0.242	0.414
$\delta = 1/4$	0.5	0.009	0.025	0.010	0.025	0.037	0.203	0.220	0.372
,	1	0.011	0.022	0.013	0.027	0.049	0.220	0.235	0.386

coefficients except the ones belonging to the first coarsest level, selecting  $j_0 = 2$  we threshold all the empirical wavelet coefficients except the ones belonging to the first two coarsest levels, and so on.

#### 3.2. Testing for joint effects

We now consider the problem of testing for joint effects through the data  $Y_{ij}$  discussed in (5) and (6). As in the problem of testing additivity, we perform a Fourier or wavelet transform of the data and then operate on the obtained matrices of corresponding empirical Fourier or wavelet coefficients. The Fourier and wavelet analysis discussed in Section 2.2 shows that, in the case of joint effects structure, only the central row of the empirical Fourier coefficient matrix (see (13)) or only the first row of the empirical wavelet coefficient matrix (see (16)) contribute to the model; all the remaining elements represent only noise.

Let  $A_1^*$  be the set of elements belonging to this significant part of the empirical Fourier and wavelet coefficient matrices, and let  $A_2^*$  be the set of elements belonging to the remaining (major) part of the corresponding matrices. The number of elements in  $A_2^*$ is  $r^* = n_1 n_2 - n_2$ . Letting  $U^*$  to be the  $r^*$ -dimensional vector consisting of the elements of the set  $A_2^*$  then, by orthogonality of the Fourier and wavelet transforms,  $U^*$  is distributed as an  $r^*$ -dimensional Gaussian distribution,  $\mathcal{N}(a, \sigma^2 I_{r^*})$ . Under the joint effects assumption,  $U^*$  is distributed as  $\mathcal{N}(0, \sigma^2 I_{r^*})$  and, therefore, testing the hypothesis (4) (joint effects) is again equivalent to testing the hypothesis (17) for the  $r^*$ -dimensional vector  $U^*$ . Therefore, the test statistics  $T_{AN}$  given in (18),  $T_H$  given in (19) and  $T_S$  given in (20) are also applied to the  $r^*$ -dimensional vector  $U^*$  when testing for joint effects.

#### 3.3. The non-equispaced design case

We now relax the assumptions that the two-dimensional lattice is equispaced and that the sample sizes  $n_1$  and  $n_2$  are powers of two. We assume that the first covariate  $t_1$  is observed at  $n_1$ distinct points  $0 \le t_{10} < t_{11} < t_{12}, \ldots < t_{1n_1-1} \le 1$  and that the second covariate  $t_2$  is observed at  $n_2$  distinct points  $0 \le t_{20} < t_{21} < t_{22}, \ldots < t_{2n_2-1} \le 1$ , with  $n_1$  and  $n_2$  being any two natural numbers. We propose a two-dimensional interpolation procedure in order to transform the original data to an equispaced two-dimensional lattice which can then be handled by the tools described in Sections 3.1 and 3.2. More specifically, let  $(Y_{ij})_{i,j}$  be the data matrix observed on the two-dimensional lattice { $(t_{1i}, t_{2j})$  :  $i = 0, 1, ..., n_1 - 1$ ;  $j = 0, 1, ..., n_2 - 1$ } in  $[0, 1]^2$ , and let  $\{(\tilde{t}_{1i}, \tilde{t}_{2j}) : i = 0, 1, \dots, m_1 - 1; j =$  $0, 1, \ldots, m_2 - 1$  be an equispaced grid design in the twodimensional lattice  $[0, 1]^2$  with  $m_1 = 2^{\lceil \log_2(n_1) \rceil}$  and  $m_2 =$  $2^{[\log_2(n_2)]}$ , where [x] denotes the integer part of x. We then apply a bivariate spline interpolation procedure over the data matrix  $(Y_{ij})_{i,j}$   $(0 \le i \le n_1 - 1; 0 \le j \le n_2 - 1)$  to obtain a new data matrix  $(\tilde{Y}_{ij})_{i,j}$   $(0 \le i \le m_1 - 1; 0 \le j \le m_2 - 1)$ , which contains an approximation of the response function over the twodimensional equispaced grid design  $(\tilde{t}_{1i}, \tilde{t}_{2j})$ . Of course, when the Fourier or wavelet transform is applied to the new data matrix, we obtain empirical coefficients that are no longer independent. However, as explained in Amato and Antoniadis (2001, Section 3.3), their variances remain bounded. Since the variance estimators and the proposed testing procedures described in Section 3.1 (testing for additivity) and Section 3.2 (testing for joint effects) require the data to be independent, some power loss is unavoidable and this is what we observe in the numerical experiments of Section 4.

Finally we mention that the above computational procedure can also be used for random designs. In other words if the first and second covariates are random, say  $T_1$  and  $T_2$  respectively, then we can assume the  $n_1$  design points for the first covariate  $T_1$  to be independent random variables with a common density  $f_1$  on [0, 1] and the  $n_2$  design points for the second covariate  $T_2$  to be independent random variables with a common density  $f_2$  on [0, 1]; the densities  $f_1$  and  $f_2$  can be of the same or different form. Then the above bivariate spline interpolation algorithm can be applied conditionally on the  $n_1 \times n_2$  design points  $T_{10}$ ,  $T_{11}$ , ...,  $T_{1n_1-1}$  and  $T_{20}$ ,  $T_{21}$ , ...,  $T_{2n_2-1}$ .

# 4. Simulation study

The purpose of this section is to shed some light on the theoretical results and to implement the algorithmic steps discussed in Section 3. We use several simulated examples to investigate the finite performance of the proposed tests for additivity and joint effects, and we make comparisons with other tests available in the literature. The computational algorithms related to wavelet analysis were performed using Version 8 of the WaveLab toolbox for MATLAB that is freely available from http://www-stat.stanford.edu/software/software. html. The entire study was carried out using the MATLAB programming environment.

We study the finite sample performance of the proposed tests over different models. We consider the following additive

**Table 3.** Empirical powers ( $\alpha = 0.01$ ) for the  $V_d$ ,  $V_f$ ,  $M_{2,\alpha}$ ,  $\hat{T}_n^{(1,2)}$ , FAN, WAN, H and S tests, obtained for the functions  $m_1 - m_5$  (additive cases), for sample size (16, 32) over an equispaced design in  $[0,1]^2$ , and for standard deviations  $\sigma = 0.1, 0.5, 1$ . The parameter  $j_0$  was ranged for both H and S in [0,3]

	σ	$V_d$	$V_f$	$M_{2,\alpha}$	$\hat{T}_n^{(1,2)}$	FAN	WAN	Н	S
$m_1$	0.1	0.984	0.985	0.993	0.989	0.962	0.821	0.816	0.903
	0.5	0.991	0.990	0.992	0.986	0.971	0.793	0.792	0.820
	1	0.989	0.988	0.990	0.994	0.969	0.824	0.774	0.784
$m_2$	0.1	0.994	0.988	0.991	1.000	1.000	0.818	0.802	0.910
	0.5	0.989	0.989	0.989	0.989	0.978	0.824	0.811	0.843
	1	0.996	0.995	0.989	0.985	0.975	0.814	0.775	0.782
$m_3$	0.1	1.000	0.992	0.989	1.000	1.000	0.821	0.818	0.903
	0.5	0.993	0.990	0.988	0.998	0.993	0.810	0.809	0.832
	1	0.985	0.984	0.989	0.991	0.979	0.784	0.740	0.750
$m_4$	0.1	0.999	0.987	0.996	1.000	0.983	0.814	0.798	0.909
	0.5	0.994	0.993	0.986	0.999	0.968	0.813	0.805	0.820
	1	0.992	0.989	0.993	0.991	0.968	0.820	0.760	0.783
$m_5$	0.1	1.000	0.985	0.990	1.000	1.000	0.814	0.813	0.903
5	0.5	0.991	0.990	0.992	0.996	0.996	0.796	0.777	0.818
	1	0.988	0.986	0.982	0.985	0.980	0.796	0.765	0.778

models studied in Derbort, Dette and Munk (2002)

$$m_1(t_1, t_2) = 0,$$
  

$$m_2(t_1, t_2) = t_1 + t_2,$$
  

$$m_3(t_1, t_2) = \exp(t_1) + \sin(\pi t_2),$$
  

$$m_4(t_1, t_2) = \sin(\pi t_1) + \sin(\pi t_2),$$
  

$$m_5(t_1, t_2) = \exp(t_1) + \exp(t_2).$$

Moreover, we consider the following non-additive models studied in Barry (1993), Eubank *et al.* (1995) and Derbort, Dette and Munk (2002)

$$m_6(t_1, t_2) = t_1 t_2,$$
  

$$m_7(t_1, t_2) = \exp(5(t_1 + t_2)) / (1 + \exp(5(t_1 + t_2))) - 1,$$
  

$$m_8(t_1, t_2) = 0.5 (1 + \sin(2\pi(t_1 + t_2))),$$
  

$$m_9(t_1, t_2) = 64 (t_1 t_2)^3 (1 - t_1 t_2)^3,$$
  

$$m_{10}(t_1, t_2) = (t_1 + t_2) / 2 + (1 \text{ outlier}),$$
  

$$m_{11}(t_1, t_2) = G(t_1)G(t_2)/36,$$

where

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

(For the description of the outlier in  $m_{10}$ , see Barry 1993.) Finally we consider the following non-additive models

 $m_{12}^{[ij]}(t_1, t_2) = h_i(t_1) + h_j(t_2) + \gamma h_i(t_1)h_j(t_2),$ 

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

**Table 4.** Empirical powers ( $\alpha = 0.01$ ) for the  $V_d$ ,  $V_f$ ,  $M_{2,\alpha}$ ,  $\hat{T}_n^{(1,2)}$ , FAN, WAN, H and S tests, obtained for the functions  $m_6 - m_{13}$  (non-additive cases), for sample size (16,32) over an equispaced design in  $[0, 1]^2$ , and for standard deviations  $\sigma = 0.1, 0.5, 1$ . The parameter  $j_0$  was ranged for both H and S in [0, 3]

	σ	$V_d$	$V_{f}$	$M_{2,\alpha}$	$\hat{T}_{n}^{(1,2)}$	FAN	WAN	Н	S
<i>m</i> <sub>6</sub>	0.1	1.000	1.000	1.000	1.000	1.000	1.000	0.994	1.000
	0.5	0.885	0.889	0.046	0.083	0.906	0.455	0.245	0.264
	1	0.238	0.243	0.017	0.022	0.285	0.221	0.235	0.230
$m_7$	0.1	1.000	1.000	0.650	0.873	1.000	0.948	0.513	0.742
	0.5	0.092	0.096	0.012	0.030	0.159	0.230	0.244	0.187
	1	0.026	0.025	0.008	0.011	0.056	0.187	0.219	0.161
$m_8$	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.5	1.000	1.000	1.000	1.000	1.000	1.000	0.998	1.000
	1	0.976	0.973	0.550	0.794	0.998	0.997	0.513	0.821
$m_9$	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.5	1.000	1.000	0.717	0.913	1.000	0.998	0.578	0.834
	1	0.326	0.333	0.057	0.122	0.636	0.659	0.293	0.320
$m_{10}$	0.1	0.033	0.012	0.346	0.182	0.022	1.000	1.000	1.000
	0.5	0.009	0.009	0.029	0.045	0.025	0.249	0.311	0.299
	1	0.007	0.007	0.013	0.020	0.032	0.181	0.227	0.213
$m_{11}$	0.1	0.379	0.702	0.530	0.002	0.997	0.958	0.744	0.901
	0.5	0.013	0.017	0.015	0.011	0.059	0.207	0.231	0.185
	1	0.010	0.009	0.012	0.017	0.031	0.179	0.204	0.158
$m_{12}^{[24]}$	0.1	0.000	0.120	0.000	0.000	0.161	0.905	0.345	0.591
$\gamma = 1$	0.5	0.007	0.008	0.004	0.000	0.036	0.216	0.249	0.106
	1	0.013	0.015	0.007	0.002	0.033	0.183	0.225	0.109
$m_{12}^{[24]}$	0.1	0.001	0.949	0.000	0.000	0.942	1.000	0.806	0.897
$\gamma = 2$	0.5	0.008	0.018	0.001	0.000	0.065	0.294	0.250	0.267
	1	0.015	0.016	0.006	0.001	0.044	0.190	0.207	0.224
$m_{12}^{[24]}$	0.1	0.011	0.019	0.000	0.000	0.045	0.232	0.211	0.231
$\delta = 1/2$	0.5	0.011	0.010	0.009	0.008	0.029	0.203	0.231	0.235
,	1	0.010	0.011	0.009	0.005	0.031	0.177	0.203	0.215
$m_{12}^{[24]}$	0.1	0.017	0.018	0.002	0.000	0.045	0.215	0.207	0.229
$\delta = 1/4$	0.5	0.011	0.010	0.009	0.012	0.029	0.203	0.231	0.235
,	1	0.010	0.011	0.010	0.005	0.031	0.177	0.202	0.215

where the parameters  $\gamma \neq 0$  and  $\delta \neq 1$  specify the deviation from additivity, and  $h_1$ ,  $h_2$ ,  $h_3$ ,  $h_4$  and  $h_5$  are the Blip, Heavisine, Spikes, Corner and Doppler functions respectively (see, for example, Antoniadis, Bigot and Sapatinas 2001). Obviously, these latter functions do not belong to the Sobolev spaces considered in Section 2. They represent, however, typical signals often considered in the literature for nonparametric regression estimation and, moreover, they are encountered in many practical applications in diverse scientific fields. Moreover, these inhomogeneous functions are very well-suited for wavelet analysis; indeed, in this case, the numerical results below show the benefits of the wavelet-based testing procedures (*WAN*, *H*, *S*) over their competitors.

When testing for additivity, we give comparisons with the  $V_d$  and  $V_f$  tests considered in Eubank *et al.* (1995), the  $\hat{T}_n^{(1,2)}$  (with order l = 1) test considered in Dette and Derbort (2001), and the  $M_{2,\alpha}$  (with order l = 2) test considered in Derbort, Dette and Munk (2002). The  $V_d$  and  $V_f$  tests are based on data-driven methods to select the order of Fourier series esti-

mators of the interaction term, and they test the hypothesis that this estimator is significantly different from zero. The power of the tests is asymptotically assessed under specific smoothness assumptions on the underlying response function. The  $\hat{T}_n^{(1,2)}$  and  $M_{2,\alpha}$  tests are based on empirical measurements of the  $L^2$ -distance between the general model (2) and the model satisfying the null hypothesis (3). These empirical distances are quadratic forms of the data, and asymptotic normality for them is showed under the hypothesis of additivity and specific smoothness assumptions for the underlying response function. The statistics  $\hat{T}_n^{(1,2)}$  and  $M_{2,\alpha}$  test the hypothesis that these estimators of the empirical distance are significantly different from zero.

When testing for joint effects, we give comparisons with the  $\hat{W}_n$  test considered in Derbort and Dette (2001). This test is again based on an estimator of an empirical  $L^2$ -distance between the general model (2) and the model satisfying the null hypothesis (4), and it tests the hypothesis that this estimator is significantly different from zero.

Finally, the above simulation study is also performed in the non-equispaced design case.

#### 4.1. Testing for additivity

In this section we report the results of the simulation study, comparing the empirical powers of  $V_d$ ,  $V_f$ ,  $M_{2,\alpha}$ ,  $\hat{T}_n^{(1,2)}$ , FAN, WAN, H and S tests. We assumed that the response functions  $m_1 - m_{13}$  are observed on the two-dimensional regular lattice  $\{(t_{1i}, t_{2j}): i = 0, 1, \dots, n_1 - 1; j = 0, 1, \dots, n_2 - 1\}$  in  $[0, 1]^2$ , and we considered the values (8, 8), (8, 16), (16, 8), (16, 16), (16, 32), (32, 16) and (32, 32) for the sample size  $(n_1, n_2)$ . However, for the sake of brevity, we only report the results for (8,8) and (16, 32), representing a small and a large sample size respectively. In the simulation study the values of  $\gamma$ and  $\delta$  associated with the response functions  $m_{12}$  and  $m_{13}$  respectively were set to  $\gamma = 1, 2$  and  $\delta = 1/2, 1/4$ , while the standard deviation and significance level were taken as  $\sigma = 0.1, 0.5, 1$ and  $\alpha = 0.01, 0.05, 0.1$  respectively. For the  $V_d, V_f, M_{2,\alpha}$  and  $\hat{T}_n^{(1,2)}$  tests,  $\sigma$  was estimated according to the consistent estimators proposed in conjunction with these tests (we refer to the appropriate papers for more details), while for the FAN, WAN, H and S tests,  $\sigma$  was estimated according to the consistent estimators proposed in Sections 3.1.1 and 3.1.2. Since the relative performances of the various procedures for the different cases of  $\alpha$  were roughly the same, we only report the results for  $\alpha =$ 0.01. The empirical power has been evaluated by the following formula

Empirical Power =  $\begin{cases} 1 - \sum_{k=1}^{M} l_k / M, & \text{if } H_0 \text{ is true,} \\ \sum_{k=1}^{M} l_k / M, & \text{if } H_0 \text{ is false,} \end{cases}$ 

where  $l_k$  is the response of the generic test statistics applied to the *k*-th run (it takes the value 1 if the null hypothesis is rejected and 0 if the null hypothesis is accepted), and *M* is the number of runs (the number of runs was taken M =1000).

Tables 1 and 3 show the empirical powers of the simulation study for the additive models  $m_1-m_5$ . It is observed that  $V_d$ ,  $V_f$ ,  $M_{2,\alpha}$  and  $\hat{T}_n^{(1,2)}$  perform better than WAN, H and S, while FAN gives comparable results. However, the performance of WAN, H and S improves significantly when the sample size increases, which qualitatively confirm the asymptotic results. Tables 2 and 4 show the empirical powers of the simulation study for the non-additive models  $m_6-m_{11}$ . It is observed that FAN, WAN, H and S perform better than  $V_d$ ,  $V_f$ ,  $M_{2,\alpha}$  and  $\hat{T}_n^{(1,2)}$  in most cases for small noise and in almost all cases for moderate and large noises. We also note that, for  $m_{10}$  and  $m_{11}$ , FAN behaves poorly relatively to WAN, H and S, and this is explained by the presence of singularity in the underlying functions which is better handled by the wavelet transform. This is not the case, however, for  $m_6-m_9$  and we stress the very good be-

**Table 5.** Empirical powers ( $\alpha = 0.01$ ) for the  $V_d$ ,  $V_f$ ,  $M_{2,\alpha}$ ,  $\hat{T}_n^{(1,2)}$ , FAN, WAN, H and S tests, obtained for the functions  $m_1 - m_5$  (additive cases), for sample size (10,10) over a lattice generated by combining two grid designs from (21) and (22), and for standard deviations  $\sigma = 0.1, 0.5, 1$ . The parameter  $j_0$  was ranged for both H and S in [0,2]

	σ	$V_d$	$V_f$	$M_{2,\alpha}$	$\hat{T}_n^{(1,2)}$	FAN	WAN	Н	S
$m_1$	0.1	0.987	0.976	0.990	0.982	0.678	0.446	0.890	0.355
	0.5	0.988	0.976	0.987	0.980	0.703	0.466	0.893	0.390
	1	0.982	0.974	0.987	0.978	0.708	0.443	0.886	0.353
$m_2$	0.1	0.999	0.978	0.994	1.000	0.940	0.453	0.891	0.356
	0.5	0.986	0.979	0.990	0.995	0.736	0.448	0.897	0.364
	1	0.990	0.980	0.995	0.981	0.687	0.439	0.901	0.347
$m_3$	0.1	1.000	0.983	1.000	1.000	0.997	0.453	0.913	0.368
	0.5	0.992	0.980	0.991	0.998	0.743	0.415	0.901	0.332
	1	0.991	0.982	0.995	0.994	0.719	0.465	0.894	0.379
$m_4$	0.1	1.000	0.974	1.000	1.000	0.998	0.451	0.994	0.369
	0.5	0.993	0.981	0.991	1.000	0.723	0.426	0.895	0.340
	1	0.993	0.983	0.992	0.992	0.715	0.399	0.892	0.330
$m_5$	0.1	1.000	0.982	0.996	1.000	0.992	0.442	0.896	0.351
	0.5	0.991	0.981	0.993	0.999	0.729	0.434	0.885	0.330
	1	0.984	0.974	0.990	0.990	0.719	0.427	0.900	0.361

haviour of *FAN* for  $m_8$  which is well-suited for Fourier analysis. Tables 2 and 4 also display the results for some combinations of the functions  $m_{12}$  and  $m_{13}$ . These results show that *WAN*, *H* and *S* outperform  $V_d$ ,  $V_f$ ,  $M_{2,\alpha}$ ,  $\hat{T}_n^{(1,2)}$  and *FAN* and clearly demonstrate the benefits of the proposed wavelet-based testing procedures over their competitors for inhomogeneous response functions.

#### 4.2. Testing for joint effects

The same framework described in Section 4.1 for testing additivity has also been considered for testing joint effects, and we have compared the empirical powers of  $\hat{W}_n$ , FAN, WAN, H and S. The numerical results obtained are very similar to those for testing additivity. Indeed, for the response functions  $m_1-m_5$ , the procedures are comparable while, for the response functions  $m_6-m_{11}$ , FAN, WAN, H and S perform better that  $\hat{W}_n$  in most cases for small noise and in almost all cases for moderate and large noises. For the response functions  $m_{12}^{[ij]}$  and  $m_{13}^{[ij]}$ (*i*, *j* = 1, 2, 3, 4, 5), WAN, H and S outperform FAN and  $\hat{W}_n$ and demonstrate once again the benefits of the proposed waveletbased testing procedures over their competitors for inhomogeneous response functions. For the sake of brevity, however, we omit the results obtained in this case.

#### 4.3. The non-equispaced design case

In this section we report the results of the simulation study obtained in the non-equispaced design case. The procedures  $V_d$ ,  $V_f$ ,  $\hat{T}_n^{(1,2)}$ ,  $\hat{W}_n$  and  $M_{2,\alpha}$  do not change, while the proposed

**Table 6.** Empirical powers ( $\alpha = 0.01$ ) for the  $V_d$ ,  $V_f$ ,  $M_{2,\alpha}$ ,  $\hat{T}_n^{(1,2)}$ , FAN, WAN, H and S tests, obtained for the functions  $m_6 - m_{13}$  (non-additive cases), for sample size (10,10) over a lattice generated by combining two grid designs from (21) and (22), and for standard deviations  $\sigma = 0.1, 0.5, 1$ . The parameter  $j_0$  was ranged for both H and S in [0,2]

	σ	$V_d$	$V_{f}$	$M_{2,lpha}$	$\hat{T}_{n}^{(1,2)}$	FAN	WAN	Н	S
<i>m</i> <sub>6</sub>	0.1	1.000	1.000	0.723	0.699	0.680	0.646	0.584	0.744
	0.5	0.113	0.134	0.021	0.037	0.304	0.537	0.460	0.630
	1	0.041	0.058	0.011	0.029	0.304	0.576	0.473	0.651
$m_7$	0.1	0.659	0.711	0.151	0.288	0.656	0.688	0.562	0.724
	0.5	0.038	0.054	0.020	0.036	0.320	0.571	0.493	0.649
	1	0.011	0.021	0.014	0.024	0.299	0.559	0.480	0.658
$m_8$	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.5	0.533	0.647	0.448	0.692	0.683	0.779	0.631	0.788
	1	0.022	0.073	0.054	0.106	0.440	0.619	0.496	0.691
$m_9$	0.1	1.000	1.000	1.000	1.000	0.995	0.971	0.924	0.985
	0.5	0.110	0.173	0.064	0.124	0.442	0.599	0.512	0.698
	1	0.019	0.037	0.017	0.037	0.355	0.576	0.481	0.676
$m_{10}$	0.1	0.139	0.128	0.473	0.364	0.244	0.767	0.721	0.829
	0.5	0.019	0.030	0.094	0.179	0.292	0.592	0.508	0.662
	1	0.012	0.022	0.011	0.027	0.308	0.542	0.462	0.646
$m_{11}$	0.1	0.003	0.177	0.000	0.000	0.315	0.646	0.563	0.732
	0.5	0.008	0.024	0.010	0.015	0.325	0.568	0.481	0.639
	1	0.0013	0.030	0.023	0.023	0.316	0.562	0.494	0.643
$m_{12}^{[21]}$	0.1	0.000	0.075	0.000	0.000	0.005	0.577	0.482	0.658
$\gamma = 1$	0.5	0.003	0.019	0.000	0.000	0.243	0.532	0.430	0.606
	1	0.009	0.022	0.002	0.005	0.273	0.558	0.470	0.643
$m_{12}^{[21]}$	0.1	0.000	0.485	0.000	0.000	0.001	0.591	0.471	0.655
$\gamma = 2$	0.5	0.003	0.034	0.000	0.000	0.188	0.568	0.480	0.656
	1	0.011	0.024	0.005	0.002	0.268	0.567	0.485	0.656
$m_{12}^{[21]}$	0.1	0.000	0.030	0.000	0.000	0.188	0.527	0.445	0.627
$\delta = 1/2$	0.5	0.008	0.016	0.006	0.017	0.282	0.549	0.467	0.625
,	1	0.010	0.025	0.007	0.024	0.271	0.557	0.473	0.641
$m_{12}^{[21]}$	0.1	0.002	0.017	0.001	0.000	0.267	0.549	0.468	0.637
$\delta = 1/4$	0.5	0.011	0.027	0.010	0.017	0.303	0.566	0.478	0.633
	1	0.012	0.020	0.007	0.016	0.303	0.588	0.498	0.651

procedures *FAN*, *WAN*, *H* and *S* are modified according to the approach described in Section 3.3. The following one-dimensional grid designs in [0, 1] were considered

$$t_i = \frac{\exp(i/n) - 1}{e - 1}, \quad i = 0, \dots, n - 1,$$
 (21)

$$t_i = \sqrt{i/n}, \quad i = 0, \dots, n-1,$$
 (22)

$$t_i \sim \text{Uniform}[0, 1], \quad i = 0, \dots, n-1,$$
 (23)

and then an equispaced two-dimensional lattice was created combining any two of these designs.

For brevity, however, we only present the empirical powers obtained when testing for additivity for some combinations of these designs and for the sample size (10, 10). The results for the additive models  $m_1 - m_5$  and the non-additive models  $m_6 - m_{13}$ , over a lattice generated by combining two gird design from (21) and (22), are reported in Tables 5 and 6 respectively. The results for the additive models  $m_1 - m_5$  and the non-additive models  $m_6 - m_{13}$ , over a lattice generated by combining two grid designs from (23), are reported in Tables 7 and 8 respectively.

From the analysis of the results, we see that the conclusions do not change with respect to the equispaced grid design case, although some power loss is unavoidable as discussed in Section 3.3. Indeed, for the additive models  $m_1 - m_5$ , *FAN* and *H* give comparable results while *WAN* and *S* perform worse than in the equispaced design. For the non-additive models  $m_6 - m_{11}$ , *FAN*, *WAN*, *H* and *S* perform better that  $V_d$ ,  $V_f$ ,  $M_{2,\alpha}$  and  $\hat{T}_n^{(1,2)}$  in most cases for small noise and in almost all cases for moderate and large noises. Furthermore, for the non-additive models  $m_{12} - m_{13}$ , as in the equispaced design, *WAN*, *H* and *S* outperform  $V_d$ ,  $V_f$ ,  $M_{2,\alpha}$ ,  $\hat{T}_n^{(1,2)}$  and *FAN*, and clearly demonstrate the benefits of the proposed wavelet-based testing procedures over their competitors for inhomogeneous response functions.

**Table 7.** Empirical powers ( $\alpha = 0.01$ ) for the  $V_d$ ,  $V_f$ ,  $M_{2,\alpha}$ ,  $\hat{T}_n^{(1,2)}$ , FAN, WAN, H and S tests, obtained for the functions  $m_1 - m_5$  (additive cases), for sample size (10,10) over a lattice generated by combining two grid designs from (23), and for standard deviations  $\sigma = 0.1$ , 0.5, 1. The parameter  $j_0$  was ranged for both H and S in [0,2]

	σ	$V_d$	$V_f$	$M_{2,\alpha}$	$\hat{T}_n^{(1,2)}$	FAN	WAN	Н	S
$m_1$	0.1	0.993	0.982	0.995	0.984	0.627	0.208	0.966	0.187
	0.5	0.985	0.977	0.994	0.984	0.597	0.207	0.972	0.128
	1	0.993	0.987	0.987	0.986	0.623	0.199	0.968	0.121
$m_2$	0.1	1.000	0.974	0.999	1.000	0.608	0.211	0.961	0.187
	0.5	0.992	0.982	0.987	0.988	0.611	0.211	0.951	0.117
	1	0.992	0.981	0.991	0.983	0.600	0.224	0.967	0.115
m <sub>3</sub>	0.1	1.000	0.983	1.000	1.000	0.595	0.191	0.951	0.174
	0.5	0.990	0.980	0.995	0.997	0.641	0.199	0.965	0.123
	1	0.993	0.982	0.995	0.991	0.605	0.185	0.959	0.120
$m_4$	0.1	1.000	0.974	1.000	1.000	0.626	0.225	0.966	0.211
	0.5	0.993	0.981	0.991	1.000	0.607	0.230	0.967	0.156
	1	0.992	0.983	0.988	0.995	0.595	0.214	0.959	0.133
$m_5$	0.1	1.000	0.982	1.000	1.000	0.593	0.208	0.968	0.196
	0.5	0.993	0.981	0.995	1.000	0.600	0.205	0.962	0.127
	1	0.985	0.974	0.989	0.990	0.606	0.199	0.963	0.113

## 5. Extension to high dimensions

Let us consider model (1) in the general case

$$m(t_1, \dots, t_d) = m_0 + \sum_{j_1=1}^d m_{j_1}(t_{j_1}) + \sum_{1 \le j_1 < j_2 \le d} m_{j_1, j_2}(t_{j_1}, t_{j_2})$$
  
+ \dots + \sum\_{1 \le j\_1 < \dots < j\_p \le d} m\_{j\_1, \dots, j\_p}(t\_{j\_1}, \dots, t\_{j\_p}),

where  $1 \leq p \leq d$ ,  $m_0$  is a constant, and  $m_{j_i}(t_{j_i})$ ,  $m_{j_1,j_2}(t_{j_1}, t_{j_2}), \ldots, m_{j_1,\ldots,j_p}(t_{j_1}, \ldots, t_{j_p})$  are unknown smooth functions; the terms  $m_{j_i}(t_{j_i})$  are called the "main effects", the terms  $m_{j_1,j_2}(t_{j_1}, t_{j_2})$  are called the "two factor interactions", and so on.

Testing for additivity is equivalent to testing the hypothesis that the interaction terms of all order are all equal to zero. We suppose the main effect  $m_{j_i}(t_{j_i})$  belongs to Sobolev space  $H_i = W^{\beta_i}[0, 1], (\beta_i > 1/2)$  for each i = 1, ..., d, and we define the tensor product Hilbert space  $W = H_1 \otimes \cdots \otimes H_d$  endowed with the following scalar product

$$\langle h_1 \otimes \cdots \otimes h_d, k_1 \otimes \cdots \otimes k_d \rangle_W = \langle h_1, k_1 \rangle_{H_1} \cdots \langle h_d, k_d \rangle_{H_d},$$

$$h_1, k_1 \in H_1, \ldots, h_d, k_d \in H_d.$$

We have that the constant function  $m_0$  can be represented in Wby a function  $M^0 \in 1^1(t) \otimes \cdots \otimes 1^d(t)$  and, for each  $1 \le p \le d$ , the main effect  $m_{j_p}(t_{j_p})$  can be represented in W as a function  $M^p(t_1, \ldots, t_d) \in 1^1(t) \otimes \cdots \otimes H_p \otimes \cdots \otimes 1^d(t)$ . If  $B^i = \{\varphi_k^i(t) : k \in \mathcal{I}_i\}, i = 1, \ldots, d$ , are d orthogonal

If  $B^i = \{\varphi_k^i(t) : k \in \mathcal{I}_i\}, i = 1, \dots, d$ , are d orthogonal and periodic basis for  $L_2[0, 1]$ , then  $B^1 \otimes \cdots \otimes B^d$  is a basis for  $L_2([0, 1]^p) = L_2[0, 1] \otimes \cdots \otimes L_2[0, 1]$ . The basis  $B^i$  can be a Fourier basis or a wavelet basis as explained for the twodimensional case in Section 2.2. In the case of additivity one has the following representation

$$m(t_1,...,t_d) = \sum_{(k_1,...,k_d)} \mu_{k_1,...,k_d} \varphi_{k_1}^1(t_1) \dots \varphi_{k_d}^d(t_d),$$

where

$$\mu_{k_1,\dots,k_d} = \langle m(t_1,\dots,t_d), \varphi_{k_1}^1 \otimes \dots \otimes \varphi_{k_d}^d \rangle \\ = m_0 \langle \mathbf{1}^1, \varphi_{k_1}^1 \rangle_{H_1} \cdots \langle \mathbf{1}^d, \varphi_{k_d}^d \rangle_{H_d} + \mu_{k_1,\dots,k_d}^1 \\ + \dots + \mu_{k_1,\dots,k_d}^d.$$

In the Fourier case,  $\varphi_k^i$  has zero first moment for  $k \neq 0$  and i = 1, ..., d, then we have

$$\begin{cases} \mu_{k_1,\dots,k_d}^1 = 0, & \text{if } k_2 \neq 0, \text{ or } k_3 \neq 0, \text{ or } k_d \neq 0 \\ \vdots \\ \mu_{k_1,\dots,k_d}^d = 0, & \text{if } k_1 \neq 0, \text{ or } k_2 \neq 0, \text{ or } k_{d-1} \neq 0. \end{cases}$$

In the wavelet case, analogous decompositions hold if proper indices are used. Under the additivity assumption, the *d*dimensional matrix of Fourier or wavelet coefficients has only *d* lines (one along each dimension) related to the respective main effect functions, all the other coefficients being zeros. The test for additivity can be performed using the vector  $U_d$  consisting of those elements of the matrix which we expect to be zero under the null hypothesis. Testing for joint effects is also easily extended to the general case  $d \ge 3$  since it only requires the appropriate elements of the vector  $U_d$  to be used in the proposed testing procedures. Finally, other interesting hypothesis testing, such as testing specific vanishing interaction terms between the components  $t_{j_1}, \ldots, t_{j_p}$  for  $1 \le p \le d$  can also be tested by appropriately adapting the proposed testing procedures.

# 6. Concluding remarks

We have considered the problem of testing for additivity and joint effects in multivariate nonparametric regression when the data are modelled as observations of an unknown response function observed on a *d*-dimensional ( $d \ge 2$ ) lattice and contaminated with additive Gaussian noise. We have proposed tests for additivity and joint effects, appropriate for both homogeneous and inhomogeneous response functions, using the particular structure of the data expanded in tensor product Fourier or wavelet bases studied recently by Amato and Antoniadis (2001) and Amato, Antoniadis and De Feis (2002). The corresponding tests were constructed by applying the adaptive Neyman truncation and wavelet thresholding procedures of Fan (1996), for testing a high-dimensional Gaussian mean, to the resulting empirical Fourier and wavelet coefficients. The empirical powers obtained from the simulation study and the comparisons made with other tests available in the literature suggest that the advantage of the proposed tests is twofold. First, they have higher empirical power in detecting additivity and joint effects for inhomogeneous response functions while maintaining a comparable behaviour for homogeneous response functions as the other available tests and,

**Table 8.** Empirical powers ( $\alpha = 0.01$ ) for the  $V_d$ ,  $V_f$ ,  $M_{2,\alpha}$ ,  $\hat{T}_n^{(1,2)}$ , FAN, WAN, H and S tests, obtained for the functions  $m_6 - m_{13}$  (non-additive cases), for sample size (10,10) over a lattice generated by combining two grid designs from (23), and for standard deviations  $\sigma = 0.1$ , 0.5, 1. The parameter  $j_0$  was ranged for both H and S in [0,2]

	σ	$V_d$	$V_f$	$M_{2,\alpha}$	$\hat{T}_{n}^{(1,2)}$	FAN	WAN	Н	S
<i>m</i> <sub>6</sub>	0.1	0.999	1.000	0.500	0.184	0.393	0.789	0.735	0.869
	0.5	0.096	0.122	0.020	0.030	0.377	0.788	0.712	0.863
	1	0.040	0.054	0.011	0.024	0.374	0.820	0.739	0.832
$m_7$	0.1	0.113	0.136	0.024	0.062	0.425	0.800	0.740	0.871
	0.5	0.017	0.031	0.016	0.031	0.382	0.792	0.726	0.866
	1	0.013	0.018	0.013	0.024	0.391	0.794	0.715	0.805
$m_8$	0.1	1.000	1.000	1.000	1.000	0.439	0.798	0.719	0.864
	0.5	0.865	0.891	0.457	0.704	0.418	0.805	0.729	0.878
	1	0.241	0.291	0.061	0.123	0.402	0.787	0.708	0.801
$m_9$	0.1	1.000	1.000	1.000	1.000	0.376	0.794	0.716	0.875
	0.5	0.457	0.568	0.124	0.197	0.398	0.793	0.718	0.804
	1	0.093	0.120	0.025	0.036	0.378	0.800	0.723	0.824
$m_{10}$	0.1	0.132	0.119	0.471	0.364	0.397	0.807	0.722	0.868
	0.5	0.013	0.024	0.100	0.159	0.387	0.772	0.701	0.869
	1	0.010	0.021	0.019	0.038	0.390	0.780	0.720	0.852
$m_{11}$	0.1	0.000	0.019	0.000	0.000	0.392	0.788	0.704	0.876
	0.5	0.008	0.019	0.005	0.009	0.388	0.783	0.716	0.884
	1	0.0012	0.029	0.003	0.017	0.384	0.779	0.713	0.879
$m_{12}^{[21]}$	0.1	0.000	0.022	0.000	0.000	0.367	0.774	0.705	0.876
$\gamma = 1$	0.5	0.003	0.017	0.000	0.000	0.366	0.787	0.715	0.874
	1	0.008	0.026	0.002	0.004	0.376	0.783	0.692	0.802
$m_{12}^{[21]}$	0.1	0.000	0.070	0.000	0.000	0.383	0.786	0.721	0.889
$\gamma = 2$	0.5	0.003	0.024	0.000	0.000	0.391	0.780	0.699	0.793
	1	0.011	0.021	0.004	0.007	0.412	0.797	0.717	0.819
$m_{12}^{[21]}$	0.1	0.000	0.022	0.000	0.000	0.404	0.778	0.708	0.883
$\delta = 1/2$	0.5	0.007	0.014	0.011	0.023	0.394	0.788	0.722	0.894
	1	0.011	0.025	0.007	0.018	0.404	0.800	0.735	0.819
$m_{12}^{[21]}$	0.1	0.001	0.017	0.001	0.000	0.375	0.791	0.721	0.864
$\delta = 1/4$	0.5	0.012	0.028	0.017	0.023	0.366	0.799	0.723	0.893
· ·/·	1	0.012	0.021	0.013	0.378	0.404	0.815	0.744	0.898

second, their empirical power deteriorates much less, as the noise level increases, than their competitors.

In summary, we have seen evidence that the Fourier-based test (FAN) is appropriate for functions obeying a homogeneous behaviour, while the wavelet-based tests (WAN, H, S) are appropriate for functions obeying a non-homogeneous behaviour. Furthermore, since these tests are also computationally simple to obtain, practitioners may sometimes prefer these tests in particular applications, especially in the case where the underlying response function obeys a non-homogeneous behaviour and the noise level is high.

# Acknowledgments

The work of Daniela De Canditiis was supported by the Agenzia Spaziale Italiana (Italian Space Agency). The work of Theofanis Sapatinas was partly supported by a 'University of Cyprus– 2002 Research Grant'. Daniela De Canditiis would like to thank Theofanis Sapatinas for warm hospitality while visiting Nicosia to carry out part of this work. The authors would like to thank Umberto Amato, Anestis Antoniadis and Italia De Feis for fruitful discussions. Helpful comments and suggestions made by the Associate Editor and two reviewers are gratefully acknowledged.

# References

- Amato U. and Antoniadis A. 2001. Adaptive wavelet series estimation in separable nonparametric regression models. Statist. Comput. 11: 373–394.
- Amato U., Antoniadis A., and De Feis I. 2002. Fourier series approximation of separable models. J. Computat. Appl. Math. 146: 459–479.
- Antoniadis A. 1984. Analysis of variance on function spaces. Math. Operationsforsch. u. Statist., Ser. Statist. 15: 59–71.
- Antoniadis A. 1994. Smoothing data with coiflets. Statistica Sinica, 4: 651–678.
- Antoniadis A., Bigot J., and Sapatinas T. 2001. Wavelet estimators in nonparametric regression: A comparative simulation study. J. Statist. Soft. 6(6): 1–83.

- Barry D. 1993. Testing for additivity of a regression function. Ann. Statist. 21: 235–254.
- Beylkin G., Coifman R.R., and Rokhlin V. 1991. The fast wavelet transform and numerical algorithms. Comm. Pure. Appl. Math. 44: 141–163.
- Buja A., Hastie T., and Tibshirani R. 1989. Linear smoothers and additive models (with discussion). Ann. Statist. 17: 453– 555.
- Derbort S., Dette H., and Munk A. 2002. A test for additivity in nonparametric regression. Ann. Inst. Statist. Math. 54: 60–82.
- Dette H. and Derbort S. 2001. Analysis of variance in nonparametric regression models. J. Multivariate Anal. 76: 110–137.
- Dette H. and Wilkau C.V.U. 2001. Testing additivity by kernelbased methods-what is a reasonable test? Bernoulli 7: 669– 697.
- Donoho D.L. and Johnstone I.M. 1994. Ideal spatial adaptation by wavelet shrinkage. Biometrika 81: 425–455.
- Eubank R.L., Hart J.D., Simpson D.G., and Stefanski L.A. 1995. Testing for additivity in nonparametric regression. Ann. Statist. 23: 1896– 1920.
- Fan J. 1996. Test of significance based on wavelet thresholding and Neyman's truncation. J. Am. Statist. Assoc. 91: 674–688.
- Fan J. and Lin S.-K. 1998. Test of significance when data are curves. J. Am. Statist. Assoc. 93: 1007–1021.
- Hastie T.J. and Tibshirani R.J. 1990. Generalized Additive Models. Chapman and Hall, London.
- Horowitz J.L. and Spokoiny V.G. 2001. An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. Econometrica 69: 599–631.

- Linton O.B. and Nielsen J.P. 1995. A kernel method of estimating structured nonparametric regression based on marginal integration. Biometrika 82: 93–100.
- Linton O.B. 1997. Efficient estimation of additive nonparametric regression models. Biometrika 84: 469–473.
- Mallat S.G. 1999. A Wavelet Tour of Signal Processing. 2nd edition. Academic Press, San Diego.
- Müller H.G. and Stadtmüller U. 1987. Estimation of heteroscedastisity in regression analysis. Ann. Statist. 15: 610–625.
- Opsomer, J.D. and Ruppert D. 1997. Fitting a bivariate additive model by local polynomial regression. Ann. Statist. 25: 186–211.
- Opsomer J.D. and Ruppert D. 1998. A fully automated bandwidth selection method for fitting additive models. J. Am. Statist. Assoc. 93: 605–619.
- Sardy S. and Tseng P. 2004. Automatic nonlinear fitting of additive and generalised additive models with wavelets. J. Comp. Graph. Statist. 13: (to appear).
- Sperlich S., Linton O.B., and Härdle W. 1999. Integration and backfitting methods in additive models—finite sample properties and comparison. Test 8: 419–458.
- Sperlich S., Tjøstheim D., and Yang L. 2002. Nonparametric estimation and testing of interaction in additive models. Econometric Theory 18: 197–251.
- Stone C.J. 1985. Additive regression and other nonparametric models. Ann. Statist. 13: 689–705.
- Tukey J. 1949. One degree of freedom test for non-additivity. Biometrics 5: 232–242.
- Zhang S. and Wong M.-Y. 2003. Wavelet threshold estimation for additive regression models. Ann. Statist. 31: 152–173.