

ON WAVELET METHODS FOR TESTING EQUALITY OF MEAN RESPONSE CURVES

PENGFEEI GUO

*Department of Mathematics and Statistics, York University
Toronto, ON, Canada, M3J 1P3
pguo@mathstat.yorku.ca*

ALWELL JULIUS OYET *

*Department of Mathematics and Statistics, Memorial University
St. John's, NL, CANADA, A1C 5S7
aoyet@math.mun.ca*

Received (Day Month Year)
Revised 27 August 2008
Communicated by Alwell Oyet

In this article, we exploit the adaptive properties of wavelets to develop some procedures for testing the equality of nonlinear and nonparametric mean response curves which are assumed by an experimenter to be the underlying functions generating several groups of data with possibly heteroscedastic errors. The essential feature of the techniques is the transformation of the problem from the domain of the input variable to the wavelet domain through an orthogonal discrete wavelet transformation or a multiresolution expansion. We shall see that this greatly simplifies the testing problem into either a wavelet thresholding problem or linear wavelet regression problem. The size and power performances of the tests are reported and compared to some existing methods. The tests are also applied to data on dose response curves for vascular relaxation in the absence or presence of a nitric oxide inhibitor.

Keywords: False Discovery Rate; Discrete Wavelet Transformation; Multiresolution Analysis.

AMS Subject Classification: 62G08, 62G10, 41A30, 62P10

1. Introduction

The problem of comparing the underlying mean response function $f(x)$ generating data y , in the presence of a covariate x , across several groups have been discussed and very well motivated by several authors. In many practical situations, the exact mathematical structure of the mean response $f(x)$ is usually unknown. Thus,

*Corresponding author: Alwell Oyet, Department of Mathematics and Statistics, Memorial University of Newfoundland, St. John's, NL, Canada A1C 5S7. Tel.: (709) 737-8075; fax.: (709) 737-3010; e-mail: aoyet@math.mun.ca.

2 *P. Guo and A. Oyet*

several authors have discussed the problem in the context of nonparametric and semiparametric regression. Dette and Neumeyer⁷ provided a brief description of tests developed in References 6, 9-12, 14, 15, 18, 22. The procedures discussed by these authors are limited to data with homoscedastic error in all groups. The work of Dette and Neumeyer⁷ extended the analysis of variance type test statistic introduced by Young and Bowman²² to the case of heteroscedastic data. They also proposed a test based on differences between a nonparametric variance estimator in the combined data from all groups $\{y_{ij}; j = 1, 2, \dots, n_i, i = 1, 2, \dots, k\}$ and the corresponding estimators in the individual groups $\{y_{ij}; j = 1, 2, \dots, n_i\}$. The idea for a test based on differences was first discussed in Ref. 21 in the case of two groups. Other authors who have discussed procedures for comparing regression functions in the presence of heteroscedastic errors are Gorgens,⁸ Koul and Schick,¹³ and Neumeyer and Dette.¹⁹

A feature that is common to the techniques discussed by these authors is the assumption that the unknown response function will be estimated by some kernel method. In this paper, our procedures are based on transforming the problem from the domain of the input variable to the wavelet domain through a discrete wavelet transformation (DWT) of the data vector or a multiresolution expansion of the response curve. Previously, Maharaj¹⁶ used wavelet coefficients obtained through DWT to construct a test statistic for the comparison of two time series. By applying a DWT matrix \mathbf{W} to the data vector \mathbf{Y} we show, in §2, that the problem of comparing pairs of mean response functions is equivalent to the problem of thresholding in wavelet analysis. This result then leads directly to the false discovery rate (FDR) and graphical procedures we shall discuss later. In the application of wavelets to nonparametric curve estimation, thresholding is one of the steps in the DWT approach. Discrete wavelet transformations are linear transformations which map data \mathbf{Y} from the domain of the input data vector to the wavelet domain. The result is a vector of wavelet coefficients \mathbf{d} of the same size as the data \mathbf{Y} . Once \mathbf{d} has been found, thresholding is applied to \mathbf{d} and the DWT inverted to obtain an estimate $\hat{\mathbf{f}} = \mathbf{W}^{-1}(\hat{\sigma}_\epsilon \delta_\lambda(\mathbf{W}\mathbf{Y}/\hat{\sigma}_\epsilon))$. The component \hat{f}_i can then be written as

$$\hat{f}_i = \sum_k w_{ki}(\hat{\sigma}_\epsilon \delta_\lambda(\mathbf{W}\mathbf{Y}/\hat{\sigma}_\epsilon)_k), \quad (1.1)$$

$i = 1, \dots, n$, where w_{ki} are the elements of \mathbf{W}^{-1} and $\hat{\sigma}^2$ is an estimate of the error variance. Thresholding is sometimes referred to as denoising or whitening because it de-correlates even highly correlated data. This property is useful when dealing with correlated data.

The most common thresholding policies are *hard* and *soft* thresholding defined respectively by $\delta^h(d_i, \lambda) = d_i \cdot \mathbf{1}(|d_i| > \lambda)$ and $\delta^s(d_i, \lambda) = (d_i - \text{sgn}(d_i) \cdot \lambda) \cdot \mathbf{1}(|d_i| > \lambda)$, where $\lambda \geq 0$ is a thresholding parameter and $d_i \in \mathbb{R}$. When $d_i \sim N(\theta, \sigma^2)$, σ^2 known, Ref. 17 among others have shown that the mean and variance of δ^h and δ^s under

squared error loss are, respectively:

$$\begin{aligned} M_\lambda^h(\theta) &= \theta + \theta[1 - \Phi(\lambda - \theta) - \Phi(\lambda + \theta)] + \phi(\lambda - \theta) - \phi(\lambda + \theta), \\ V_\lambda^h(\theta) &= (\theta^2 + 1)[1 - \Phi(\lambda - \theta) - \Phi(\lambda + \theta)] + (\lambda + \theta)\phi(\lambda - \theta) + (\lambda - \theta)\phi(\lambda + \theta) - (M_\lambda^h(\theta))^2, \\ M_\lambda^s(\theta) &= M_\lambda^h(\theta) - \lambda[\Phi(\lambda + \theta) - \Phi(\lambda - \theta)], \quad V_\lambda^s(\theta) = V_\lambda^h(\theta) - \lambda[v(\lambda, \theta) + v(\lambda, -\theta)], \end{aligned} \quad (1.2)$$

where ϕ and Φ are the standard normal density and its cumulative distribution function and $v(\lambda, \theta) = [1 + \Phi(\lambda - \theta) - \Phi(\lambda + \theta)] \cdot [(2\theta - \lambda)(1 - \Phi(\lambda - \theta)) + 2\phi(\lambda - \theta)]$. Under certain conditions, Brillinger³ showed that, for each i , \hat{f}_i is asymptotically Gaussian with standard errors estimated by $s_i = \hat{\sigma} \sqrt{\sum_k w_{ki}^2 V_\lambda \left(\frac{\hat{\theta}_k}{\hat{\sigma}} \right)}$. It then follows that for a fixed confidence level α we can construct an approximate Bonferroni-type simultaneous confidence interval for \hat{f}_i given by

$$\left[\hat{f}_i - t_{n-1, \alpha/2n} \cdot s_i, \hat{f}_i + t_{n-1, \alpha/2n} \cdot s_i \right]. \quad (1.3)$$

If the mean response function $f(x)$ is not too irregular, the coefficients d_i at the finest scale of the discrete wavelet transformation should contain mainly noise and the square of the signal to noise ratio will be small. Thus, the wavelet coefficients d_i at the finest scale are used in the estimation of σ^2 . Two variance estimators we have used in this paper are $\hat{\sigma}_s = \sqrt{\frac{1}{n/2-1} \sum_{i=1}^{n/2} [d_i^{(J-1)} - \bar{d}^{(J-1)}]^2}$, and a more robust median absolute deviation estimator $\hat{\sigma}_{MAD} = 1/0.6745 \cdot \text{MAD} [\mathbf{d}^{(J-1)}] = 1.4826 \cdot \text{MEDIAN} [|\mathbf{d}^{(J-1)} - \text{MEDIAN}(\mathbf{d}^{(J-1)})|]$. In the expressions for $\hat{\sigma}_s$ and $\hat{\sigma}_{MAD}$, $\mathbf{d}^{(J-1)}$ is the vector of finest detail coefficients associated with the multiresolution subspace W_{J-1} described in the next paragraph.

Aside from the FDR and graphical methods, we shall also discuss a test we have called the F_1 test. The F_1 test is based on a multiresolution analysis (MRA) of the $\mathbb{L}_2(\mathbb{R})$ space. The MRA consists of an increasing sequence of closed subspaces V_j , $j \in \mathbb{Z}$ satisfying the conditions: (a) $\bigcap_j V_j = 0$; (b) $\overline{\bigcap_j V_j} = \mathbb{L}_2(\mathbb{R})$; (c) there exists a scaling function $\phi \in V_0$ such that $\phi(x - k)$, $k \in \mathbb{Z}$ is an orthonormal basis of V_0 ; (d) for all $k \in \mathbb{Z}$, $f(x) \in V_j \iff f(x - k) \in V_j$, and (e) $f(x) \in V_j \iff f(2x) \in V_{j+1}$. Properties (c) - (e) leads to a dilation equation for the scaling function $\phi(x)$, $\phi(x) = \sum_{p \in \mathbb{Z}} h_p \phi(2x - p)$, where h_r are constants called filter coefficients. In order to ensure the existence of a unique solution to the dilation equation and orthogonality of the translates of $\phi(x)$, it is required that $\sum_{p \in \mathbb{Z}} h_p = 2$ and $\sum_{p \in \mathbb{Z}} h_p h_{p-2j} = \delta_j$, $j \in \mathbb{Z}$. Define the difference space W_j to be the orthogonal complement of V_j , satisfying the property that $W_j \oplus V_j = V_{j+1}$, $W_j \perp V_j$. Then, it can be shown that $V_j = V_0 \oplus \bigoplus_{i=0}^{j-1} W_i$ and that for any $\psi(x) \in W_0$, we can write $\psi(x) = \sum_{r \in \mathbb{Z}} g_r \phi(2x - r)$, where $g_r = (-1)^r h_{-r+1}$ with $\int \phi(x) dx = 1$, $\int \psi(x) dx = 0$, and $\int \phi^2(x) dx = 1$. This implies that any function $f \in \mathbb{L}_2(\mathbb{R})$ can be written as a linear combination of

4 *P. Guo and A. Oyet*

functions in V_0 and W_i , $i = 0, \dots, \infty$ (convergent in $\mathbb{L}_2(\mathbb{R})$) given by

$$f(x) = \sum_{k \in \mathbb{Z}} c_{j_0 k} \phi_{j_0 k}(x) + \sum_{j=j_0}^{\infty} \sum_{k \in \mathbb{Z}} d_{jk} \psi_{jk}(x), \quad (1.4)$$

where $c_{j_0 k}$, d_{jk} are unknown constant coefficients, and $\psi_{jk}(x) = 2^{j/2} \psi(2^j x - k)$ and $\phi_{jk}(x) = 2^{j/2} \phi(2^j x - k)$, $k \in \mathbb{Z}$ are dilated and translated versions of $\phi(x)$ and $\psi(x)$ respectively. We note that (1.4) can be written as $f(x) = \mathbf{q}_m^T(x) \beta + g(x)$ where m is a finite level at which to terminate the multiresolution expansion, $g(x)$ are the remainder terms which we assume is negligible, $\mathbf{q}_m^T(x) = (\phi_0(x), \psi_{0,0}(x), \psi_{1,0}(x), \psi_{1,1}(x), \dots, \psi_{m,2^m-1}(x))$ and β is a vector of unknown constants. Given repeated observations y_{ij} , an alternative estimate of $f(x)$ can be written as $\hat{f}(x) = \sum_{i,j} w_{ij} y_{ij}$. Various weights w_{ij} which correspond to wavelet versions of the Gasser Müller estimator and weighted least squares have been discussed by Oyet and Sutradhar²⁰. They also discussed various methods for estimating the error variance σ^2 based on these weights.

Two wavelet bases we used in our simulation studies are the Haar and the Daubechies wavelet systems. The scaling functions and primary wavelets of the Daubechies⁴ wavelet systems, commonly represented as ${}_N\phi(x)$ and ${}_N\psi(x)$ respectively, have no closed forms. They are constructed numerically for different values of the wavelet number N , which identify the number of nonvanishing coefficients in the "dilation equation" ${}_N\phi(x) = \sum h_k \cdot {}_N\phi(2x - k)$. The choice $N = 1$ yields the Haar wavelets. In our studies, the Daubechies-Lagarias pyramidal algorithm of Daubechies and Lagarias⁵ was used in the construction of ${}_N\phi(x)$. Once ${}_N\phi(x)$ was constructed, the corresponding primary wavelet was obtained from the relation ${}_N\psi(x) = \sum_k (-1)^k h_{1-k} \cdot {}_N\phi(2x - k)$. The functions ${}_N\phi(x)$ and ${}_N\psi(x)$, which we shall refer to as DaubN, have compact support with vanishing moments of order 1 to N and support width $2N - 1$. This property, commonly referred to as a moment condition, guarantees good approximation properties of the corresponding wavelet expansion of a response function $f(x)$ in ${}_N\phi(x)$ and ${}_N\psi(x)$. That is, it determines how quickly the wavelet expansion will converge to the true response $f(x)$.

This article is organized as follows. In §2 we develop the FDR and F_1 test procedures for comparing pairs of mean response curves followed by a simulation study of the size and power performances of the proposed methods. We also discuss a graphical method similar to the method of Bowman and Young,² which can be used when the number of observations is small. The graphical procedure is extended to compare more than 2 curves in §3. In the case where the number of observations and the input points may not be the same for each of the groups, we develop a more general test statistic, obtain their critical values by simulation and study their power performances. We end with some concluding remarks on the results of comparing our methods to the kernel-based method of Ref. 12. We also discuss the effect of

deviations from normality of the error distribution on the proposed tests and apply the techniques to data on dose response curves for vascular relaxation in the absence or presence of a nitric oxide inhibitor obtained from a perinatal research.

2. Test Procedures Based On Discrete Wavelet Transformation And Multiresolution Analysis

Consider k groups of data arising from the nonparametric model

$$y_{ij} = f_j(x_{ij}) + \sigma_j \varepsilon_{ij}, \quad j = 1, 2, \dots, k; \quad i = 1, 2, \dots, n_j, \quad (2.1)$$

where for a fixed group j , $f_j(x)$ is a nonlinear response function and ε_{ij} is a sequence of independent and identically distributed normal random variables with mean 0 and $\text{Var}(\varepsilon_{ij}) = 1$. An experimenter is interested in determining whether the mean response functions f_1, f_2, \dots, f_k are the same. In statistical notations, the problem is to test the hypothesis $H_0 : f_1(x) = f_2(x) = \dots = f_k(x)$ against the alternative $H_a : f_j(x) \neq f_k(x), j \neq k$, for some x . These hypotheses can also be written as

$$H_0 : f_1(x) - f_2(x) = f_2(x) - f_3(x) = \dots = f_{k-1}(x) - f_k(x) = 0, \text{ at each } x, \text{ against} \\ H_a : f_{j-1}(x) - f_j(x) \neq 0, \text{ for some } x, j = 2, \dots, k. \quad (2.2)$$

In the sections that follow, we shall discuss wavelet methods for testing H_0 .

2.1. A false discovery rate procedure (FDR)

The FDR approach we are about to describe is applicable when (a) $k = 2$, (b) $n_1 = n_2 = n$ and (c) $x_{i1} = x_{i2} = x_i, i = 1, 2, \dots, n$. Define $\mathbf{y}_1 = (y_{11}, y_{21}, \dots, y_{n,1})'$, $\mathbf{f}_1(x) = (f_1(x_1), f_1(x_2), \dots, f_1(x_n))'$, and $\varepsilon_1 = (\varepsilon_{11}, \varepsilon_{21}, \dots, \varepsilon_{n,1})'$. Correspondingly, we also define \mathbf{y}_2 , $\mathbf{f}_2(x)$ and ε_2 from (2.1). Thus, in vector notations (2.1) can be written as $\mathbf{y} = \mathbf{f}(x) + \varepsilon$ where $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2)^T$, $\mathbf{f} = (\mathbf{f}_1, \mathbf{f}_2)^T$ and $\varepsilon = (\varepsilon_1, \varepsilon_2)^T$. Let \mathbf{I}_n be the $n \times n$ identity matrix and $\mathbf{D} = [\mathbf{I}_n \quad -\mathbf{I}_n]$. The FDR approach begins with the computation of the difference vector $\mathbf{u} = \mathbf{D}\mathbf{y} = \mathbf{y}_1 - \mathbf{y}_2$. We then let $\mathbf{h}(x) = \mathbf{D}\mathbf{f}(x) = \mathbf{f}_1(x) - \mathbf{f}_2(x)$ and $\mathbf{e} = \mathbf{D}\varepsilon = \varepsilon_1 - \varepsilon_2$. It follows that

$$\mathbf{u} = \mathbf{h}(x) + \mathbf{e}. \quad (2.3)$$

Now, let \mathbf{W} be a discrete wavelet transformation matrix and define $\mathbf{d} = \mathbf{W}\mathbf{u}$, $\theta(x) = \mathbf{W}\mathbf{h}(x)$ and $\xi = \mathbf{W}\mathbf{e}$. By applying the DWT matrix \mathbf{W} to (2.3) we obtain $\mathbf{d} = \theta(x) + \xi$, where the components of \mathbf{d} are wavelet coefficients representing the differenced vector \mathbf{u} in the wavelet domain.

Clearly, if H_0 in (2.2) is true, then $\theta(x) = \mathbf{W}\mathbf{h}(x) = \mathbf{0}$ at each point x and the components of the differenced data $\mathbf{d} = \mathbf{W}\mathbf{u}$, in the wavelet domain will be dominated by the noise component. Thus, in the wavelet domain the hypotheses

$$H_0 : \theta_i = 0, \text{ at each } i, \text{ against the alternative, } H_a : \theta_i \neq 0, \text{ for some } i, \quad (2.4)$$

6 *P. Guo and A. Oyet*

is equivalent to the hypotheses (2.2), where θ_i are the components of $\theta(x)$. We note that if H_0 is rejected for fixed i , the wavelet coefficient d_i is retained in the model. Otherwise, it is set to zero. Thus, in the wavelet domain the problem of testing equality of two response curves is in fact the wavelet thresholding problem discussed earlier in §1. Suppose we consider the universal threshold $\lambda = \sigma\sqrt{2\log n}$ as a critical value of the test, then it can be shown that the level of significance α can be expressed as

$$\alpha = P(|d_i| > \lambda | H_0) = 2\Phi\left(-\frac{\lambda}{\sigma}\right) \approx \left(n\sqrt{\pi \log n}\right)^{-1}. \quad (2.5)$$

The critical value λ_k for testing the hypotheses (2.4) can then be obtained in two steps: (1). For each d_i and its two-sided p -value, p_i in testing $H_0 : \theta_i = 0$, $p_i = 2[1 - \Phi(|d_i|/\sigma)]$. (2). Order the p_i according to their size, $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(n)}$. Find $k = \max\{i | p_{(i)} < (i/n) \cdot \alpha\}$. For this k calculate $\lambda_k = \sigma\Phi^{-1}(1 - p_{(k)}/2)$. Following Benjamini and Hochberg¹ we refer to this approach for obtaining the critical value as the false discovery rate approach.

2.2. The F_1 test

In this section, we use the MRA and the well known general linear regression approach to develop a test which can be applied if the number of observations are not the same and the input points at which the observations were measured are distinct. The procedure also allow for repeated measurements at some input points.

Following our discussion on MRA in §1, the model (2.1) can be expressed as a general linear regression model given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \boldsymbol{\varepsilon}, \quad \text{where} \quad \text{Var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_\varepsilon^2 \mathbf{I}_{n_1} & \mathbf{0} \\ \mathbf{0} & \sigma_\nu^2 \mathbf{I}_{n_2} \end{pmatrix}. \quad (2.6)$$

In the representation (2.6),

$$\boldsymbol{\tau} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{Q}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_2 \end{pmatrix}, \quad \text{where} \quad \mathbf{Q}_1 = \begin{pmatrix} \mathbf{q}_m^T(x_{11}) \\ \mathbf{q}_m^T(x_{21}) \\ \dots \\ \dots \\ \mathbf{q}_m^T(x_{n_{11}}) \end{pmatrix}, \quad \text{and} \quad \mathbf{Q}_2 = \begin{pmatrix} \mathbf{q}_m^T(x_{12}) \\ \mathbf{q}_m^T(x_{22}) \\ \dots \\ \dots \\ \mathbf{q}_m^T(x_{n_{22}}) \end{pmatrix} \quad (2.7)$$

Let A be a well-defined $2^{m+1} \times 2^{m+2}$ matrix with elements -1, 0 and 1. Then, H_0 in (2.2) can be written as $H_0 : \beta_1 = \beta_2$ or $H_0 : \mathbf{A}\boldsymbol{\tau} = \mathbf{0}$. Define $\mathbf{y}^* = \hat{\boldsymbol{\Sigma}}^{-1/2}\mathbf{y}$, $\mathbf{X}^* = \hat{\boldsymbol{\Sigma}}^{-1/2}\mathbf{X}$, and $\boldsymbol{\varepsilon}^* = \mathbf{y}^* - \mathbf{X}^*\hat{\boldsymbol{\tau}}$. From the theory of linear models, it can be shown that an approximate test statistic for H_0 is

$$F_1 = \frac{(n-k)(\mathbf{A}\hat{\boldsymbol{\tau}})^T [\mathbf{A}(\mathbf{X}^{*T}\mathbf{X}^*)^{-1}\mathbf{A}^T]^{-1}(\mathbf{A}\hat{\boldsymbol{\tau}})}{p\boldsymbol{\varepsilon}^{*T}\boldsymbol{\varepsilon}^*}, \quad k = 2^{m+2}, \quad p = 2^{m+1}, \quad (2.8)$$

and that the approximate null distribution of F_1 is $F_{p,(n-k)}$. The F_1 test can also be used when the design points in both groups are the same and the number of

observations are equal. In that case, the multiresolution expansion is applied to $\mathbf{h}(x)$ in (2.3) and the model written as $u_{ij} = \mathbf{q}_m^T(x_i)\beta + e_{ij}$, $i = 1, \dots, L$, $j = 1, \dots, r_i$. Here, \mathbf{A} becomes the identity matrix, $\tau = \beta$, and $k = p = 2^{m+1}$.

2.3. A simulation study of the FDR and F_1 procedures

We note that when $n_1 = n_2 = n$, and $x_{ij} = x_i$ for all j , and H_0 is true, we have that $h(x) = 0$ for the FDR and F_1 tests. In this case, the difference between the data from the two groups $u_i = h(x) + e_i$ is simply noise. It follows that specific expressions for $f_1(x)$ and $f_2(x)$ are not required when generating data needed to assess the size of the tests. Thus, to study the size of the tests we set $h(x) = 0$ and generate observations u_i of length n following (2.3). We then used the Haar, Daub2 and Daub4 wavelet filters to construct three discrete wavelet transformation matrices needed for the FDR test. These matrices were applied to the data vector \mathbf{u} to construct the wavelet coefficients. Our next step was to compute the critical value λ_k and determine whether to reject H_0 . For the F_1 test, the data vector \mathbf{u} was used to estimate τ by weighted least squares (see Ref. 20). This procedure was repeated 1000 times and the proportion of rejections were evaluated. In order to examine the effect of sample size, the computations were performed for various combinations of $n = 8, 16$, and 32 , $(\sigma_1, \sigma_2) = (0.5, 0.5), (1, 0.5), (1, 1)$ and levels of significance $\alpha = 0.01, 0.025$ and 0.05 . The input points used for generating the data were chosen to be equally spaced from 0.001 to 0.999 . Some results obtained under the Haar wavelet representation are shown in Table 1.

Table 1: Simulated size for FDR test based on 1000 replications and the $m = 0$ Haar wavelet with $\epsilon_i \sim N(0, \sigma_\epsilon^2)$ and $\nu_j \sim N(0, \sigma_\nu^2)$ at significance levels $\alpha = 0.01, 0.025$ and 0.05 .

sample size	α	$(\sigma_\epsilon^2, \sigma_\nu^2) = (0.5, 0.5)$			$(\sigma_\epsilon^2, \sigma_\nu^2) = (1, 0.5)$			$(\sigma_\epsilon^2, \sigma_\nu^2) = (1, 1)$		
		0.01	0.025	0.05	0.01	0.025	0.05	0.01	0.025	0.05
8		.071	.097	.137	.083	.111	.130	.085	.094	.105
16		.031	.043	.058	.028	.047	.052	.024	.038	.053
32		.013	.024	.046	.015	.025	.050	.017	.026	.046

Simulated Size of F_1 test based on 1000 replications and the $m = 0$ Haar wavelet model with $\epsilon_i \sim N(0, \sigma_\epsilon^2)$ and $\nu_j \sim N(0, \sigma_\nu^2)$ at significance levels $\alpha = 0.01, 0.025$ and 0.05 .

sample size	α	$(\sigma_\epsilon^2, \sigma_\nu^2) = (0.5, 0.5)$			$(\sigma_\epsilon^2, \sigma_\nu^2) = (1, 0.5)$			$(\sigma_\epsilon^2, \sigma_\nu^2) = (1, 1)$		
		0.01	0.025	0.05	0.01	0.025	0.05	0.01	0.025	0.05
8		.009	.028	.052	.011	.024	.042	.011	.030	.044
16		.010	.020	.059	.010	.025	.049	.007	.019	.043
32		.011	.025	.057	.011	.027	.059	.006	.027	.049

Data for examining the power of the FDR and F_1 tests were generated by specifying expressions for the difference $h(x) = f_1(x) - f_2(x)$. We examined the sensitivity of the test to differences of small magnitude by considering the following functions:

8 *P. Guo and A. Oyet*

(1). $h_1(x) = (a) \sqrt{1.5}$, (b) 2.5; (2). $h_2(x) = (a) 3(x - 0.4)^2$, (b) $3(x - 0.4)^2 + 1.2$, (c) $3(x - 0.4)^2 + 2$; (3). $h_3(x) = 4 \sin 4\pi x - \text{sgn}(x - 0.3) - \text{sgn}(0.72 - x)$.

Table 2: Power simulations for FDR test under 1000 replications and the Haar wavelet with $\epsilon_i \sim N(0, \sigma_\epsilon^2)$ and $\nu_j \sim N(0, \sigma_\nu^2)$ at Significance level $\alpha = 0.01, 0.025, 0.05$.

$h(x)$	n	α	$(\sigma_\epsilon^2, \sigma_\nu^2) = (0.5, 0.5)$			$(\sigma_\epsilon^2, \sigma_\nu^2) = (1, 0.5)$			$(\sigma_\epsilon^2, \sigma_\nu^2) = (1, 1)$		
			0.01	0.025	0.05	0.01	0.025	0.05	0.01	0.025	0.05
$h_1(a)$	8		.636	.689	.732	.521	.530	.589	.401	.437	.493
	16		.878	.921	.953	.706	.766	.819	.559	.651	.674
	32		1.00	.997	1.00	.957	.974	.983	.870	.914	.942
$h_1(b)$	8		.992	.998	.999	.948	.962	.984	.885	.930	.938
	16		1.00	1.00	1.00	1.00	1.00	1.00	.995	.999	.998
	32		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$h_2(a)$	8		.125	.150	.165	.110	.137	.167	.117	.115	.145
	16		.073	.098	.157	.058	.070	.127	.049	.076	.094
	32		.090	.108	.164	.064	.082	.128	.031	.067	.081
$h_2(b)$	8		.808	.827	.865	.639	.693	.775	.520	.562	.688
	16		.979	.986	.990	.882	.916	.937	.760	.827	.867
	32		1.00	1.00	1.00	.997	1.00	1.00	.982	.982	.993
$h_2(c)$	8		.980	.991	.994	.915	.947	.974	.856	.896	.912
	16		1.00	1.00	1.00	.998	.999	1.00	.991	.999	.999
	32		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
h_3	8		.000	.000	.000	.000	.000	.001	.000	.001	.001
	16		.886	.962	.991	.817	.891	.971	.742	.829	.908
	32		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

2.3.1. Size and power of FDR test

The result in Table 1 show that the performance of the FDR test in controlling the size improves as the number of observations increases. For instance, the size of the FDR test reduces from 0.094 to 0.026 as the number of observations increases from 8 to 32 at 2.5% level of significance and $(\sigma_1^2, \sigma_2^2) = (1, 1)$. This pattern can be observed in all combinations of $\sigma_1^2, \sigma_2^2, \alpha$, and n . The power of the FDR test, shown in Table 2 also improves with an increase in the sample size. As an example, with $(\sigma_1^2, \sigma_2^2) = (1, 0.5)$ and 1.0% level of significance, the power of the test goes from 52.1% to 70.6% and then to 95.7% as the sample size increases from 8 to 16 and then to 32 respectively.

Furthermore, the magnitude of the error variances appear to have some effect on the power shown in Table 2. More specifically, the power of the test reduces as the error variance becomes larger. The reduction is, however, minimal when the sample size is large. For instance, consider the case where the difference is constant denoted

by $h_1(a)$ and representing the case in which $h(x) = \sqrt{1.5}$. When $n = 8$, $\alpha = 5\%$, and $(\sigma_1^2, \sigma_2^2) = (0.5, 0.5)$ the power was 73.2%; whereas the power reduces to 49.3% when the magnitude of the error variances are increased to $(\sigma_1^2, \sigma_2^2) = (1, 1)$. However, at the same 5% level of significance and same increases in error variances, the power reduces slightly from 100% to 94.2% when $n = 32$. We also note that the larger the magnitude of the difference $h(x)$, the higher the power. This can be clearly seen in Table 2 by comparing the power for the difference functions $h_1(a)$ and $h_1(b)$ and by also comparing the power under the functions $h_2(a)$, $h_2(b)$ and $h_2(c)$ for a fixed level of significance and fixed error variances.

Table 3: Power of F_1 test under 1000 replications and the Haar ($m = 0$), Daub2(*) ($m = 2$) and Daub4(**) ($m = 2$) wavelet with $\epsilon_i \sim N(0, \sigma_\epsilon^2)$, $\nu_j \sim N(0, \sigma_\nu^2)$ and $\alpha = 0.01, 0.025$, and 0.05 .

$h(x)$	n	α	$(\sigma_\epsilon^2, \sigma_\nu^2) = (0.5, 0.5)$			$(\sigma_\epsilon^2, \sigma_\nu^2) = (1, 0.5)$			$(\sigma_\epsilon^2, \sigma_\nu^2) = (1, 1)$		
			0.01	0.025	0.05	0.01	0.025	0.05	0.01	0.025	0.05
$h_1(a)$	8		.293	.492	.669	.159	.324	.510	.129	.224	.362
	16		.897	.960	.981	.681	.802	.903	.543	.689	.814
	32		1.00	1.00	1.00	.989	.998	1.00	.948	.981	.991
$h_1(b)$	8		.943	.991	.997	.802	.926	.988	.658	.841	.926
	16		1.00	1.00	1.00	1.00	1.00	1.00	.999	1.00	.998
	32		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$h_2(a)$	8		.013	.041	.098	.016	.035	.085	.013	.026	.068
	16		.044	.100	.165	.034	.063	.118	.027	.060	.114
	32		.144	.202	.305	.066	.169	.234	.046	.110	.180
$h_2(b)$	8		.449	.670	.823	.308	.498	.684	.210	.360	.553
	16		.980	.990	1.00	.865	.941	.984	.738	.867	.940
	32		1.00	1.00	1.00	.999	1.00	1.00	.993	.999	.999
$h_2(c)$	8		.881	.957	.991	.720	.884	.950	.566	.779	.882
	16		1.00	1.00	1.00	1.00	1.00	1.00	.994	.997	1.00
	32		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
h_3	8		.000	.000	.010	.003	.005	.017	.003	.015	.040
	16		.828	.973	.994	.719	.898	.967	.613	.820	.924
	32		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
h_3^*	16		.880	.967	.995	.762	.910	.966	.633	.810	.924
	32		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
h_3^{**}	16		.886	.976	.994	.727	.920	.975	.641	.807	.933
	32		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

2.3.2. Size and power of the F_1 test

By comparing the results in Table 1, we see that the F_1 test offers some improvement over the FDR test in controlling the size for small values of n . However, our results show that the power performances of the FDR and F_1 tests are comparable. When

the structure of the difference function $h(x)$ is simple (e.g. constants and quadratic functions), the power of the F_1 test tend to be higher with the $m = 0$ Haar wavelet &ter; whereas when the difference function is more complicated (e.g. $h_3(x)$), the power tends to be higher under the $m = 2$ Daub2 wavelet and Daub4 wavelets as shown in Table 3.

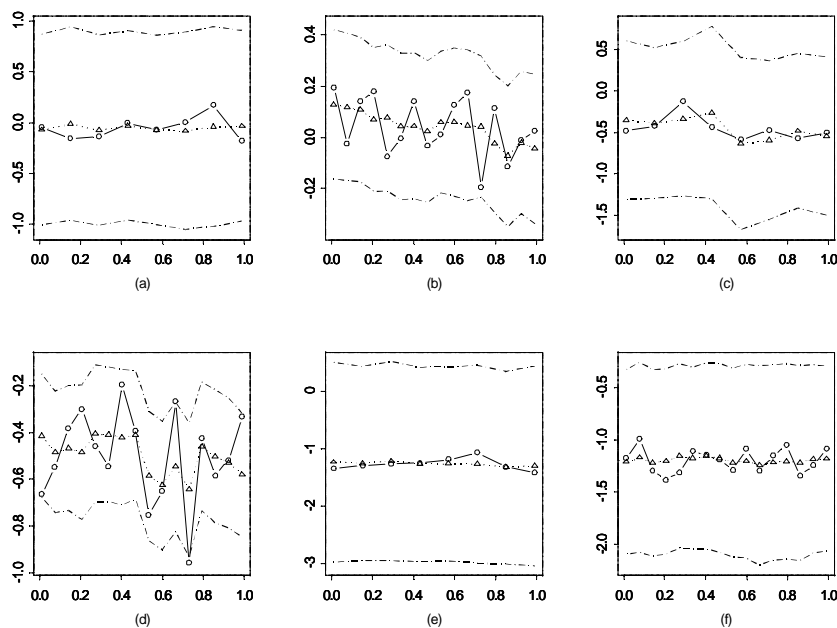


Fig. 1. A plot of the data $D_i = Y_i - Z_i$ $i = 1, \dots, n$ (solid line), Haar wavelet estimate $\hat{D} = \hat{h}(x) = f - g(x)$ (dotted line), and approximate confidence bounds (dotted broken line) for difference $h(x) = f(x) - g(x)$ with parameters (a), (b) $n = 8$ and 16 , respectively, with $\alpha = 0.05$, $(\sigma_\epsilon, \sigma_\nu) = (1.6, 2.5)$ and $h(x) = 0$; (c), (d) $n=8$ and 16 , respectively, with $\alpha = 0.025$, $(\sigma_\epsilon, \sigma_\nu) = (3.5, 0.8)$, and $h(x) = -0.5$; (e), (f) $n = 8$ and 16 , respectively with $\alpha = 0.01$, $(\sigma_\epsilon, \sigma_\nu) = (2, 2)$, and $h(x) = -\sqrt{1.5}$.

If the input points are distinct and the number of observations are not equal, the size of the F_1 test can be examined by specifying only one of the functions $f_1(x)$, $f_2(x)$ since under H_0 , $f_1(x) = f_2(x)$. We studied the size by generating data of size $(n_1, n_2) = (12, 16)$ and $(n_1, n_2) = (16, 20)$ using each of the functions, $f(x) = 0$, $f(x) = h_2(x)(a)$, and $f(x) = h_3(x)$, de&ned in §2.3. The input points x_{ij} and u_{lk} were chosen to be equally spaced from 0.001 to 0.999. Using the Haar wavelet with $m = 1$ and the Daub4 wavelet with $m = 3$, we found that the F_1 test performs well in controlling the size of the test when $\sigma_\epsilon^2 = \sigma_\nu^2$; whereas when $\sigma_\epsilon^2 \neq \sigma_\nu^2$, the test

tends to overestimate the size. The significance levels considered in the simulation study were $\alpha = 0.01, 0.025, \text{ and } 0.05$. We further examined the sensitivity of the test to large and small differences between $f_1(x)$ and $f_2(x)$. We found that both the magnitude of the difference and the complexity of the structure of the difference function has effect on the power of the test.

2.4. A graphical method

Recall that the FDR test requires using \mathbf{d}^{J-1} to estimate σ^2 which is then used in computing the critical value λ_k . For small samples, the estimate of σ_ξ^2 tend to be highly biased. The bias in $\hat{\sigma}^2$ then leads to inaccurate conclusions in most of the cases. When n is small we found that using estimates of $h(x)$ and confidence bounds computed from (1.3) produced more reliable results. The main idea is to compute an estimate $\hat{h}(x)$ of the difference function $h(x)$ as in (1.1) and construct a confidence bound for $h(x)$ as in (1.3). Our simulation results show that if $\hat{h}(x)$ is approximately zero for all x and the confidence bound is approximately symmetric about the line $h(x) = 0$, we cannot reject the null hypothesis H_0 .

Figures 1(a) and (b) show that when $h(x) = 0$ the confidence bound covers zero and is approximately symmetric about the line $h(x) = 0$ with some distortion when the error variances σ_ε^2 and σ_ν^2 are large. Figures 1(c) - (f) provide a good illustration of the fact that when $h(x)$ is non-zero but a constant for all x , the estimate $\hat{h}(x)$ is close to the true $h(x)$ and the confidence bound is not symmetric about the line $h(x) = 0$. However, if the sample size is small, $n = 8$ and $h(x) = \text{constant}$, the confidence bound may still cover the line $h(x) = 0$ (see Figures 1(c), (e)). Figures 1(d) and (f) then show that with a sample size of $n = 16$ the confidence bound will exclude the zero line. These patterns were repeated when $h(x)$ is a function rather than a constant. In this case, we used the Daubechies wavelet for estimating the functions instead of the Haar wavelet because it is more suitable for estimating smooth functions.

3. Extension To $k > 2$ Groups

When $k \geq 2$, a preliminary step in examining whether $f_1 = f_2 = \dots = f_k$ is to estimate the error variance σ_j^2 and response curve f_j for each of the k groups. We can then plot and visually assess the shape of the graph of the estimated response curves \hat{f}_j , $j = 1, \dots, k$. If $x_{ij} = x_i$, for all j , which is usually the situation in most cases, then the graph of the estimated response functions should provide some indication of differences or equality as in Figures 2(a), 2(b). The data used in constructing Figures 2(a)-(d) was generated from (2.1) with (a) $f_1 = \dots = f_4 = h_3(x)$ of §2.3, (b) $f_1 = f_3 = h_3(x)$, $f_2 = f_4 = t(x) = 3 * \min\{2x, -2(x-1)\}$, (c) $f_1 = \dots = f_4 = f_5 = h_3(x)$ and (d) $f_1 = f_2 = f_5 = t(x) = 3 * \min\{2x, -2(x-1)\}$, $f_3 = f_4 = d(x) = 5(x-0.4)^2 + 1.2$ where $\sigma_1 = \sqrt{5.5}$, $\sigma_2 = 1$, $\sigma_3 = 3$, and $\sigma_4 = \sqrt{1.5}$. The graph is the average from 5000 simulations of the estimated functions using

12 *P. Guo and A. Oyet*

(1.1) and the Daub4 wavelet in an Splus code. It is clear that the functions in Figure 2(a) are the same, whereas the functions in Figure 2(b) are not the same. If in addition to $x_{ij} = x_i$ we also have that $n_j = n$ for all j , then we can compute $z_{ij} = y_{ij} - y_{i,j+1}$, $j = 1, 2, \dots, k - 1$. Let $h_j(x) = f_j(x) - f_{j+1}(x)$ and define,

$$T_n(x) = \sum_{j=1}^{k-1} |h_j(x)|. \tag{3.1}$$

By (2.2), $h_j(x) = 0$, $j = 1, \dots, k - 1$, implying that $T_n(x) = 0$ at each x . We then use the data z_{ij} to obtain an interval estimate of $T_n(x)$ as in (1.1) and (1.3) where the standard error of $\hat{T}_n(x)$ is $s_{T_n} = \sqrt{\sum_{j=1}^{k-1} s_j^2 / (k - 1)}$ and s_j is the standard error of \hat{h}_j defined in (1.3). We illustrate this technique in Figures 2(c), 2(d). We found that if the response functions are not the same but the values of the difference is close to zero, the confidence intervals will still cover zero but may not be approximately symmetric about the line $T_n(x) = 0$ as in Figure 2(d).

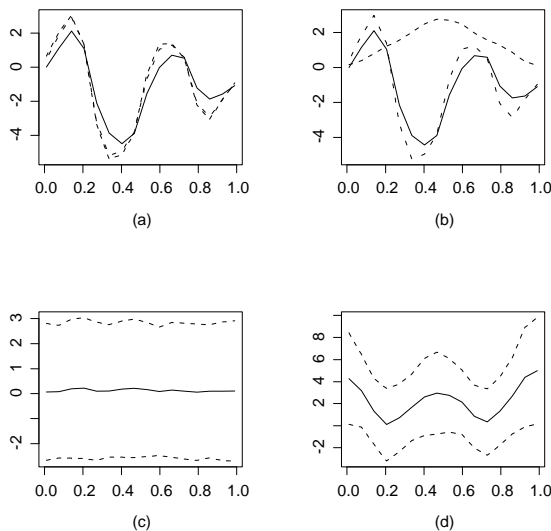


Fig. 2. A plot of (a) estimated functions $\hat{f}_1(x) = \hat{f}_2(x) = \hat{f}_3(x) = \hat{f}_4(x) = \hat{h}_3(x)$ (b) $\hat{f}_1(x) = \hat{f}_2(x) = \hat{h}_3(x)$ and $\hat{f}_3(x) = \hat{f}_4(x) = \hat{t}(x)$. Graph of $\hat{T}_n(x)$ with approximate confidence bounds when (a) $f_c(x) = h_3(x)$ and (b) $f_1 = f_2 = f_5 = t(x)$, $f_3 = f_4 = d(x)$.

In the general case, we begin by computing $v_{ij} = y_{ij}/\sigma_j$ and merge the data v_{ij} from all groups. Let r be the number of distinct x_i values in the merged data and t_i the total number of repeated observations v_{il} corresponding to each x_i . Now, we

observe that if H_0 is true, we can find a common function, say $f_c(x)$, such that $f_1(x) = f_2(x) = \dots = f_k(x) = f_c(x)$. Then, using the combined data we can estimate $f_c(x)$ as $\hat{f}_c(x) = \mathbf{q}^T(x)\hat{\beta}_c$, where $\hat{\beta}_c = [(1/n) \sum_{i=1}^r t_i w_i \mathbf{q}(x_i) \mathbf{q}^T(x_i)]^{-1} [(1/n) \sum_{i=1}^r \sum_{l=1}^{t_i} w_i \mathbf{q}(x) v_{il}]$ and the weights w_i are the minimum variance unbiased weights discussed in Ref. 20. For fixed j , define $e_{ij} = y_{ij} - \hat{f}_j$, $e_{c,ij} = y_{ij} - \hat{f}_{c,j}$, $R_j = \sum_{i=1}^{n_j} w_i e_{ij}^2 = \mathbf{e}_j^T \mathbf{W} \mathbf{e}_j$ and $R_j^c = \sum_{i=1}^{n_j} w_i e_{c,ij}^2 = \mathbf{e}_{c,j}^T \mathbf{W} \mathbf{e}_{c,j}$, where $\hat{f}_j(x_{ij}) = \mathbf{q}^T(x_{ij})\hat{\beta}_j$ and $\hat{f}_{c,j}(x_{ij}) = \mathbf{q}^T(x_{ij})\hat{\beta}_{c,j}$ are estimated response functions for group j . Here, $\hat{\beta}_{c,j}$ is the component of $\hat{\beta}_c$ corresponding to x_{ij} . Suppose that for fixed j , $y_{ij} \sim N(0, \sigma_j^2)$. Then, it can be shown that the distribution of $W^{1/2} \mathbf{e}_j$ is a degenerate normal distribution with mean 0 and singular covariance matrix $\Sigma = W^{1/2}(\mathbf{I}_{n_j} - \mathbf{H}_w)(\mathbf{I}_{n_j} - \mathbf{H}_w)^T W^{1/2}$, where $\mathbf{H}_w = \mathbf{Q}(\mathbf{Q}^T \mathbf{W} \mathbf{Q})^{-1} \mathbf{Q}^T \mathbf{W}$ and the rank of Σ is $n_j - 2^{m+1}$.

Table 4: $(1 - \alpha)100\%$ Quantiles of the distributions of T_1 and T_2 for comparing $k = 4$ functions with the $m = 0$ Daub2 wavelet model obtained from 10,000 simulations

(n_1, n_2, n_3, n_4)	statistic	$(1 - \alpha)100\%$				
		90	95	97.5	98	99
8	T_1	0.2175	0.3088	0.4111	0.4416	0.5556
	T_2	0.7634	0.9051	1.0231	1.0659	1.1773
10	T_1	0.1637	0.2311	0.3102	0.3303	0.4120
	T_2	0.6561	0.7803	0.8927	0.9273	1.0329
15	T_1	0.0874	0.1254	0.1685	0.1827	0.2256
	T_2	0.4816	0.5711	0.6583	0.6811	0.7704
30	T_1	0.0249	0.0354	0.0470	0.0499	0.0641
	T_2	0.2578	0.3053	0.3518	0.3649	0.4074
64	T_1	0.0057	0.0085	0.0111	0.0120	0.0151
	T_2	0.1251	0.1492	0.1699	0.1763	0.1969
120	T_1	0.0017	0.0024	0.0031	0.0035	0.0042
	T_2	0.0675	0.0794	0.0917	0.0954	0.1061
200	T_1	0.0006	0.0008	0.0011	0.0012	0.0015
	T_2	0.0408	0.0478	0.0545	0.0570	0.0629
(8,10,10,15)	T_1	0.0521	0.0727	0.0971	0.1045	0.1356
	T_2	0.3733	0.4392	0.5058	0.5287	0.5991
(15,30,30,64)	T_1	0.0030	0.0043	0.0060	0.0065	0.0084
	T_2	0.0885	0.1044	0.1213	0.1273	0.1400
(64,120,64,120)	T_1	0.0007	0.0010	0.0014	0.0015	0.0019
	T_2	0.0455	0.0531	0.0610	0.0633	0.0709

Clearly, $R_j^c - R_j = 0$ when H_0 is true and $R_j^c - R_j \neq 0$ otherwise. It follows that,

two test statistic that can be used for testing H_0 are

$$T_1 = \sum_{j=1}^k (R_j^c - R_j)^2 \text{ and } T_2 = \sum_{j=1}^k |R_j^c - R_j|. \quad (3.2)$$

As an illustration, in 10000 simulations we found that when $n_1 = n_3 = 8, n_2 = n_4 = 10$ and $f_1 = f_2 = f_3 = f_4 = h_3(x)$, the average values of R_j and R_j^c were $R_1 = 3.343, R_1^c = 3.413, R_2 = 4.079, R_2^c = 4.145, R_3 = 3.324, R_3^c = 3.394, R_4 = 4.100, R_4^c = 4.165$. However, with $f_1 = d(x), f_2 = f_3 = f_4 = h_3(x)$, the average values of R_j and R_j^c become $R_1 = 1.591, R_1^c = 3.043, R_2 = 4.079, R_2^c = 4.272, R_3 = 3.324, R_3^c = 3.531, R_4 = 4.100, R_4^c = 4.292$. Clearly, the large difference in the values of R_1 and R_1^c shows that f_1 is different from f_2, f_3 , and f_4 . Furthermore when we replace f_3 in the above computation with the triangular function $t(x)$, the average values become $R_1 = 1.591, R_1^c = 2.469, R_2 = 4.079, R_2^c = 4.634, R_3 = 0.617, R_3^c = 1.384, R_4 = 4.100, R_4^c = 4.655$. Again, we see that the difference in the response curves are reflected in the values of R_1, R_1^c and R_3 and R_3^c . In general, we found that in cases where H_0 is not true and the values of the functions at the input points are not close in magnitude, the difference will be clearly reflected in the computed values of R_j and R_j^c . A careful examination of these values will also lead to identification of the functions that are different. Such differences become much more evident when the sample size is large.

Table 5: Power of T_1 and T_2 in 10,000 simulations under the $m = 0$ Daub2 wavelet model with $\alpha = 0.05$

Functions	Statistic	$(n_1, n_2), (n_3, n_4)$					
		10	64	200	(8,10)	(15,30)	(64,120)
$f_1 = f_2 = f_4 = h_3(x)$	T_1	13.41	76.48	99.97	11.21	42.91	84.09
$f_3 = h_3(x) - 0.5$	T_2	13.36	74.02	99.95	10.82	41.87	78.33
$f_1 = f_2 = f_4 = h_3(x)$	T_1	88.70	100	100	85.60	100	100
$f_3 = h_3(x) - 1.5$	T_2	86.70	100	100	82.54	100	100
$f_1 = f_2 = f_4 = h_3(x)$	T_1	100	100	100	99.99	100	100
$f_3 = h_3(x) - 2.5$	T_2	99.97	100	100	99.99	100	100
$f_1 = f_2 = f_4 = d(x)$	T_1	56.26	99.04	100	49.56	81.91	99.55
$f_3 = t(x)$	T_2	53.46	98.63	100	46.21	79.33	99.08

3.1. Quantiles and power of T_1 and T_2

Given that the sampling distributions of T_1 and T_2 are unknown, it is possible to obtain the quantiles of their distributions by simulation. An example of the quantiles of the distributions of T_1 and T_2 for comparing four functions we obtained from 10000 simulations are shown in Table 4 for various values of n_1, n_2, n_3, n_4 . The quantiles were obtained by generating data $y_{ij}, i = 1, \dots, n_j; j = 1, 2, 3, 4$ from

(2.1) under the null hypothesis with $f_c(x) = h_3(x)$ or $d(x)$ or $t(x)$ after which we computed the values of T_1 and T_2 based on the $m = 0$ Daub2 wavelet model. We repeated the computations 10,000 times using an Splus code and then computed the quantiles in Table 4 from the 10,000 values of T_1 and T_2 . Using the 95% quantile we computed the power of the tests shown in Table 5. The values in Table 5 show that if the difference between the values of the functions is small, the size of the data has to be large for the power of the tests to be high.

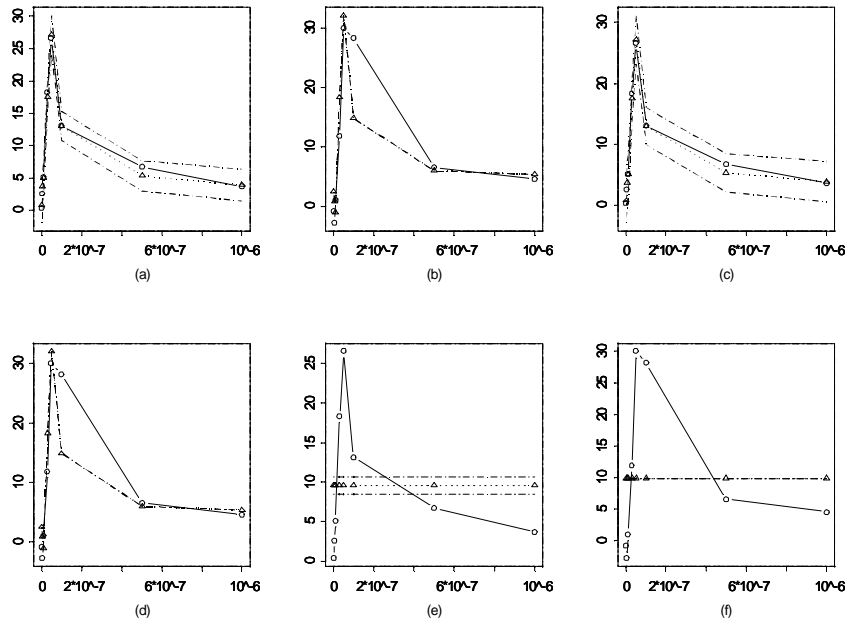


Fig. 3. A plot of the difference $h(x)$ (solid line), Daubechies wavelet estimate of difference $\hat{h}(x)$ (dotted line) and confidence intervals for methacholine CRC data in mesenteric arteries for (a), (b) 4-month and 7-month data, respectively with 95% CI; (c), (d) 4-month and 7-month data, respectively with 99% CI. Haar wavelet estimate of $h(x)$ (dotted line) for (e), (f) 4-month and 7-month data, respectively with 95% CI.

4. Case Study: Methacoline CRC In Mesenteric Arteries

In this section, we apply the graphical method to dose response measurements for vascular relaxation in the absence and presence of a nitric oxide inhibitor from 4-month and 7-month old male rats. The pairs of observations from both the 4-month and 7-month old rats are of size $n = 8$ and measured at the same concentration x_i , $i = 1, 2, \dots, 8$. In our simulation studies, we found that the graphical method was

more reliable when the sample size is small. The graph of the difference function $h(x)$, $\hat{h}(x)$ and the confidence bands shown in Figure 3 were constructed using the Daub2 wavelet. The graphs show that the response curves generating the data are not the same since the confidence bands do not cover zero. Furthermore, Figures 3(e) and 3(f) indicates that even if the difference $h(x)$ were to be constant for all x , it is nonzero. Consequently, based on the data shown in Figure 3 the decision is to reject H_0 .

5. Concluding Remarks

Previously, we have used data from a normal population in our simulation studies. We now discuss the effect of deviations of the error distribution from normality on the size and power performances of the FDR and F_1 tests. We also comment on the results of comparing these two tests to the kernel approach of King *et al.*¹². In generating the data for examining robustness of the tests, we use the functions defined in §2.3 and replace the error distribution by one of the following: (a) $\varepsilon_{ij} \sim \frac{1}{\sqrt{2}}\sigma_\varepsilon \cdot t_4$, $\nu_{lk} \sim \frac{1}{\sqrt{2}}\sigma_\nu \cdot t_4$; (b) $\varepsilon_{ij} \sim \frac{1}{\sqrt{2}}\sigma_\varepsilon \cdot (\chi_1^2 - 1)$, $\nu_{lk} \sim \frac{1}{\sqrt{2}}\sigma_\nu \cdot (\chi_1^2 - 1)$. These distributions were chosen so that the error variances will be the same as those used under normal error assumption.

We found that when data is generated from a t -distribution with 4 degrees of freedom, the size and power performances of the FDR test were identical to the size and power of the FDR test under normal errors. When the error distributions are from a skewed χ^2 distribution, the FDR test performs poorly in controlling the size of the test. However, the power of the test remains at approximately the same level. Unlike the FDR test, the F_1 test was found to be robust against deviations from normality to both the t -distribution and the χ^2 distribution in terms of size and power when the design points are the same. When the design points are not the same, the power of the F_1 test appear to be unaffected by the change in distribution. However, the test appear to fail in controlling the size of the test once the distribution is changed to the skewed χ^2 distribution.

The results of the comparison between the FDR and F_1 tests and the kernel approach show that (a) the F_1 test and the kernel method in Ref. 12 performs better in controlling the size than the FDR test, and (b) the power of both the FDR and F_1 tests are comparable to the power of the kernel method. Now, the null distribution of the test statistic under the kernel approach is unknown, so a simulated p-value is needed for the test. King *et al.*¹² recommended 8000 simulations. Attempts at computing these p-values using an Splus code highlighted the inefficiency of this approach in terms of the number of days it took to complete 1000 simulations. This appear to be a weakness of the kernel approach. In terms of robustness, we found that the kernel approach also failed in controlling the size of the test under the skewed χ^2 distribution. The effect of the t -distribution appear to be minimal. But

the reduction in power of the test based on the kernel approach appear to be much higher than the reduction in both the FDR and F_1 tests. When the sample size is small, $n = 8$, the power of the kernel based method was found to be very low. Thus, for small samples, the method will also be highly unreliable. In this case, we have shown that the graphical method we have proposed is more reliable.

Acknowledgement

This research was supported by the Natural Sciences and Engineering Research Council of Canada.

References

- (1) Y. Benjamini and Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society. Series B*, **57** (1995) 289-300.
- (2) A. Bowman and S. Young, Graphical comparison of nonparametric curves, *Applied Statistics*, **45**(1) (1996) 83-98.
- (3) D. R. Brillinger, Uses of cumulants in wavelet analysis, *Technical report*, Department of Statistics, University of California at Berkeley (1995).
- (4) I. Daubechies, *Ten Lectures on Wavelets*, Number 61 in CBMS-NSF Series in applied mathematics, Society for Industrial and Applied Mathematics, Philadelphia (1992).
- (5) I. Daubechies and J. C. Lagarias, Two-scale difference equations. II. Local regularity, infinite products of matrices and fractals, *SIAM Journal on Mathematical Analysis*, **23**(4) (1992) 1031-1079.
- (6) M. A. Delgado, Testing the equality of nonparametric regression curves, *Statistics and Probability Letters*, **17** (1993) 199-204.
- (7) H. Dette and N. Neumeyer, Nonparametric analysis of covariance, *The Annals of Statistics*, **29**(5) (2001) 1361-1400.
- (8) T. Görgens, Nonparametric comparison of regression curves by local linear testing, *Statistics and Probability Letters*, **60** (2002) 81-89.
- (9) P. Hall and J. D. Hart, Bootstrap test for difference between means in nonparametric regression, *Journal of the American Statistical Association*, **85** (1990) 1039-1049.
- (10) P. Hall, C. Huber and P. L. Speckman, Covariate-matched one-sided tests for the difference between functional means, *Journal of the American Statistical Association*, **92** (1997) 1074-1083.
- (11) W. Härdle and J. S. Marron, Semiparametric comparisons of regression curves, *The Annals of Statistics*, **18** (1990) 63-89.
- (12) E. C. King E. C., J. D. Hart and T. E. Wehrly, Testing the equality of two regression curves using linear smoothers, *Statistics and Probability Letters*, **12** (1991) 239-247.

- (13) H. L. Koul and A. Schick, Testing by superiority among two regression curves, *Journal of Statistical Planning and Inference*, **117** (2003) 15-23.
- (14) K. B. Kulasekera, Comparison of regression curves using quasi-residuals, *Journal of the American Statistical Association*, **90** (1995) 1085-1093.
- (15) K. B. Kulasekera and J. Wang, Smoothing parameterselection for power optimality in testing regression curves, *Journal of the American Statistical Association*, **92** (1997) 500-511.
- (16) E. A. Maharaj (2005), Using wavelets to compare time series patterns, *International Journal of Wavelets, Multiresolution and Information Processing* **3**(4) (2005) 511-521.
- (17) S. J. Marron, S. Adak, I. M. Johnstone, M. H. Neumann and P. Patil, Exact risk analysis of wavelet regression. *Journal of Computational and Graphical Statistics*, **7**(3) (1998) 278-309.
- (18) A. Munk and H. Dette, Nonparametric comparison of several regression functions: exact and asymptotic theory, *The Annals of Statistics*, **26**(6) (1998) 2339-2368.
- (19) N. Neumeier and H. Dette, Nonparametric comparison of regression curves: an empirical process approach, *The Annals of Statistics*, **31**(3) (2003) 880-920.
- (20) A. J. Oyet and B. Sutradhar, Testing variances in wavelet regression models, *Statistics and Probability Letters*, **61** (2003) 97-109.
- (21) A. Yatchew, An elementary nonparametric differencing test of equality of regression functions, *Economics Letters*, **62** (1999) 271-278.
- (22) S. Young and A. Bowman, Non-parametric analysis of covariance, *Biometrics*, **51**(3) (1995) 920-931.