*Research Article*

# Transfer Function Models with Time-Varying Coefficients

## Maria Sílvia de A. Moura,[1] Pedro A. Morettin,[2] Clélia M. C. Toloi,[2] and Chang Chiann[2]

[1] *Department of Statistics, Federal University of São Carlos, 13565-905 São Carlos, SP, Brazil*
[2] *Department of Statistics, University of São Paulo, 05508-090 São Paulo, SP, Brazil*

Correspondence should be addressed to Pedro A. Morettin, pam@ime.usp.br

We consider a transfer function model with time-varying coefficients. We propose an estimation procedure, based on the least squares method and wavelet expansions of the time-varying coefficients. We discuss some statistical properties of the estimators and assess the validity of the methodology through a simulation study. We also present an application of the proposed procedure to a real pair of series.

## 1. Introduction

In this paper, we consider the transfer function model

$$Y_{t,T} = \sum_{i=1}^{m} \delta_i\left(\frac{t}{T}\right) Y_{t-i,T} + \sum_{j=0}^{n} \omega_j\left(\frac{t}{T}\right) X_{t-j,T} + \varepsilon_t, \quad t = 1, \ldots, T, \tag{1.1}$$

where $T$ is the number of observations and $\varepsilon_t$ is i.i.d. $(0, \sigma^2)$. We assume that the error and the input series are independent. The functions $\omega_j(u)$, $j = 0, 1, 2, \ldots, n$, and $\delta_j(u)$, $j = 1, 2, \ldots, m$, are supported on the interval $[0,1]$ and connected to the underlying series by an appropriate rescaling of time, $u = t/T$.

As noted by Nason and Sapatinas [1], when both time series $X_t$ and $Y_t$ belong to the class of ARMA models, the transfer function models made popular by Box and Jenkins are appropriate. See Box et al. [2].

Model (1.1) is important because it can be used when either or both time series are not stationary, but it is perhaps more suitable when the series exhibit a locally stationary behaviour, for example, in the sense of Dahlhaus [3] or of Nason et al. [4]. In several areas, such as signal analysis (speech analysis in particular) and economics, empirical and theoretical works suggest that models with time-varying coefficients are useful, reflecting, for example, the evolution of the economy over time.

Chiann and Morettin [5, 6] considered a special case of (1.1), namely,

$$Y_{t,T} = \sum_{j=0}^{\infty} a_j\left(\frac{t}{T}\right) X_{t-j,T} + \sigma\left(\frac{t}{T}\right)\varepsilon_t, \tag{1.2}$$

while Dahlhaus et al. [7] entertained the model

$$X_{t,T} = \sum_{j=1}^{p} a_j\left(\frac{t}{T}\right) X_{t-j,T} + \sigma\left(\frac{t}{T}\right)\varepsilon_t. \tag{1.3}$$

These authors used wavelet expansions of the time-varying coefficients in order to obtain estimates.

Related references are Brockwell et al. [8, 9], who used a state-space approach to transfer function modelling, allowing for nonstationary input and output sequences, and Nason and Sapatinas [1] who show how a nondecimated wavelet packet transform (NDWPT) can be used to model a response time series $Y_t$ in terms of an explanatory time series $X_t$, both assumed to be nonstationary. Rather than building a model directly between $X_t$ and $Y_t$ some regression-type model is built between $Y_t$ and an NDWPT version of $X_t$.

We consider the problem of estimating $\omega_j(u)$, $j = 0, 1, 2, \ldots, n$ and $\delta_j(u)$, $j = 1, 2, \ldots, m$, in time domain, using wavelet expansions. We use least squares to obtain the estimators of the wavelet coefficients. Then the empirical detail coefficients are shrunk before the inverse wavelet transform is applied to obtain the final estimates of $\omega_j(u)$ and $\delta_j(u)$. This results in a nonlinear smoothing procedure. See Section 2 for further details.

We will apply the methodology to real data, namely, a pair of hourly wind speeds recorded at two Welsh Meteorological Office stations, Valley and Aberporth. Here the idea is to relate the target wind speeds (Valley) to wind speeds measured at a nearby reference site (Aberporth). See Nason and Sapatinas [1] for a complete description of this data and Section 6.

The plan of the paper is as follows. In Section 2, we introduce some background material on wavelets. In Section 3, we present the estimation procedure, and in Section 4, we give some properties of the empirical wavelet coefficients. In Section 5, we perform some simulations and in Section 6, we apply the methodology to the data described above. The paper ends with some remarks in Section 7.

## 2. Wavelets

In this section, we discuss some basic ideas on wavelets. For more details, see the books by Vidakovic [10], Percival and Walden [11], and Nason [12]. From two basic functions, the *scaling function* $\phi(x)$ and the *wavelet* $\psi(x)$, we define infinite collections of translated

and scaled versions, $\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k)$, $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$, $j,k \in \mathbb{Z} = \{0,\pm 1,\pm 2,\ldots\}$. We assume that $\{\phi_{\ell,k}(\cdot)\}_{k\in\mathbb{Z}} \cup \{\psi_{j,k}(\cdot)\}_{j\geq\ell;k\in\mathbb{Z}}$ forms an orthonormal basis of $L_2(\mathbb{R})$, for some coarse scale $\ell$. A key point [13] is that it is possible to construct compactly supported $\phi$ and $\psi$ that generate an orthonormal system and have space-frequency localization, which allows parsimonious representations for wide classes of functions in wavelet series.

In some applications, the functions involved are defined in a compact interval, such as $[0,1]$. This will be the case of our functions $\omega_j(u)$ and $\delta_j(u)$ in (1.1). So it will be necessary to consider an orthonormal system that spans $L_2([0,1])$. Several solutions were proposed, the most satisfactory one being that by Cohen et al. [14]. As in any proposal, close to the boundaries the basis functions are modified to preserve orthogonality. Cohen et al. [14] use a preconditioned step. Periodized wavelets are defined by

$$\tilde{\phi}_{j,k}(x) = \sum_{n\in\mathbb{Z}} \phi_{j,k}(x - n), \qquad \tilde{\psi}_{j,k}(x) = \sum_{n\in\mathbb{Z}} \psi_{j,k}(x - n), \tag{2.1}$$

and these generate a multiresolution level ladder $\tilde{V}_0 \subset \tilde{V}_1 \subset \cdots$, in which the spaces $\tilde{V}_j$ are generated by the $\tilde{\psi}_{j,k}$. Negative values of $j$ are not necessary, since $\tilde{\phi} = \tilde{\phi}_{0,0} = 1$ and if $j \leq 0$, $\tilde{\psi}_{j,k}(x) = 2^{-j/2}$. See Vidakovic [10] for details and Walter and Cai [15] for a different approach. Other instances are the boundary corrected Meyer wavelets [16]. From now on, we denote the periodized wavelets simply by $\psi_{j,k}$.

Accordingly, for any function $f \in L_2([0,1])$, we can expand it in an orthogonal series

$$f(x) = \alpha_{0,0}\phi(x) + \sum_{j\geq 0} \sum_{k\in I_j} \beta_{j,k}\psi_{j,k}(x), \tag{2.2}$$

where we take $\ell = 0$ and $I_j = \{k : k = 0,\ldots,2^j - 1\}$. For each $j$, the set $I_j$ brings the values $k$, such that $\beta_{jk}$ belongs to scale $2^j$. For example, for $j = 3$, we have 8 wavelet coefficients in scale $2^3$, while for $j = 2$ we have half of this number, namely, 4 coefficients in scale $2^2$. The wavelet coefficients are given by

$$\alpha_{0,0} = \int f(x)\phi(x)dx, \qquad \beta_{j,k} = \int f(x)\psi_{j,k}(x)dx. \tag{2.3}$$

Often we consider the sum in (2.2) for a maximum level $J$,

$$f(x) \simeq \alpha_{00}\phi(x) + \sum_{j=0}^{J-1} \sum_{k\in I_j} \beta_{jk}\psi_{jk}(x), \tag{2.4}$$

in such a way that we approximate $f$ in the $\tilde{V}_J$ space.

Wavelet shrinkage provides a simple tool for nonparametric function estimation. It is an active research area where the methodology is based on optimal shrinkage estimators for

the location parameters. Some references are Donoho and Johnstone [17, 18] and Donoho et al. [19]. In this paper, we focus on the simplest, yet most important shrinkage strategy-wavelet thresholding. Earlier models considered nonparametric regression with i.i.d. normal errors. These were further extended by Johnstone and Silverman [20] for correlated errors. In this context, the paper by Opsomer at al. [21] is a good survey of methods using kernels, splines, and wavelets.

The thresholding technique consists of reducing the noise included in a signal through the application of a threshold to the estimated wavelets coefficients $\widehat{\beta}_{jk}$. Some commonly used forms are the soft and hard thresholds, given by

$$\delta^{(s)}\left(\widehat{\beta}_{jk}, \lambda\right) = \left(\left|\widehat{\beta}_{jk}\right| - \lambda\right)_{+} \operatorname{sgn}\left(\widehat{\beta}_{jk}\right),$$

$$\delta^{(h)}\left(\widehat{\beta}_{jk}, \lambda\right) = \widehat{\beta}_{jk} I\left(\left|\widehat{\beta}_{jk}\right| \geq \lambda\right), \tag{2.5}$$

respectively. One crucial issue is the choice of the parameter $\lambda$. If the empirical wavelet coefficients $\widehat{\beta}_{jk}$ are assumed to be Gaussian with mean $\beta_{jk}$ and constant variance, a popular method is to apply the so-called universal threshold $\lambda = \sigma\sqrt{2\log 2^J}$, where $\sigma^2$ is the noise variance. Other methods include the SURE procedure, cross validation, and Bayesian approaches. See Percival and Walden [11] and Vidakovic [10] for details.

But despite the good properties of these procedures, which are spatially adaptive and nearly optimum over wide classes of function spaces, they are not appropriate when the design is not regular or it is random. Recently, several approaches were proposed dealing with nonregular designs. Delouille [22] introduced the concept of design adapted wavelets, and Kerkyacharian and Picard [23] introduced the warped wavelets. Cai and Brown [24, 25] consider wavelet estimation for models with nonequispaced or random uniform designs, and Sweldens [26] introduced the so-called lifting scheme; now the wavelets are no longer obtained as dilations and translations of some mother wavelet. See also Jansen and Oonincx [27]. Some applications and extensions of these approaches are given in Porto et al. [28, 29] and Morettin and Chiann [30].

In this paper, we will use ordinary wavelets, as in Dahlhaus et al. [7], since these wavelets work well for our purposes. See also a related discussion in Morettin and Chiann [30]. Concerning the choice of the mother wavelet, there are no established rules. In our experience, we agree with Nason and Sapatinas [1], when they say that "how the choice of of wavelets affects the final model and its interpretation is an area of further research." We will make use in particular of the Daubechies compactly supported wavelets in the simulations and application. Concerning thresholding, we use hard thresholds and the parameter $\lambda$ chosen as explained in Sections 4 and 5.

## 3. Estimators

The aim is to estimate the functions $\delta_i(u)$, $i = 1, \ldots, m$, and $\omega_i(u)$, $i = 0, 1, \ldots, n$, appearing in model (1.1), given $T$ observations of the series $X_{t,T}$ and $Y_{t,T}$. We assume that the
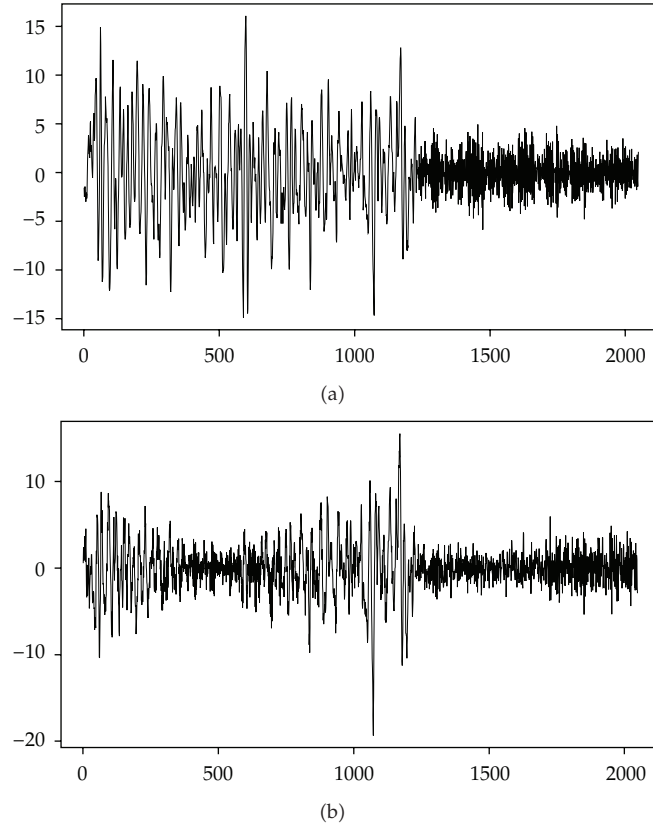
Figure 1: Example 5.1: (a) input series and (b) output series.

orders $m$ and $n$ are fixed and known. The idea is to expand these functions in wavelet series

$$\delta_i(u) = \alpha_{00}^{(\delta_i)}\phi(u) + \sum_{j=0}^{J-1} \sum_{k \in I_j} \beta_{jk}^{(\delta_i)}\psi_{jk}(u), \quad i = 1,\ldots,m,$$

$$\omega_i(u) = \alpha_{00}^{(\omega_i)}\phi(u) + \sum_{j=0}^{J-1} \sum_{k \in I_j} \beta_{jk}^{(\omega_i)}\psi_{jk}(u), \quad i = 0,\ldots,n.$$

(3.1)

The empirical wavelet coefficients are obtained minimizing

$$\sum_{t=v+1}^{T} \left( Y_{t,T} - \sum_{i=1}^{m} \delta_i(u)Y_{t-i,T} - \sum_{j=0}^{n} \omega_j(u)X_{t-j,T} \right)^2,$$

(3.2)

with $\delta_i(u)$ and $\omega_i(u)$ replaced by (3.1), $v = \max(m,n)$. These empirical wavelet coefficients are then modified using a hard threshold, and finally we build estimators of $\delta_i(u)$ and $\omega_i(u)$ by applying the inverse wavelet transform to these thresholded coefficients.
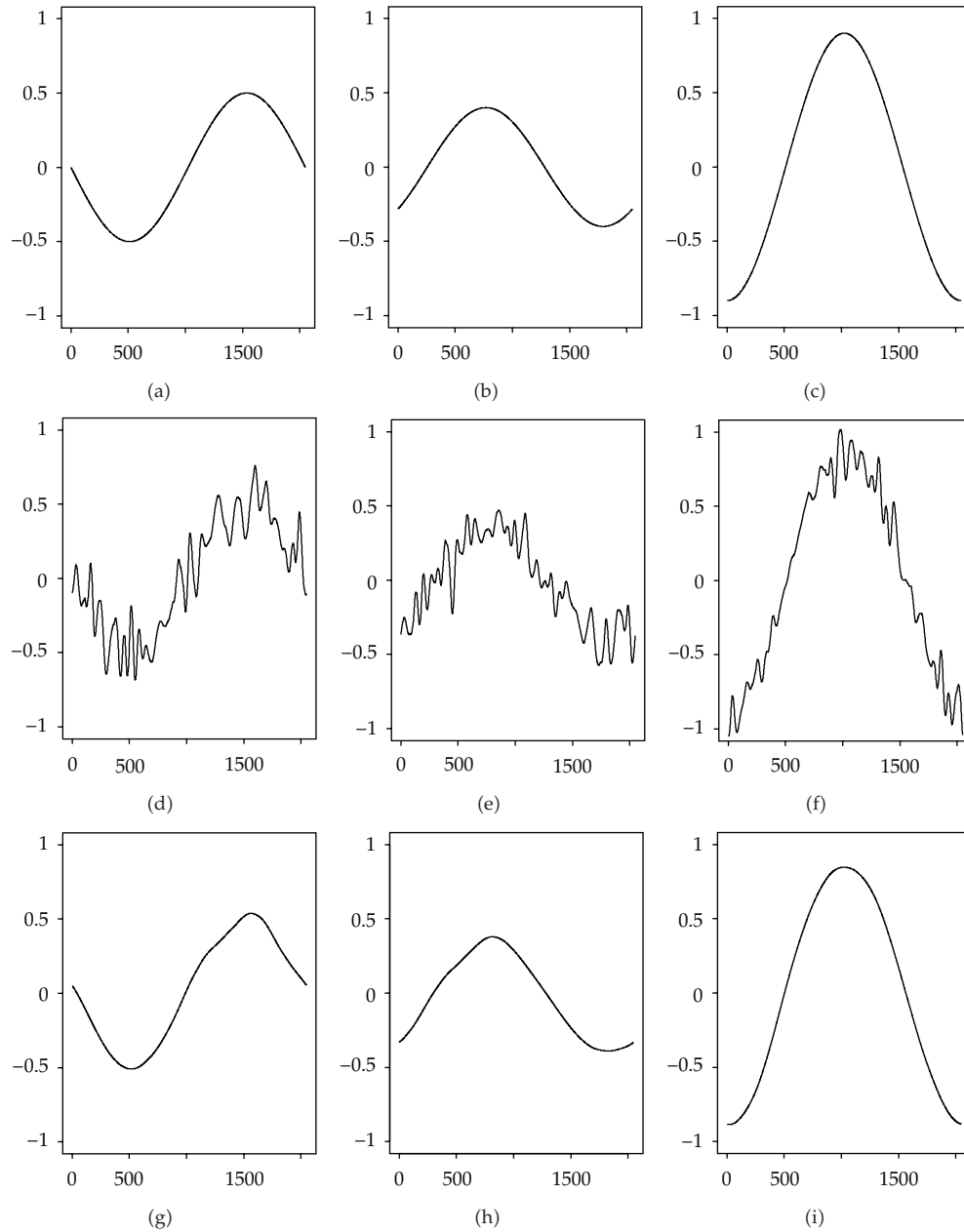
**Figure 2:** Example 5.1: True parameters and estimates using the Daubechies D12 wavelet: (a) $\delta_1$, (b) $\delta_2$, (c) $\omega_0$, (d) $\hat{\delta}_1$, (e) $\hat{\delta}_2$ (f) $\hat{\omega}_0$, (g) $\tilde{\delta}_1$, (h) $\tilde{\delta}_2$ (i) $\tilde{\omega}_0$.

For ease of exposition, we restrict our attention from now on to the simple model with $m = 1$ and $n = 0$, namely,

$$Y_{t,T} = \delta_1\left(\frac{t}{T}\right)Y_{t-1,T} + \omega_0\left(\frac{t}{T}\right)X_{t,T} + \varepsilon_t. \tag{3.3}$$
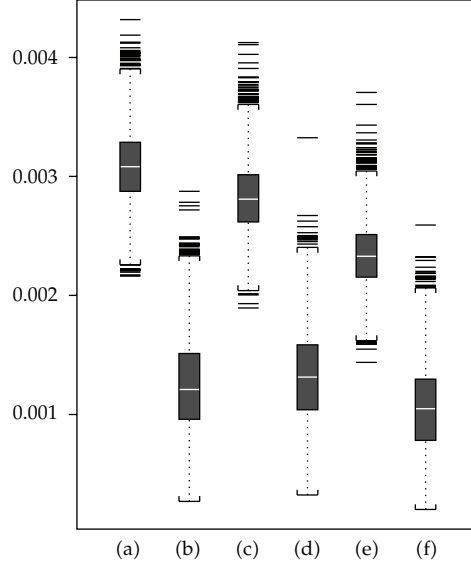
**Figure 3:** Example 5.1: Boxplots of the 5000 RMSE-values using estimators with the Daubechies D12 wavelet, for linear estimates (a) $\hat{\delta}_1(\cdot)$, (c) $\hat{\delta}_2$, (e) $\hat{w}_0(\cdot)$ and nonlinear estimators (b) $\tilde{\delta}_1(\cdot)$, (d) $\tilde{\delta}_2$, (f) $\tilde{w}_0(\cdot)$.

So we regress $Y_{t,T}$ on $X_{t,T}$ and $Y_{t-1,T}$, using the expansion (3.1) for $\delta_1(t/T)$ and $\omega_0(t/T)$. Let $\Delta = 2^J - 1$. In matrix notation, we have

$$
\begin{bmatrix} Y_{2,T} \\ Y_{3,T} \\ \vdots \\ Y_{T,T} \end{bmatrix} =
\begin{bmatrix}
\phi_{00}\left(\dfrac{2}{T}\right)Y_{1,T} & \psi_{00}\left(\dfrac{2}{T}\right)Y_{1,T} & \cdots & \psi_{J-1,\Delta}\left(\dfrac{2}{T}\right)Y_{1,T} \\
\phi_{00}\left(\dfrac{3}{T}\right)Y_{2,T} & \psi_{00}\left(\dfrac{3}{T}\right)Y_{2,T} & \cdots & \psi_{J-1,\Delta}\left(\dfrac{3}{T}\right)Y_{2,T} \\
\vdots & \vdots & \ddots & \vdots \\
\phi_{00}\left(\dfrac{T}{T}\right)Y_{T-1,T} & \psi_{00}\left(\dfrac{T}{T}\right)Y_{T-1,T} & \cdots & \psi_{J-1,\Delta}\left(\dfrac{T}{T}\right)Y_{T-1,T}
\end{bmatrix}
\begin{bmatrix} \alpha_{00}^{(\delta_1)} \\ \beta_{00}^{\delta_1} \\ \vdots \\ \beta_{J-1,\Delta}^{(\delta_1)} \end{bmatrix}
$$

$$
+ \begin{bmatrix}
\phi_{00}\left(\dfrac{2}{T}\right)X_{2,T} & \psi_{00}\left(\dfrac{2}{T}\right)X_{2,T} & \cdots & \psi_{J-1,\Delta}\left(\dfrac{2}{T}\right)X_{2,T} \\
\phi_{00}\left(\dfrac{3}{T}\right)X_{3,T} & \psi_{00}\left(\dfrac{3}{T}\right)X_{3,T} & \cdots & \psi_{J-1,\Delta}\left(\dfrac{3}{T}\right)X_{3,T} \\
\vdots & \vdots & \ddots & \vdots \\
\phi_{00}\left(\dfrac{T}{T}\right)X_{T,T} & \psi_{00}\left(\dfrac{T}{T}\right)X_{T,T} & \cdots & \psi_{J-1,\Delta}\left(\dfrac{T}{T}\right)X_{T,T}
\end{bmatrix}
\begin{bmatrix} \alpha_{00}^{(\omega_0)} \\ \beta_{00}^{(\omega_0)} \\ \vdots \\ \beta_{J-1,\Delta}^{(\omega_0)} \end{bmatrix}
+ \begin{bmatrix} \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_T \end{bmatrix}
$$

$$(3.4)$$

or

$$
\begin{aligned}
\mathbf{Y} &= \begin{bmatrix} \boldsymbol{\Phi}_Y & \boldsymbol{\Psi}_Y^{(0)} & \boldsymbol{\Psi}_Y^{(1)} & \cdots & \boldsymbol{\Psi}_Y^{(J-1)} \end{bmatrix} \boldsymbol{\beta}^{(\delta_1)} \\
&+ \begin{bmatrix} \boldsymbol{\Phi}_X & \boldsymbol{\Psi}_X^{(0)} & \boldsymbol{\Psi}_X^{(1)} & \cdots & \boldsymbol{\Psi}_X^{(J-1)} \end{bmatrix} \boldsymbol{\beta}^{(\omega_0)} + \varepsilon,
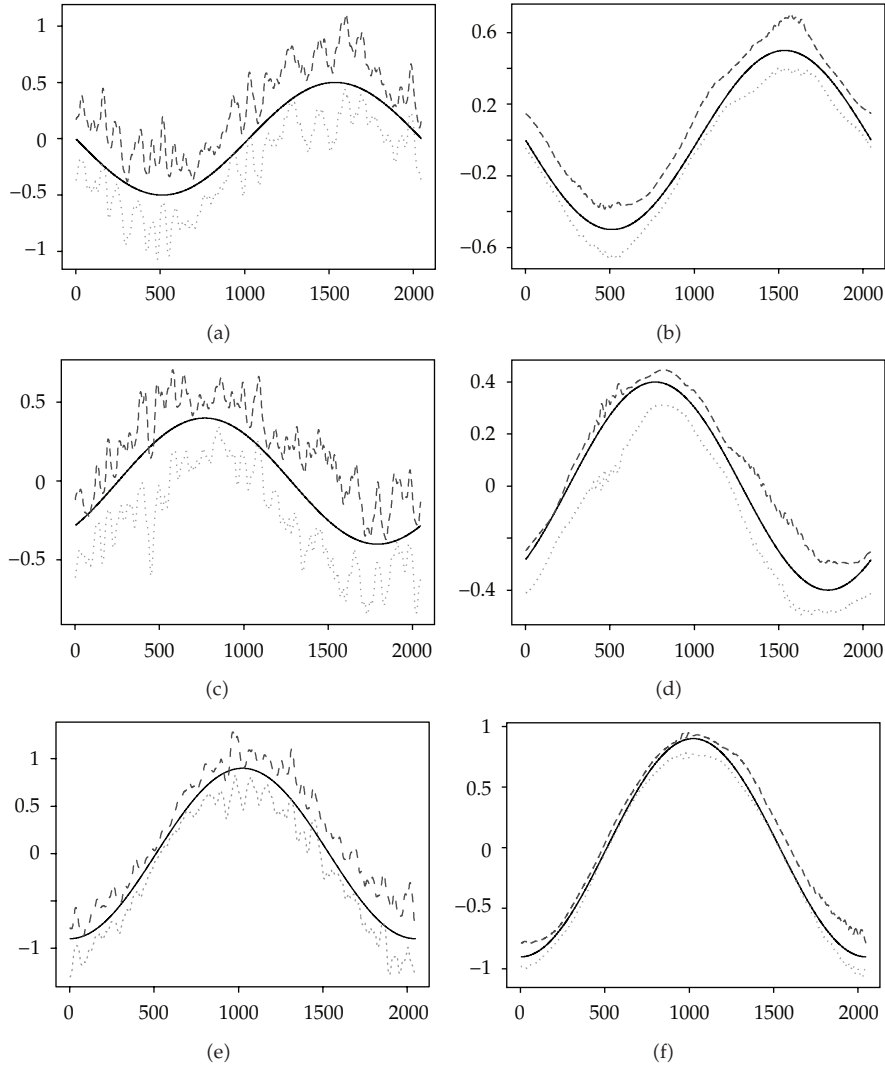\end{aligned}
$$

$$(3.5)$$

**Figure 4:** Example 5.1: Bootstrap 95% confidence regions for linear estimators, (a), (c), and (e) and nonlinear estimators, (b), (d), and (f), with true curves inside.

where

$$\mathbf{Y} = \begin{pmatrix} Y_{2,T} & Y_{3,T} & \cdots & Y_{T,T} \end{pmatrix}',$$

$$\boldsymbol{\beta}^{(\delta_1)} = \begin{pmatrix} \alpha_{00}^{(\delta_1)} & \beta_{00}^{(\delta_1)} & \beta_{10}^{(\delta_1)} & \beta_{11}^{(\delta_1)} & \cdots & \beta_{J-1,0}^{(\delta_1)} & \cdots & \beta_{J-1,\Delta}^{(\delta_1)} \end{pmatrix}',$$

$$\boldsymbol{\beta}^{(\omega_0)} = \begin{pmatrix} \alpha_{00}^{(\omega_0)} & \beta_{00}^{(\omega_0)} & \beta_{10}^{(\omega_0)} & \beta_{11}^{(\omega_0)} & \cdots & \beta_{J-1,0}^{(\omega_0)} & \cdots & \beta_{J-1,\Delta}^{(\omega_0)} \end{pmatrix}',$$

$$\varepsilon = \begin{pmatrix} \varepsilon_2 & \varepsilon_3 & \cdots & \varepsilon_T \end{pmatrix}'.$$

(3.6)
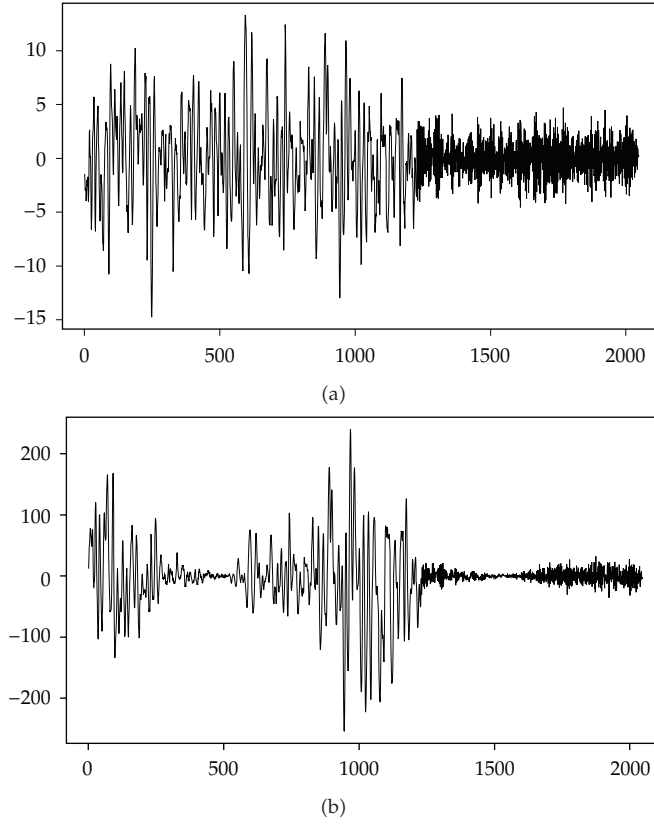
Figure 5: Example 5.2: (a) input series and (b) output series.

The vector $\mathbf{\Phi}_Y$ is given by

$$\mathbf{\Phi}_Y = \begin{bmatrix} \phi_{00}\left(\dfrac{2}{T}\right)Y_{1,T} \\[2ex] \phi_{00}\left(\dfrac{3}{T}\right)Y_{2,T} \\[1ex] \vdots \\[1ex] \phi_{00}\left(\dfrac{T}{T}\right)Y_{T-1,T} \end{bmatrix}, \tag{3.7}$$

and the matrices $\mathbf{\Psi}_Y^{(m)}$ are given, for $0 \le m \le J - 1$, by

$$\mathbf{\Psi}_Y^{(m)} = \begin{bmatrix} \psi_{m0}\left(\dfrac{2}{T}\right)Y_{1,T} & \psi_{m1}\left(\dfrac{2}{T}\right)Y_{1,T} & \cdots & \psi_{m,2^m-1}\left(\dfrac{2}{T}\right)Y_{1,T} \\[2ex] \psi_{m0}\left(\dfrac{3}{T}\right)Y_{2,T} & \psi_{m1}\left(\dfrac{3}{T}\right)Y_{2,T} & \cdots & \psi_{m,2^m-1}\left(\dfrac{3}{T}\right)Y_{2,T} \\[1ex] \vdots & \vdots & \ddots & \vdots \\[1ex] \psi_{m0}\left(\dfrac{T}{T}\right)Y_{T-1,T} & \psi_{m1}\left(\dfrac{T}{T}\right)Y_{T-1,T} & \cdots & \psi_{m,2^m-1}\left(\dfrac{T}{T}\right)Y_{T-1,T} \end{bmatrix}. \tag{3.8}$$
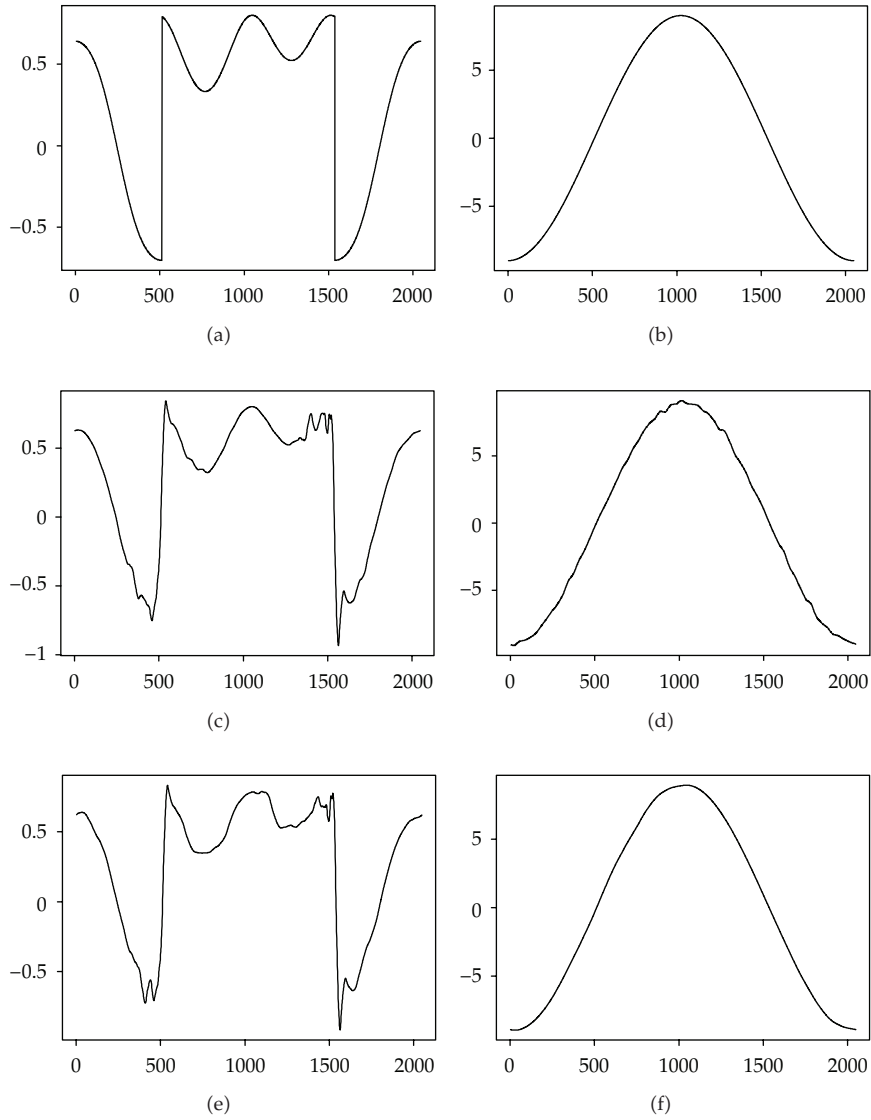
**Figure 6:** Example 5.2: True parameters and estimates using the Daubechies D8 wavelet: (a) $\delta_1$, (b) $\omega_0$, (c) $\widehat{\delta}_1$, (d) $\widehat{\omega}_0$, (e) $\widetilde{\delta}_1$, (f) $\widetilde{\omega}_0$.

The vector $\mathbf{\Phi}_X$ is given by

$$
\mathbf{\Phi}_X = \begin{bmatrix} \phi_{00}\left(\dfrac{2}{T}\right)X_{2,T} \\ \phi_{00}\left(\dfrac{3}{T}\right)X_{3,T} \\ \vdots \\ \phi_{00}\left(\dfrac{T}{T}\right)X_{T,T} \end{bmatrix},
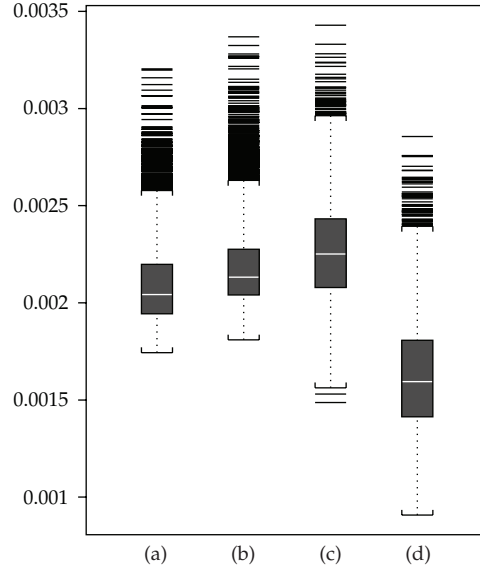\tag{3.9}
$$

**Figure 7:** Example 5.2: Boxplots of the 5000 RMSE-values using estimators with the Daubechies D8 wavelet, for linear estimates (a) $\widehat{\delta}_1(\cdot)$, (c) $\widehat{w}_0(\cdot)$, and nonlinear estimators (b) $\widetilde{\delta}_1(\cdot)$, (d) $\widetilde{w}_0(\cdot)$.

and the matrices $\boldsymbol{\Psi}_X^{(m)}$ are given, for $0 \leq m \leq J-1$, by

$$
\boldsymbol{\Psi}_X^{(m)} = \begin{bmatrix} \psi_{m0}\left(\dfrac{2}{T}\right)X_{2,T} & \psi_{m1}\left(\dfrac{2}{T}\right)X_{2,T} & \cdots & \psi_{m,2^m-1}\left(\dfrac{2}{T}\right)X_{2,T} \\ \psi_{m0}\left(\dfrac{3}{T}\right)X_{3,T} & \psi_{m1}\left(\dfrac{3}{T}\right)X_{3,T} & \cdots & \psi_{m,2^m-1}\left(\dfrac{3}{T}\right)X_{3,T} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{m0}\left(\dfrac{T}{T}\right)X_{T,T} & \psi_{m1}\left(\dfrac{T}{T}\right)X_{T,T} & \cdots & \psi_{m,2^m-1}\left(\dfrac{T}{T}\right)X_{T,T} \end{bmatrix}.
\tag{3.10}
$$

Let $\boldsymbol{\Psi}_X = [\boldsymbol{\Phi}_X\ \boldsymbol{\Psi}_X^{(0)}\ \cdots\ \boldsymbol{\Psi}_X^{(J-1)}]$ and $\boldsymbol{\Psi}_Y = [\boldsymbol{\Phi}_Y\ \boldsymbol{\Psi}_Y^{(0)}\ \cdots\ \boldsymbol{\Psi}_Y^{(J-1)}]$.

It follows easily that the least squares estimators of the coefficients are then given by

$$
\begin{bmatrix} \widehat{\boldsymbol{\beta}}^{(\delta_1)} \\ \widehat{\boldsymbol{\beta}}^{(\omega_0)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Psi}_Y'\boldsymbol{\Psi}_Y & \boldsymbol{\Psi}_Y'\boldsymbol{\Psi}_X \\ \boldsymbol{\Psi}_X'\boldsymbol{\Psi}_Y & \boldsymbol{\Psi}_Y'\boldsymbol{\Psi}_X \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{\Psi}_Y'\mathbf{Y} \\ \boldsymbol{\Psi}_X'\mathbf{Y} \end{bmatrix}.
\tag{3.11}
$$

Having obtained the estimates given by (3.11), we plug them in (3.1), resulting in linear estimates $\widehat{\delta}_i(u)$ and $\widehat{w}_i(u)$. Finally, nonlinear smoothed estimators $\widetilde{\delta}_i(u)$ and $\widetilde{w}_i(u)$, respectively, are obtained applying some threshold to the empirical detail coefficients $\widehat{\beta}_{jk}^{(i)}$ and taking the inverse wavelet transform.
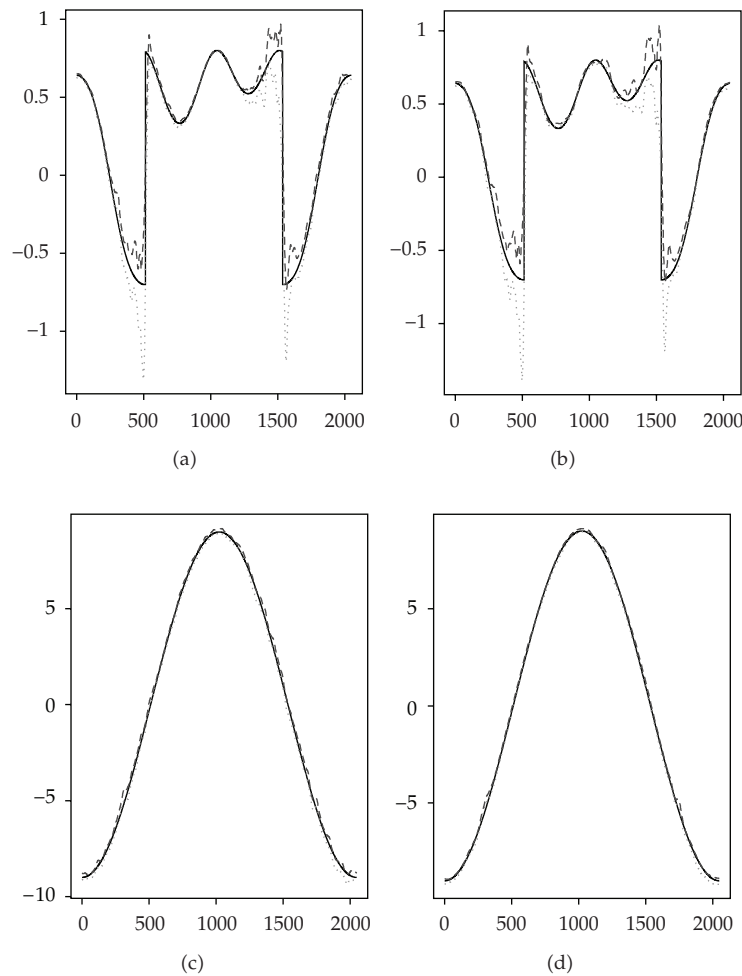
**Figure 8:** Example 5.2: Bootstrap 95% confidence regions for linear estimators, (a) and (c), and nonlinear estimators, (b) and (d), with true curves inside.

We might think of using other procedures, as the LASSO (least absolute shrinkage and selection operator) procedure of Tibshirani [31] or the nonnegative garotte of Breiman [32]. For the case of orthogonal design, LASSO has a relationship with the soft shrinkage procedure of Donoho and coworkers. But using these procedures would make the derivations of the properties of the estimators above considerably more difficult, or even impossible.

## 4. Properties of Empirical Coefficients

Now we present some properties of the linear empirical wavelet coefficients. The techniques used to prove the results are quite involved and are based on function space theory. Basically we adapt the results of Dahlhaus et al. [7] for the transfer function model (1.1).
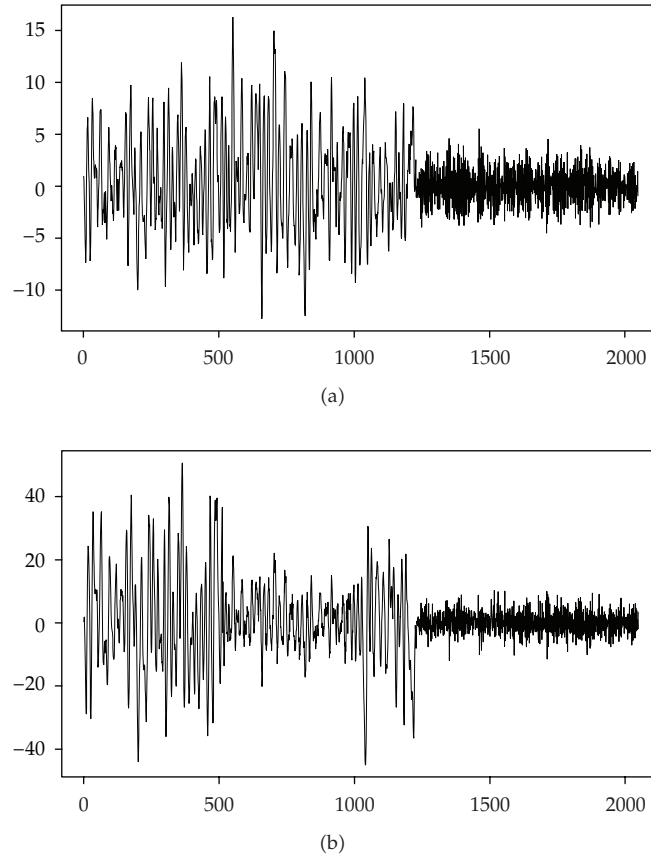
(a)



(b)

**Figure 9:** Example 5.3: (a) input series and (b) output series.

The proofs may be obtained from the authors upon request or from the site of the journal.

We assume that functions $\omega_i(u)$, $i = 0, 1, \ldots, n$ and $\delta_i(u)$, $i = 1, \ldots, m$ belong to some function spaces $\mathcal{F}_i$ given by

$$\mathcal{F}_i = \left\{ f_i = \alpha_{00}^{(i)} \phi + \sum_{j,k} \beta_{jk}^{(i)} \psi_{jk} : \left\| \alpha_{00}^{(i)} \right\| \leq C_{i1}, \left\| \beta_{\cdot\cdot}^{(i)} \right\|_{\ell, p, q} \leq C_{i2} \right\}, \quad (4.1)$$

where

$$\left\| \beta_{\cdot\cdot}^{(i)} \right\|_{\ell_i, p_i, q_i} = \left( \sum_{j \geq 0} \left[ 2^{j s_i p_i} \sum_{k \in I_j} \left| \beta_{jk}^{(i)} \right|^{p_i} \right]^{q_i / p_i} \right)^{1/q_i}, \quad (4.2)$$
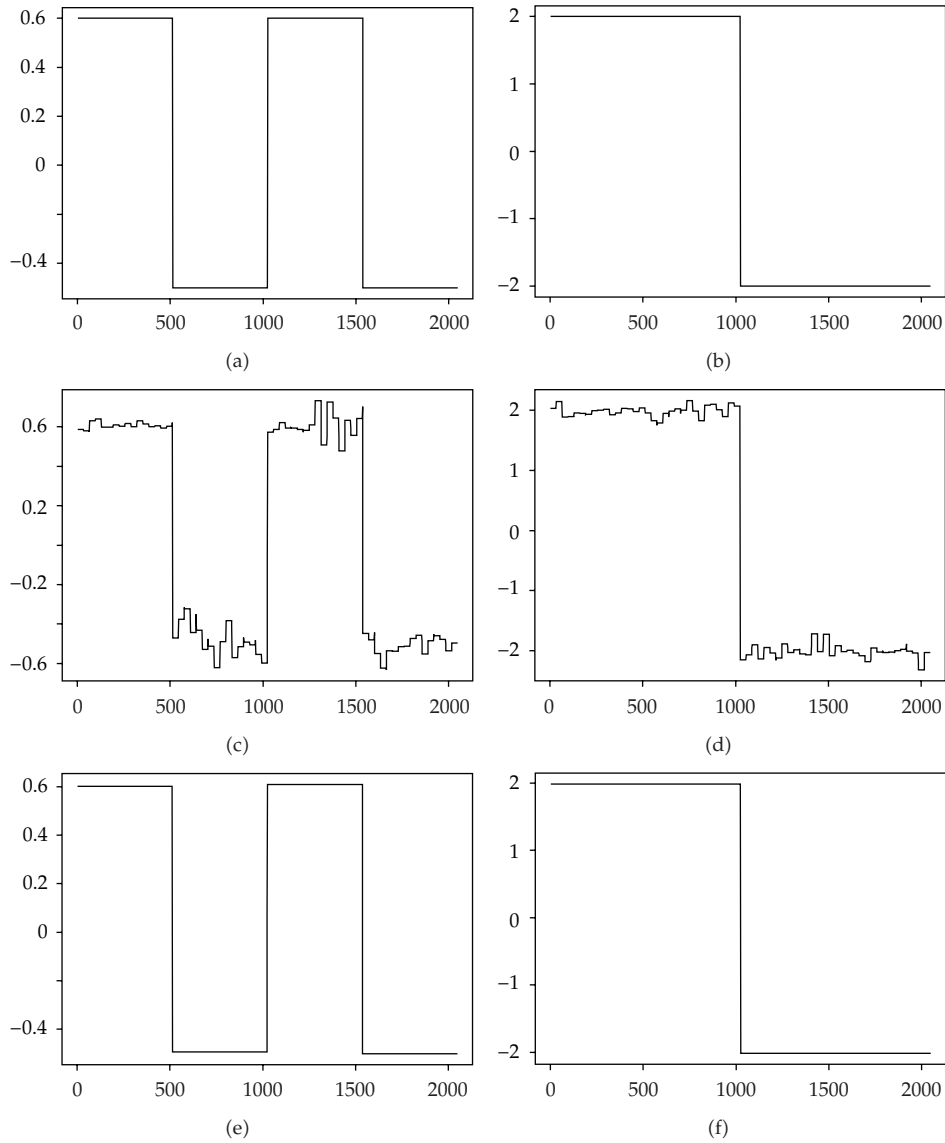
**Figure 10:** Example 5.3: True parameters and estimates using the Haar wavelet: (a) $\delta_1$, (b) $\omega_0$, (c) $\widehat{\delta}_1$, (d) $\widehat{\omega}_0$, (e) $\widetilde{\delta}_1$, (f) $\widetilde{\omega}_0$.

$s_i = \ell_i + 1/2 - 1/p_i$. Here, $\ell_i$ is the smoothness degree of each space, $p_i$ and $q_i$ ($1 \leq p_i$, $q_i \leq \infty$) specify the norm, and $C_{i1}$ and $C_{i2}$ are positive constants. For these function spaces, the following result is valid [19]:

$$\sup_{f_i \in \mathcal{F}_i} \left\{ \sum_{j \geq J} \sum_{k} \left| \beta_{jk}^{(i)} \right|^2 \right\} = O\left( 2^{-2J\widetilde{s}_i} \right), \tag{4.3}$$

where $\widetilde{s}_i = \ell_i + 1/2 - 1/\widetilde{p}_i$, with $\widetilde{p}_i = \min\{p_i, 2\}$.
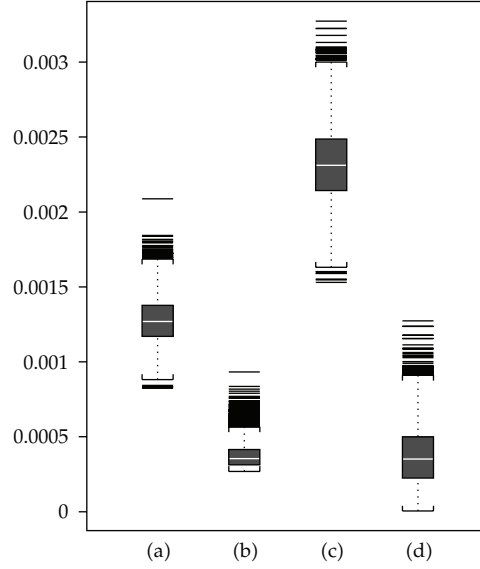
**Figure 11:** Example 5.3: Boxplots of the 5000 RMSE-values using estimators with the Haar wavelet, for linear estimates (a) $\hat{\delta}_1(\cdot)$, (c) $\hat{w}_0(\cdot)$ and nonlinear estimators (b) $\tilde{\delta}_1(\cdot)$, (d) $\tilde{w}_0(\cdot)$.

These classes contain Besov, Hölder, and $L_2$-Sobolev spaces, see, for example, Vidakovic [10], Dahlhaus et al. [7], and Triebel [33].

It can be shown, see Donoho et al. [19], that the loss incurred by truncating at level $J$ is of order $T^{-2\ell_i/(2\ell_i+1)}$ if we choose $J$ such that $2^{J-1} \le T^{1/2} \le 2^J$. This is achieved if $\tilde{s}_i > 1$.

Concerning the wavelets, we assume that $\phi$ and $\psi$ are compactly supported on $[0,1]$ and have continuous derivatives up to order $r > l$, with $\ell = \max \ell_i$. We denote the spectral norm by $\| \cdot \|_2$ and the sup norm by $\| \cdot \|_\infty$.

In order to analyze the statistical behavior of the estimated coefficients, it is convenient to assume that we take expansions of these coefficients as linear combinations of functions in the $\tilde{V}_J$ spaces, generated by $\{\phi_{J,1}, \phi_{J,2}, \ldots, \phi_{J,2^J}\}$. With this basis, we can write (3.3) as

$$Y_{t,T} = \sum_{i=1}^{2^J} \zeta_{J,i}^{(\delta_1)} \phi_{J,i}\left(\frac{t}{T}\right) Y_{t-1,T} + \sum_{i=1}^{2^J} \zeta_{J,i}^{(\omega_0)} \phi_{J,i}\left(\frac{t}{T}\right) X_{t,T} + \gamma_{t,T}, \tag{4.4}$$

where

$$\gamma_{t,T} = \sum_{j \ge J} \sum_{k \in I_j} \beta_{jk}^{(\delta_1)} \psi_{jk}\left(\frac{t}{T}\right) Y_{t-1,T} + \sum_{j \ge J} \sum_{k \in I_j} \beta_{jk}^{(\omega_0)} \psi_{jk}\left(\frac{t}{T}\right) X_{t,T} + \varepsilon_t. \tag{4.5}$$

Equation (4.4) in matrix form becomes

$$\mathbf{Y} = \left[ \varphi_Y \ \vdots \ \varphi_X \right] \begin{bmatrix} \boldsymbol{\zeta}^{(\delta_1)} \\ \cdots \\ \boldsymbol{\zeta}^{(\omega_0)} \end{bmatrix} + \gamma, \tag{4.6}$$
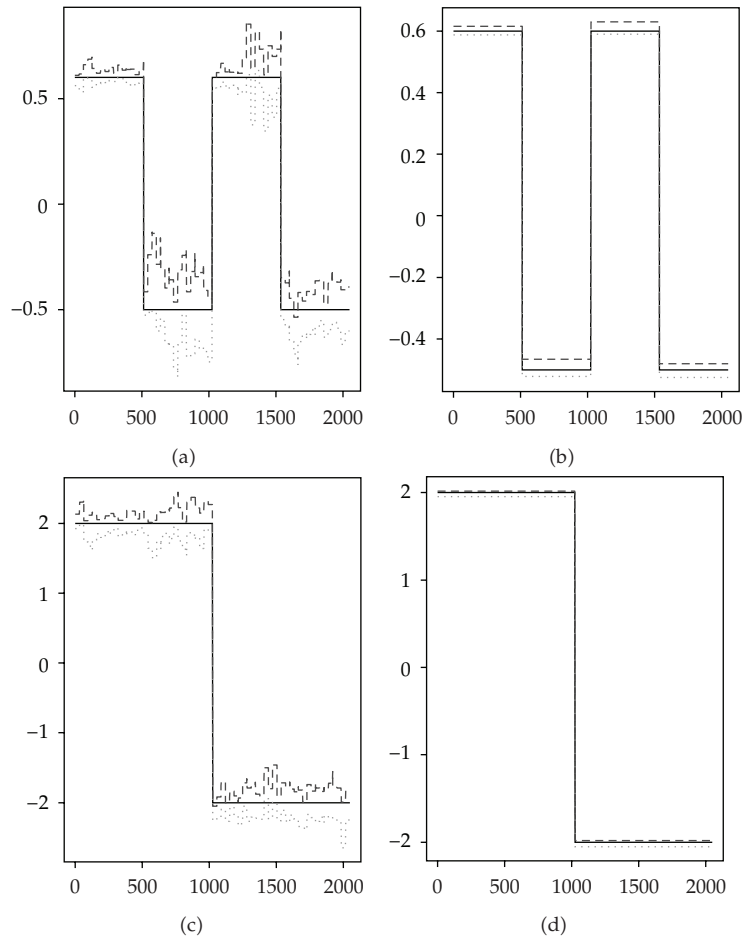
**Figure 12:** Example 5.3: Bootstrap 95% confidence regions for linear estimators, (a) and (c), and nonlinear estimators, (b) and (d), with true curves inside.

with

$$
\varphi_Y = \begin{bmatrix}
\phi_{J,1}\left(\dfrac{2}{T}\right)Y_{1,T} & \phi_{J,2}\left(\dfrac{2}{T}\right)Y_{1,T} & \cdots & \phi_{J,2^J}\left(\dfrac{2}{T}\right)Y_{1,T} \\[2ex]
\phi_{J,1}\left(\dfrac{3}{T}\right)Y_{2,T} & \phi_{J,2}\left(\dfrac{3}{T}\right)Y_{2,T} & \cdots & \phi_{J,2^J}\left(\dfrac{3}{T}\right)Y_{2,T} \\[1ex]
\vdots & \vdots & \ddots & \vdots \\[1ex]
\phi_{J,1}\left(\dfrac{T}{T}\right)Y_{T-1,T} & \phi_{J,2}\left(\dfrac{T}{T}\right)Y_{T-1,T} & \cdots & \phi_{J,2^J}\left(\dfrac{T}{T}\right)Y_{T-1,T}
\end{bmatrix},
$$

$$
\varphi_X = \begin{bmatrix}
\phi_{J,1}\left(\dfrac{2}{T}\right)X_{2,T} & \phi_{J,2}\left(\dfrac{2}{T}\right)X_{2,T} & \cdots & \phi_{J,2^J}\left(\dfrac{2}{T}\right)X_{2,T} \\[2ex]
\phi_{J,1}\left(\dfrac{3}{T}\right)X_{3,T} & \phi_{J,2}\left(\dfrac{3}{T}\right)X_{3,T} & \cdots & \phi_{J,2^J}\left(\dfrac{3}{T}\right)X_{3,T} \\[1ex]
\vdots & \vdots & \ddots & \vdots \\[1ex]
\phi_{J,1}\left(\dfrac{T}{T}\right)X_{T,T} & \phi_{J,2}\left(\dfrac{T}{T}\right)X_{T,T} & \cdots & \phi_{J,2^J}\left(\dfrac{T}{T}\right)X_{T,T}
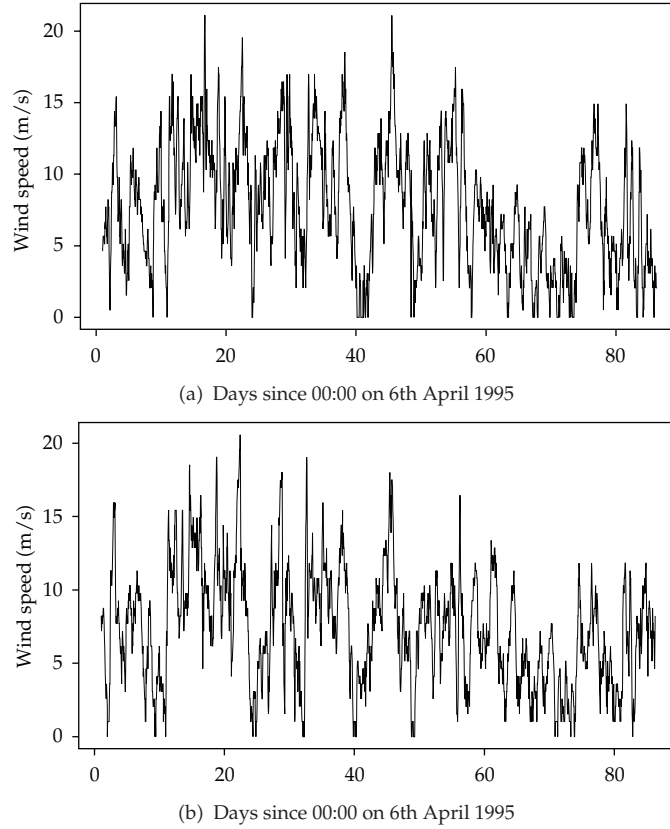\end{bmatrix}.
$$

$$(4.7)$$

(a) Days since 00:00 on 6th April 1995



(b) Days since 00:00 on 6th April 1995

**Figure 13:** Hourly wind speeds from 00:00 on 6th April 1995: (a) Aberporth (b) Valley.

In (4.6), $\boldsymbol{\zeta}^{(\delta_1)}$ and $\boldsymbol{\zeta}^{(\omega_0)}$ contain the coefficients in the expansion (4.4) and $\gamma$ contains lines given by (4.5).

Analogously to (3.11), we have the least squares estimates

$$
\begin{bmatrix} \widehat{\boldsymbol{\zeta}}^{(\delta_1)} \\ \widehat{\boldsymbol{\zeta}}^{(\omega_0)} \end{bmatrix} = \begin{bmatrix} \varphi_Y' \varphi_Y & \varphi_Y' \varphi_X \\ \varphi_X' \varphi_Y & \varphi_X' \varphi_X \end{bmatrix}^{-1} \begin{bmatrix} \varphi_Y' \mathbf{Y} \\ \varphi_X' \mathbf{Y} \end{bmatrix}.
\tag{4.8}
$$

The relationship between $\begin{pmatrix} \boldsymbol{\beta}^{(\delta_1)} \\ \boldsymbol{\beta}^{(\omega_0)} \end{pmatrix}$ and $\begin{pmatrix} \boldsymbol{\zeta}^{(\delta_1)} \\ \boldsymbol{\zeta}^{(\omega_0)} \end{pmatrix}$ is

$$
\begin{pmatrix} \boldsymbol{\beta}^{(\delta_1)} \\ \boldsymbol{\beta}^{(\omega_0)} \end{pmatrix} = \boldsymbol{\Gamma} \begin{pmatrix} \boldsymbol{\zeta}^{(\delta_1)} \\ \boldsymbol{\zeta}^{(\omega_0)} \end{pmatrix},
\tag{4.9}
$$

where the $\boldsymbol{\Gamma}$ is a $(2^{J+1} \times 2^{J+1})$ block diagonal matrix. The matrix $\boldsymbol{\Gamma}$ does the transformation $(\widehat{\alpha}_{00}^{(i)},$ $\widehat{\beta}_{00}^{(i)}, \widehat{\beta}_{10}^{(i)}, \widehat{\beta}_{11}^{(i)}, \ldots, \widehat{\beta}_{J-1,0}^{(i)}, \ldots, \widehat{\beta}_{J-1,\Delta}^{(i)})' = \boldsymbol{\Gamma}(\widehat{\zeta}_{J,1}^{(i)}, \ldots, \widehat{\zeta}_{J,2^J}^{(i)})'$; so each coefficient estimate $\widehat{\beta}_{jk}^{(i)}, i = \delta_1, \omega_0$ in (3.11) can be written as $\widehat{\beta}_{jk}^{(i)} = \Gamma_{i,jk}' \widehat{\zeta}$ and $\|\boldsymbol{\Gamma}_{\delta_1,j,k}\|_{L_2} = \|\boldsymbol{\Gamma}_{\omega_0,j,k}\|_{L_2} = 1$.
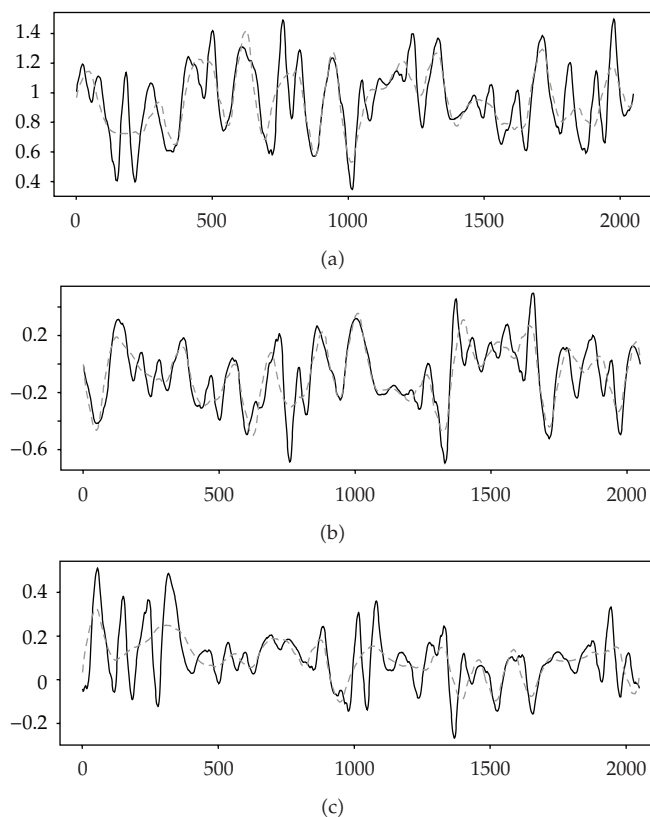
(a)

(b)

(c)

**Figure 14:** Model $M_3$: (a) $\widehat{\delta}_1(u)$ (full line) and $\widetilde{\delta}_1(u)$ (dashed line), (b) $\widehat{\delta}_2(u)$ (full line) and $\widetilde{\delta}_2(u)$ (dashed line), (c) $\widehat{\omega}_0(u)$ (full line) and $\widetilde{\omega}_0(u)$ (dashed line).

Equation (4.6) can be written in the form

$$\Upsilon = \Upsilon\zeta + \gamma, \tag{4.10}$$

with $\Upsilon = [\varphi_Y \vdots \varphi_X]$, $\zeta = \begin{pmatrix} \zeta^{(\delta_1)} \\ \zeta^{(\omega_0)} \end{pmatrix}$ and the vector $\gamma$ as defined above. The least squares estimator of $\zeta$ is given by

$$\widehat{\zeta} = (\Upsilon'\Upsilon)^{-1}\Upsilon'\Upsilon. \tag{4.11}$$

We notice that the error term $\gamma$ is not independent from the regressors $\Upsilon$, which means that the estimator is biased and its squared error is of order $O_p(2^J T^{-1/2} + \text{bias})$. But this order can be improved, as shown below. To take care of the bias, we might use some robust procedure, as in Martin et al. [34], for example.

From (4.10) and (4.11), we obtain

$$\widehat{\zeta} = (\Upsilon'\Upsilon)^{-1}\Upsilon'(\Upsilon\zeta + \gamma). \tag{4.12}$$
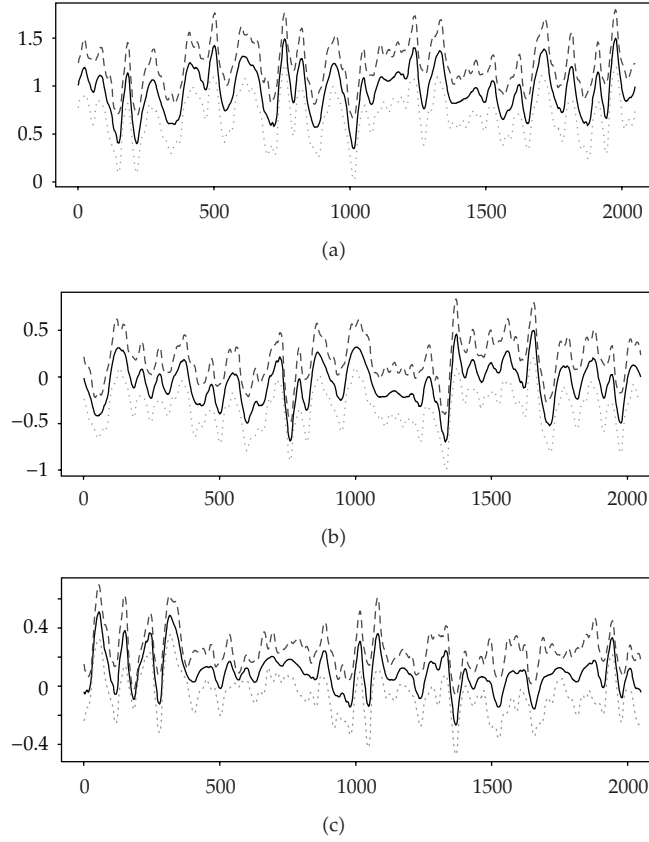
(a)



(b)



(c)

**Figure 15:** Model $M_3$: Bootstrap 95% confidence regions for (a) $\delta_1(u)$, (b) $\delta_2(u)$, (c) $\omega_0(u)$, using linear estimates.

If we write $\gamma = \varepsilon + \mathbf{S}$, we get

$$\widehat{\boldsymbol{\zeta}} = \left(\boldsymbol{\Upsilon}'\boldsymbol{\Upsilon}\right)^{-1}\boldsymbol{\Upsilon}'(\boldsymbol{\Upsilon}\boldsymbol{\zeta} + \varepsilon + \mathbf{S}), \tag{4.13}$$

where the matrix $\mathbf{S}$ is an $((T-1) \times 1)$ matrix, given by

$$\mathbf{S} = \begin{bmatrix} \displaystyle\sum_{j \geq J}\sum_{k \in I_j}\beta_{jk}^{(\delta_1)}\psi_{jk}\left(\frac{2}{T}\right)Y_{1,T} + \sum_{j \geq J}\sum_{k \in I_j}\beta_{jk}^{(\omega_0)}\psi_{jk}\left(\frac{2}{T}\right)X_{2,T} \\ \displaystyle\sum_{j \geq J}\sum_{k \in I_j}\beta_{jk}^{(\delta_1)}\psi_{jk}\left(\frac{3}{T}\right)Y_{2,T} + \sum_{j \geq J}\sum_{k \in I_j}\beta_{jk}^{(\omega_0)}\psi_{jk}\left(\frac{3}{T}\right)X_{3,T} \\ \vdots \\ \displaystyle\sum_{j \geq J}\sum_{k \in I_j}\beta_{jk}^{(\delta_1)}\psi_{jk}\left(\frac{T}{T}\right)Y_{T-1,T} + \sum_{j \geq J}\sum_{k \in I_j}\beta_{jk}^{(\omega_0)}\psi_{jk}\left(\frac{T}{T}\right)X_{T,T} \end{bmatrix}. \tag{4.14}$$
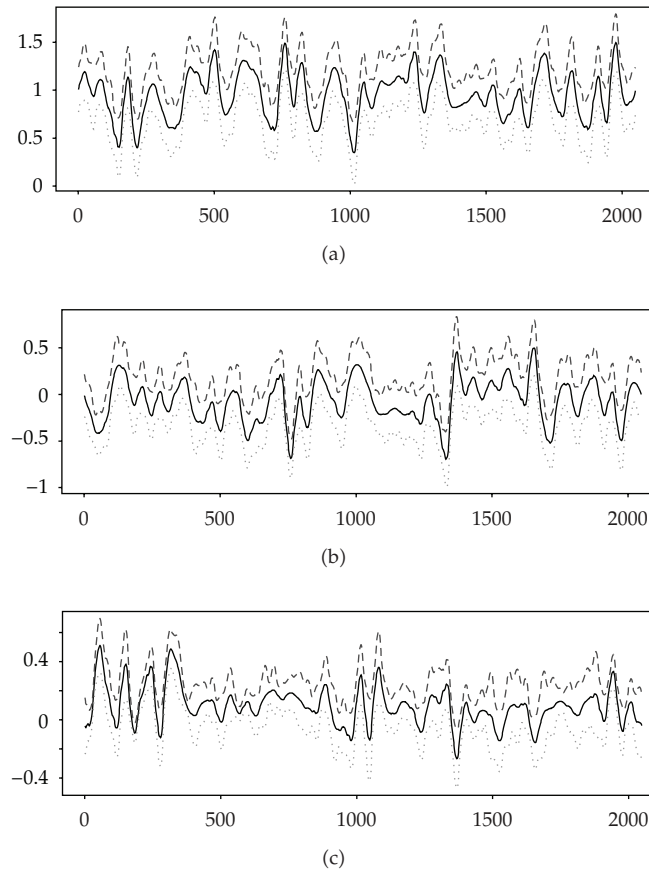
(a)



(b)



(c)

**Figure 16:** Model $M_3$: Bootstrap 95% confidence regions for (a) $\delta_1(u)$, (b) $\delta_2(u)$, (c) $\omega_0(u)$, using nonlinear estimates.

We now state results on the squared error and mean squared error of the linear empirical wavelet coefficients. The following assumptions are needed.

(A1) The wavelets $\phi$ and $\psi$ are compactly supported on $[0,1]$ and have continuous derivatives up to order $r > l$, with $\ell = \max \ell_i$.

(A2) In the estimation procedure, we have used first a linear estimator, truncating the wavelet expansion at scale $2^J$. In order to get functions with the appropriate smoothness, $J$ was chosen such that $\tilde{s}_i > 1$.

(A3) The matrix $\boldsymbol{\Upsilon}$ in (4.10) satisfies $E\|(\boldsymbol{\Upsilon}'\boldsymbol{\Upsilon})^{-1}\|_{L_2}^{2+\delta} = O(T^{-2-\delta})$, for some $\delta > 0$.

The Daubechies Extremal Phase wavelets $D2N$, for an appropriate choice of $N$, satisfy (A1). See Härdle et al. [35] for details. The number of continuous derivatives has a connection with the number of null moments of $\psi$. Also, the support of $\psi$ is contained in the interval $[-N+1, N]$.

Since $\beta_{jk}$ is the coefficient of the regressor $\psi_{jk}$, the squared error of the least squares estimator is $\|\hat{\beta}_{jk} - \beta_{jk}\|^2 = O_p(T^{-1})$, which implies that the square error $\|\hat{\delta}(u) - \delta(u)\|^2 = O_p(2^J T^{-1/2} + \text{bias})$, for example. We can improve this rate as in the next result.

**Proposition 4.1.** *Under Assumptions (A1)–(A3), we have that $\|\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}\|_\infty^2 = O_p(2^J T^{-1} \log(T))$.*

This proposition allows us to derive the rate of convergence of the mean square error of the original linear estimators of the wavelet coefficients.

**Proposition 4.2.** *If Assumptions (A1)–(A3) hold, then $E(\hat{\beta}_{jk}^{(i)} - \beta_{jk}^{(i)})^2 = O(T^{-1})$ holds uniformly in $i$, $k$ and $j < J$.*

## 5. Some Simulations

We now present three simulation examples for the linear and nonlinear estimates $\hat{\delta}_i(\cdot)$, $\hat{w}_i(\cdot)$ and $\tilde{\delta}_i(\cdot)$, $\tilde{w}_i(\cdot)$, respectively. In the first, the coefficients are very smooth functions, the second presents an intermediary case, and, in the third, the functions are piecewise constants.

The performance of the estimators are assessed via the square root of the mean squared errors (RMSE), namely,

$$
\begin{aligned}
\text{RMSE}(\delta)_j &= \left\{ \frac{1}{T-v} \sum_{t=v+1}^{T} \left[ \hat{\delta}_j\left(\frac{t}{T}\right) - \delta_j\left(\frac{t}{T}\right) \right]^2 \right\}^{1/2}, \\
\text{RMSE}(w)_j &= \left\{ \frac{1}{T-v} \sum_{t=v+1}^{T} \left[ \hat{w}_j\left(\frac{t}{T}\right) - w_j\left(\frac{t}{T}\right) \right]^2 \right\}^{1/2},
\end{aligned}
\tag{5.1}
$$

the same for the nonlinear estimators. Recall that $v = \max(m, n)$.

Concerning the choice of the wavelet, we have tried the Daubechies Extremal Phase D8 and D12, the Daubechies Least Asymmetric S8 and S12 and the Haar wavalets. The best results were obtained with the wavelet mentioned in each example. Of course, in a real problem situation, this cannot be done, since we do not know the true function coefficients. In our experience, the Haar and D8 wavelets have proven useful.

Other important issues are the choice of $J$, how to estimate $\sigma^2$, the noise variance, and the choice of the threshold parameter $\lambda$. Concerning $J$, we have used the rule specified in Section 4, below (4.3), resulting $J = 6$ for all examples. The variance was estimated using the empirical wavelet coefficients in all scales. Hard thresholding was used with the parameter $\lambda$ chosen as $\lambda = \hat{\sigma}\sqrt{2\log(2^J)}$.

In all cases, the input time series was generated as an AR(2) model, $X_{t,T} = a_1(t/T) X_{(t-1),T} + a_2(t/T) X_{(t-2),T} + \varsigma_t$, where $\varsigma_t$ are independent, normally distributed random variables, with zero mean and variance one, for $t = 1, 2, \ldots, T$, and the coefficients are given by

$$
a_1(u) = \begin{cases} 1.69, & 0 < u \le 0.6, \\ -0.3, & 0.6 < u \le 1, \end{cases}
\tag{5.2}
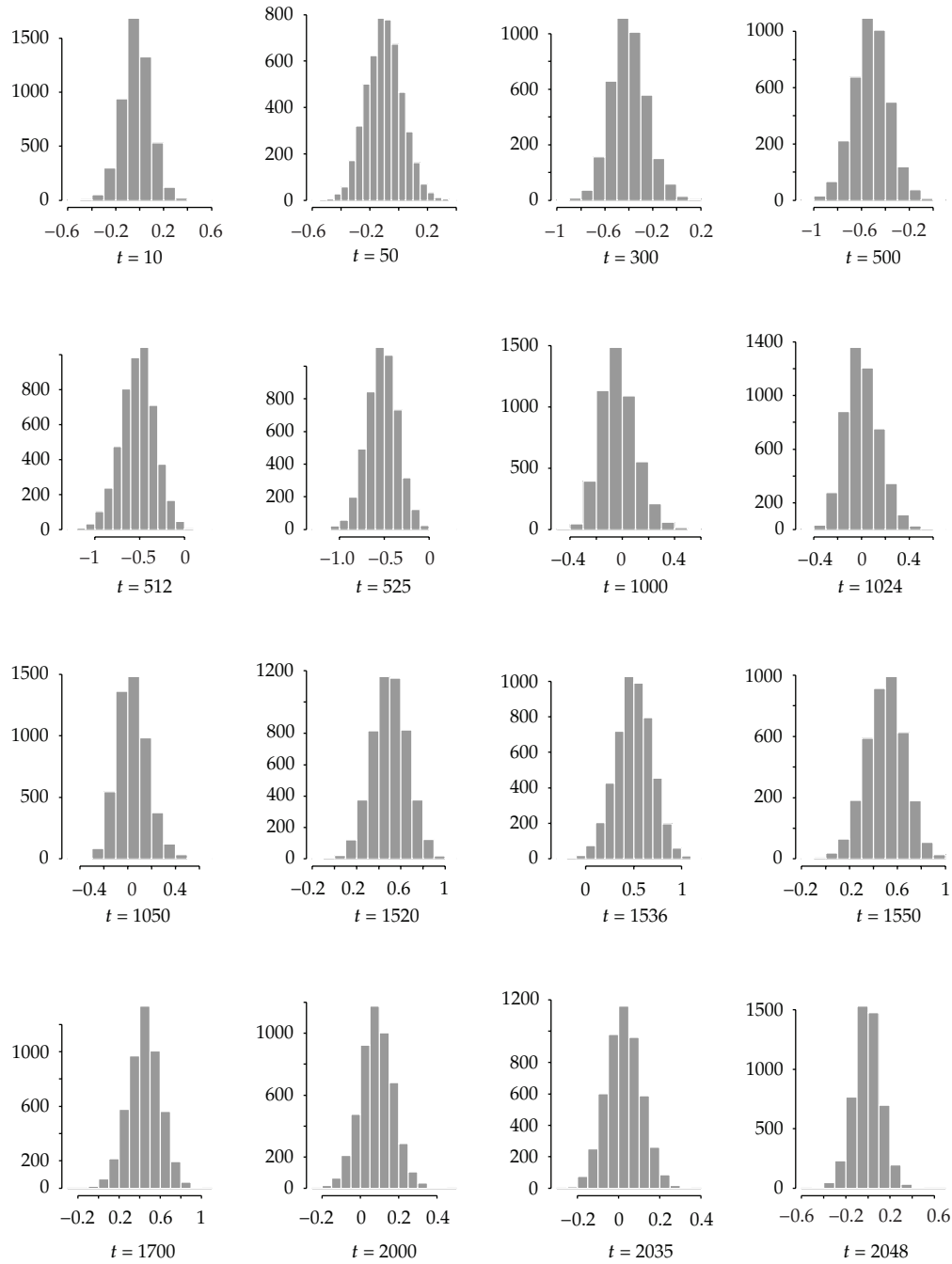$$

$$
a_2(u) = -0.81 \quad 0 < u \le 1.
$$

**Figure 17:** Example 5.1: Histograms of linear estimate of $\delta_1$ at some points of [0,1].

*Example 5.1.* Let $m = 2$ and $n = 0$, so the transfer function model is

$$Y_{t,T} = \delta_1\left(\frac{t}{T}\right)Y_{t-1,T} + \delta_2\left(\frac{t}{T}\right)Y_{t-2,T} + \omega_0\left(\frac{t}{T}\right)X_{t,T} + \epsilon_t, \quad t = 1, 2, \ldots, T. \tag{5.3}$$
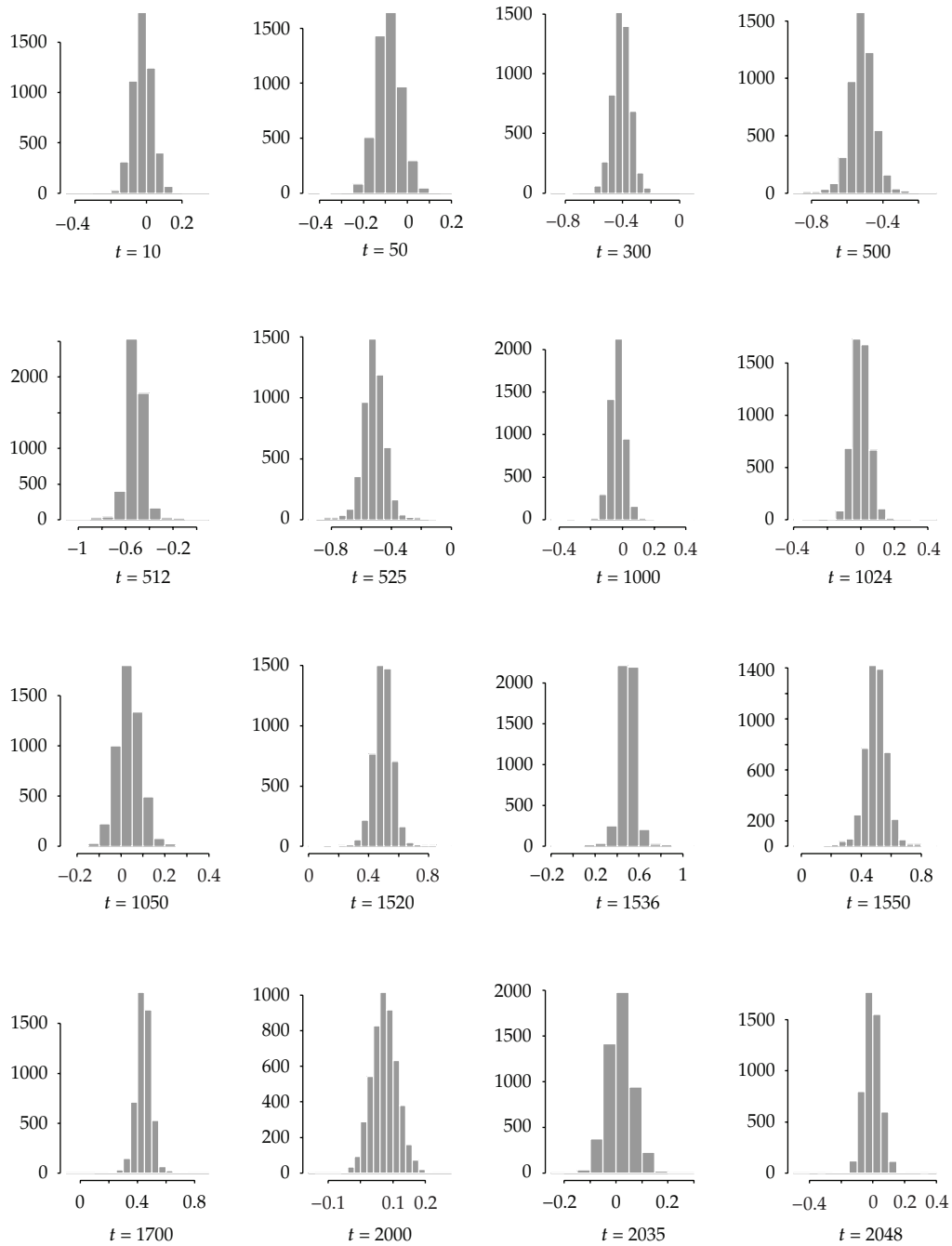
**Figure 18:** Example 5.1: Histograms of nonlinear estimate of $\delta_1$ at some points of $[0, 1]$.

We choose $\omega_0(u)$, $\delta_1(u)$, and $\delta_2(u)$ as $\omega_0(u) = 0.9\cos(2\pi u + \pi)$, $\delta_1(u) = -0.5\sin(2\pi u)$, and $\delta_2(u) = -0.4\cos(2\pi u + \pi/4)$.

For the simulation, we generated $T = 2048$ data values for $X_{t,T}$ and $Y_{t,T}$ with 5000 replicates. For this example, the Daubechies D12 wavelet is employed for our estimation procedure.
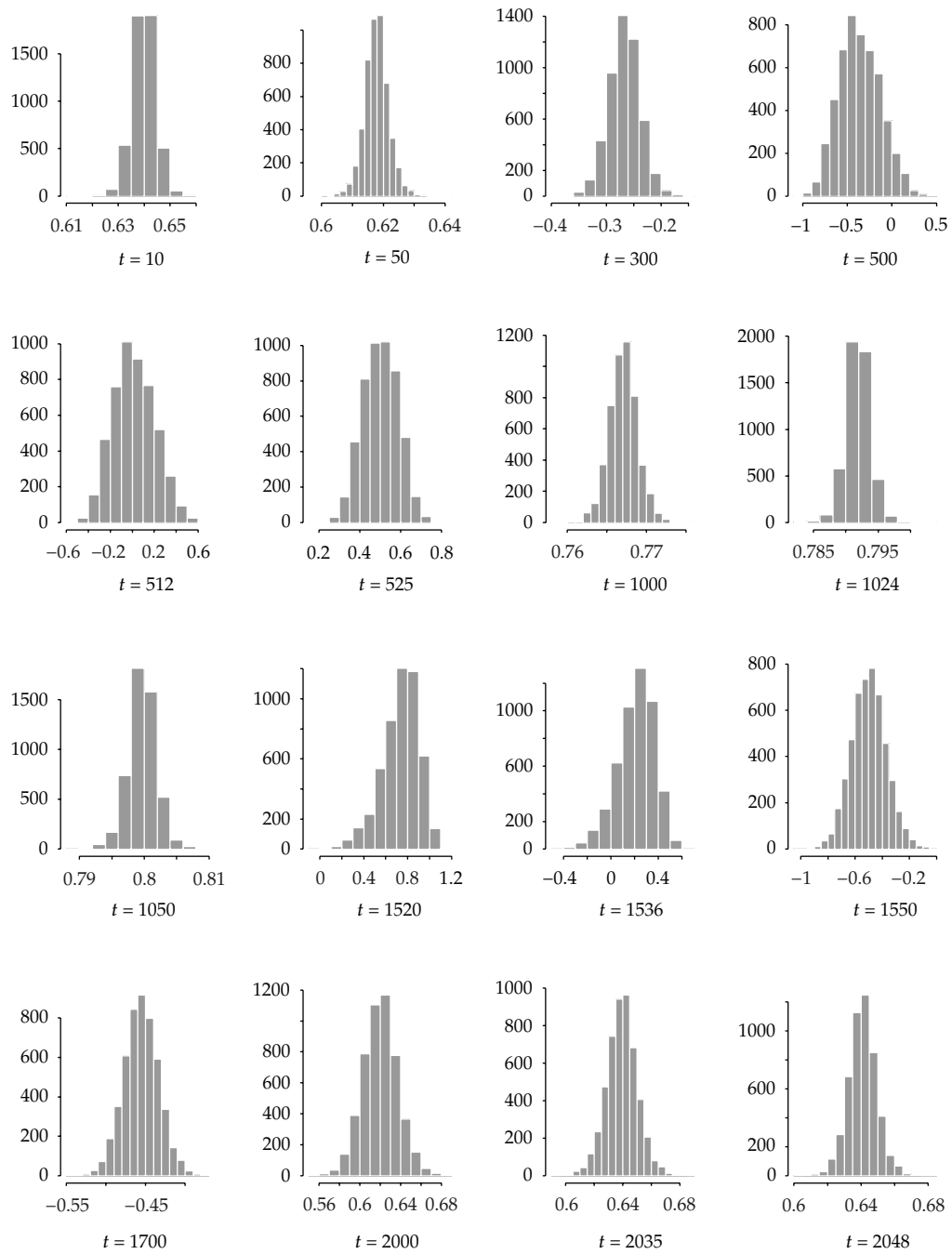
**Figure 19:** Example 5.2: Histograms of linear estimate of $\delta_1$ at some points of $[0, 1]$.

Figure 1 presents the input and output series, and Figure 2 presents the true and estimated coefficients from a typical sample. The boxplots for the 5000 RMSE values are presented in Figure 3. We see that both types of estimators perform well, but clearly the nonlinear estimator perform, better, both visually and in RMSE sense.
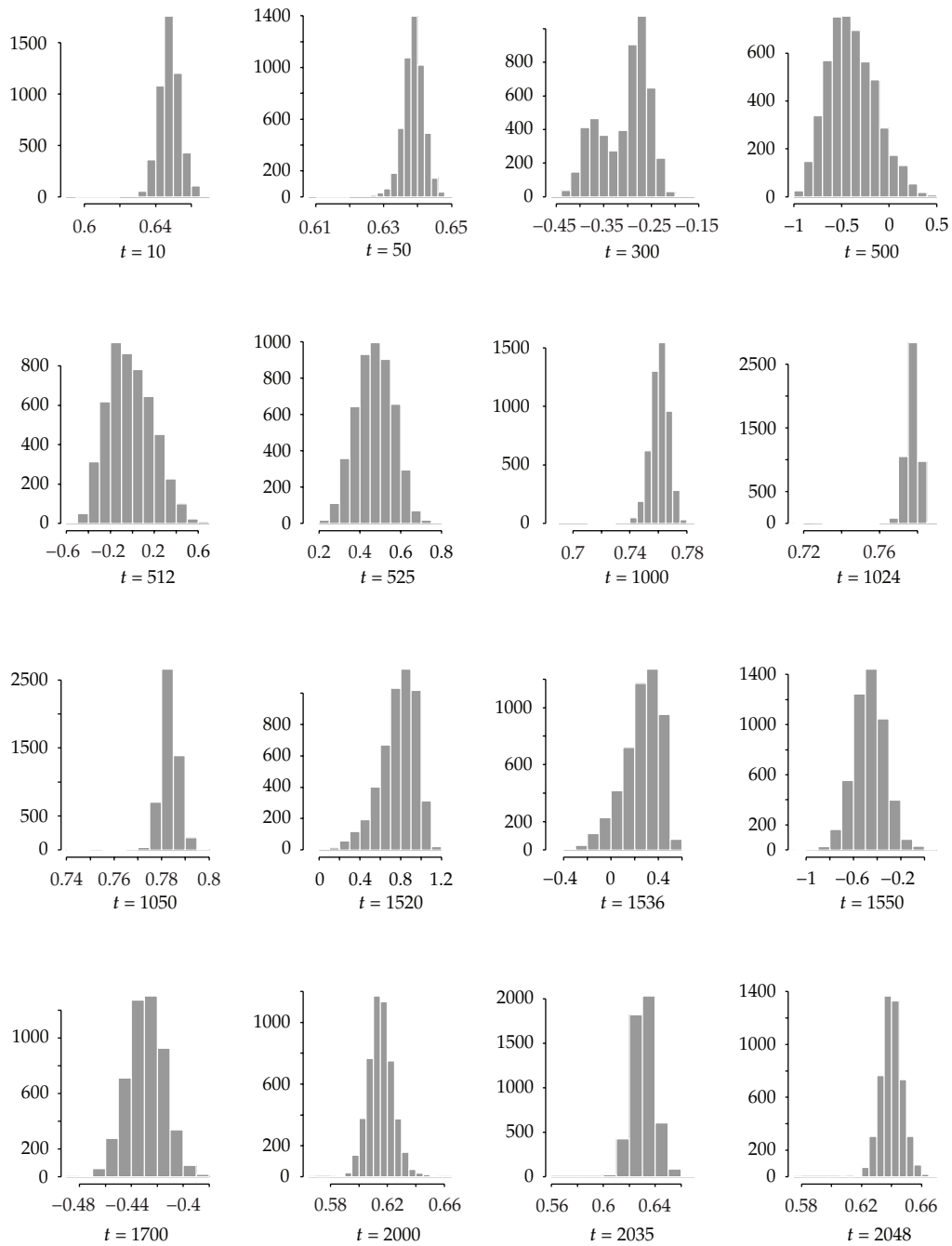
**Figure 20:** Example 5.2: Histograms of nonlinear estimate of $\delta_1$ at some points of $[0, 1]$.

In Figure 4, we use bootstrap with $B = 300$ replications to see that the true transfer function coefficients lie within the 95% confidence regions. See Section 6 for the description of the bootstrap procedure.
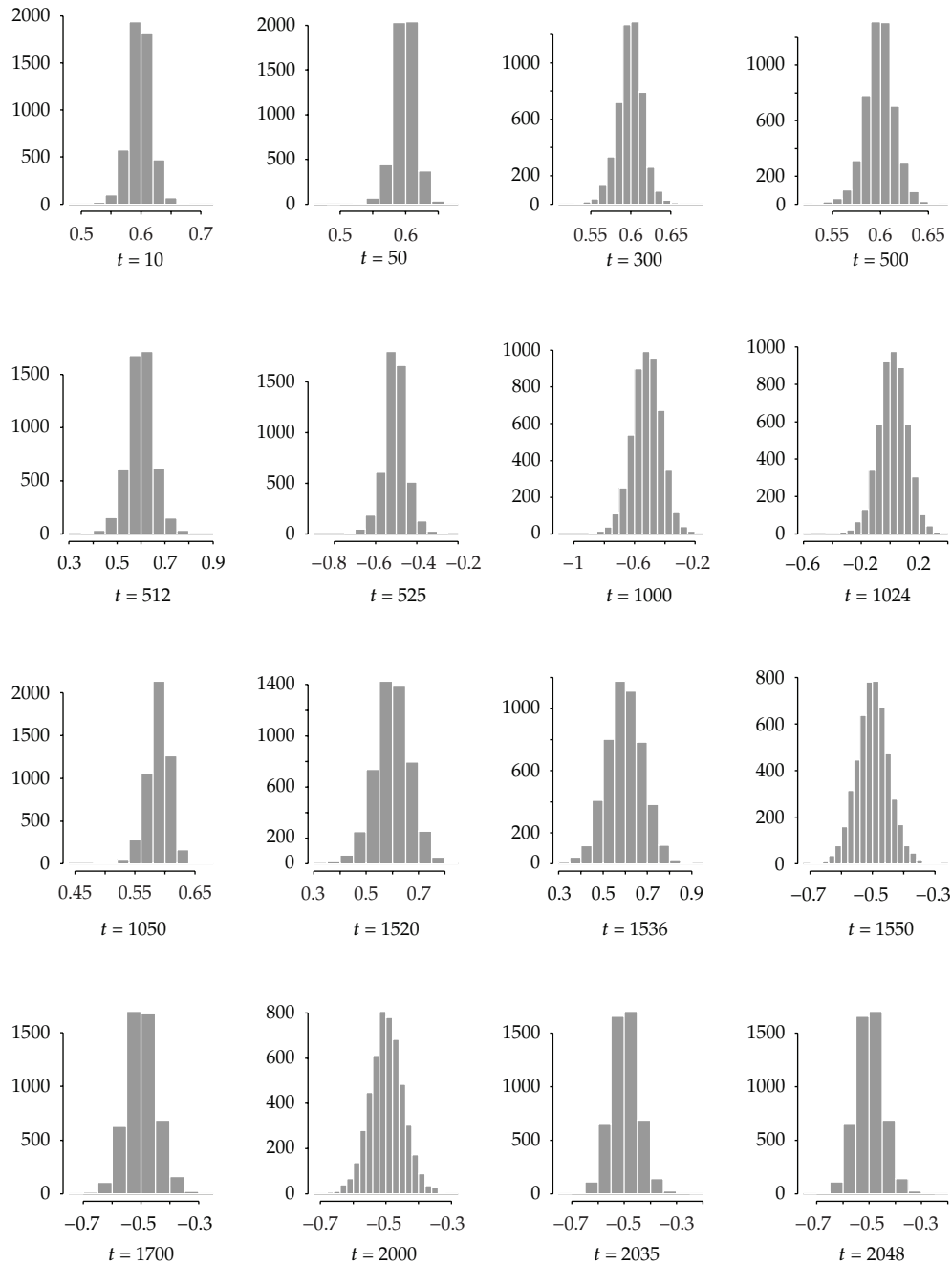
**Figure 21:** Example 5.3: Histograms of linear estimate of $\delta_1$ at some points of $[0, 1]$.

In Figures 17 and 18 of the appendix, we show histograms for linear and nonlinear estimators of the parameter $\delta_1$ at some points of the interval $[0, 1]$. Notice that the normal approximation for the estimators is a possibility to be considered under the theoretical point of view.
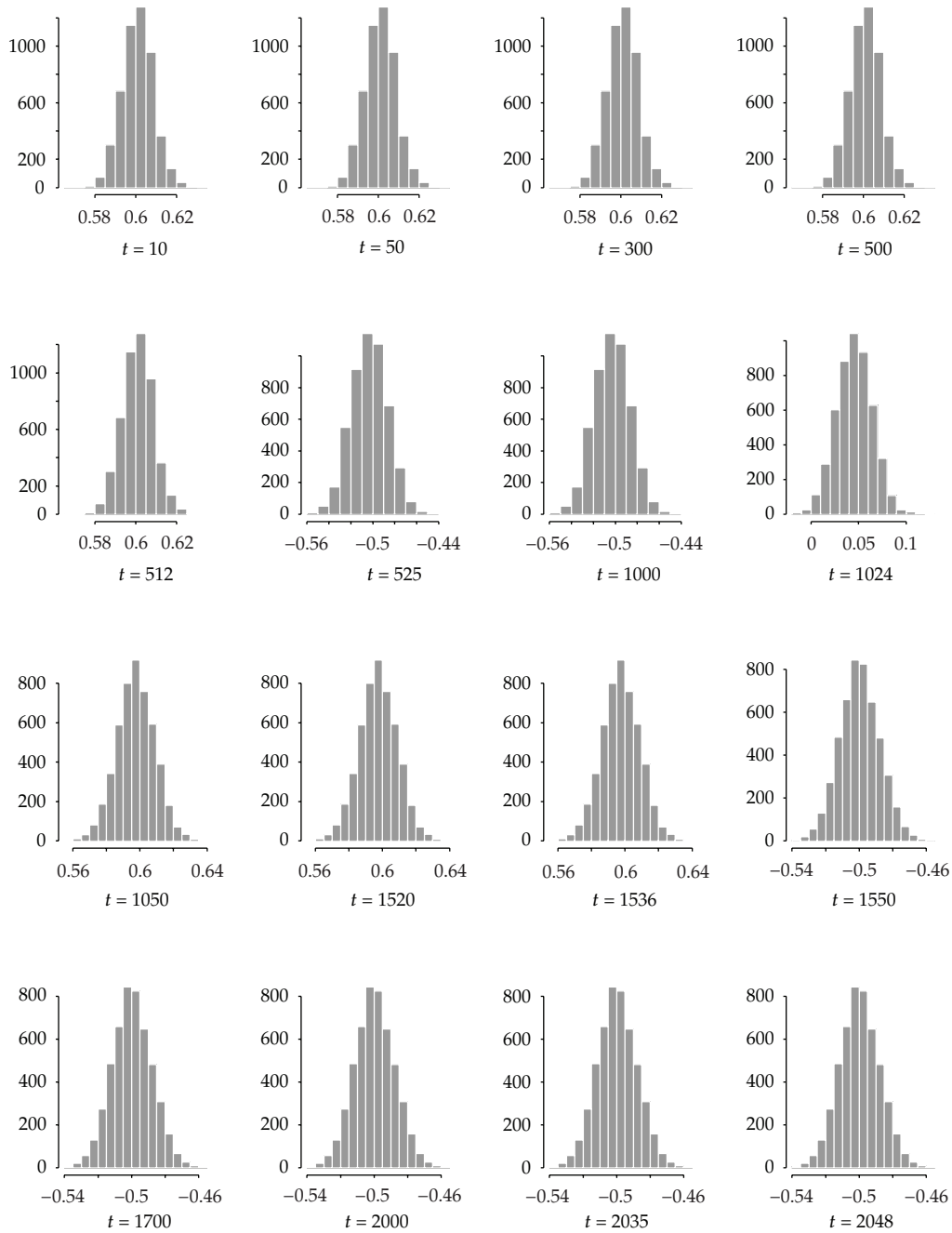
**Figure 22:** Example 5.3: Histograms of nonlinear estimate of $\delta_1$ at some points of $[0, 1]$.

*Example 5.2.* Let $m = 1$ and $n = 0$; so the transfer function model is

$$Y_{t,T} = \delta_1\left(\frac{t}{T}\right)Y_{t-1,T} + \omega_0\left(\frac{t}{T}\right)X_{t,T} + \epsilon_t, \quad t = 1, 2, \dots, T. \tag{5.4}$$

We choose $\omega_0(u)$ and $\delta_1(u)$ as $\omega_0(u) = 9\cos(2\pi u + \pi)$, $u \in (0,1]$; $\delta_1(u) = -0.8\cos(1.5 - \cos(4\pi u + \pi))$, for $0 < u \leq 0.25$ and $0.75 < u \leq 1$, and $\delta_1(u) = -0.8\cos(3 - \cos(4\pi u + \pi/2))$, for $0.25 < u \leq 0.75$.

As in the Example 5.1, we generated $T = 2048$ data values for $X_{t,T}$ and $Y_{t,T}$ with 5000 replicates and the Daubechies Extremal Phase wavelet D8 is employed.

From a typical sample, the input and output series are shown in Figure 5 and the estimated coefficients in Figure 6. The boxplots for the 5000 RMSE values are presented in Figure 7. We see that, visually, both types of estimators perform well, but the thresholded estimator is better for the smoother parameter $\omega_0$ in terms of RMSE. This is a well-known fact, see, for example, Percival and Walden [11]. In Figure 8, we use bootstrap with $B = 300$ replications to see that the true transfer function coefficients lie within the 95% confidence regions.

In Figures 6(c) and 6(e), we see some unwanted artifacts next to discontinuity points. These do not disappear, even with expansions with many terms. But this Gibbs effect is less severe than in Fourier expansions. See Vidakovic [10] and Percival and Walden [36] for details.

In Figures 19 and 20 of the appendix, we show histograms for linear and nonlinear estimators of the parameter $\delta_1$ at some points of the interval $[0,1]$.

*Example 5.3.* Let Let $m = 1$ and $n = 0$; so the transfer function model is as in Example 5.2.

We choose the coefficients $\omega_0(u)$ and $\delta_1(u)$ as piecewise constant functions: $\omega_0(u) = 2$, for $0 < u \leq 0.5$ and $\omega_0(u) = -2$, for $0.5 < u \leq 1$; $\delta_1(u) = 0.6$, for $0 < u \leq 0.25$ or $0.50 < u \leq 0.75$, and $\delta_1(u) = -0.5$, for $0.25 < u \leq 0.5$ or $0.75 < u \leq 1$.

We simulated $T = 2048$ data values for $X_{t,T}$ and $Y_{t,T}$ with 5000 replicates, and, due to the nature of the functions, the Haar wavelet is appropriate in this situation.

Figure 9 presents the input and output series, and Figure 10 presents the true and estimated coefficients from a typical sample. The boxplots for the 5000 RMSE values are presented in Figure 11. We see that both types of estimators perform well, but clearly the nonlinear estimator perform better, both visually and in RMSE sense.

In Figure 12, we use bootstrap with $B = 300$ replications to see that the true transfer function coefficients lie within the 95% confidence regions.

In Figures 21 and 22 of the appendix, we present histograms for linear and nonlinear estimators of $\delta_1$ at some points of the interval $[0,1]$.

## 6. An Application

The data that we analyze are hourly wind speeds recorded from 00:00 on 6th April 1995 at two Welsh Meteorological Office stations: Valley and Aberporth, with 2048 observations. Valley is located approximately 120 km north of Aberporth, and they are mostly separated by Cardigan Bay. Part of these data were analysed by Nason and Sapatinas [1]. Figure 13 shows the data we analyze. We are grateful to N&M Wind Power for providing the data.

In this application, our aim is to model Valley's wind speeds, $\{Y_{t,T}\}$, in terms of its lagged values and present and lagged values of wind speeds at Aberporth, $\{X_{t,T}\}$. We should

remember that both series are not stationary and so classical methods should not be used. We fitted four potential models, namely,

$$
\begin{aligned}
M_1 : \ & y_{t,T} = \delta_1\!\left(\frac{t}{T}\right)y_{t-1,T} + \omega_0\!\left(\frac{t}{T}\right)x_{t,T} + \epsilon_t, \\[4pt]
M_2 : \ & y_{t,T} = \delta_1\!\left(\frac{t}{T}\right)y_{t-1,T} + \omega_0\!\left(\frac{t}{T}\right)x_{t,T} + \omega_1\!\left(\frac{t}{T}\right)x_{t-1,T} + \epsilon_t, \\[4pt]
M_3 : \ & y_{t,T} = \delta_1\!\left(\frac{t}{T}\right)y_{t-1,T} + \delta_2\!\left(\frac{t}{T}\right)y_{t-2,T} + \omega_0\!\left(\frac{t}{T}\right)x_{t,T} + \epsilon_t, \\[4pt]
M_4 : \ & y_{t,T} = \delta_1\!\left(\frac{t}{T}\right)y_{t-1,T} + \delta_2\!\left(\frac{t}{T}\right)y_{t-2,T} + \omega_0\!\left(\frac{t}{T}\right)x_{t,T} + \omega_1\!\left(\frac{t}{T}\right)x_{t-1,T} + \epsilon_t.
\end{aligned}
\tag{6.1}
$$

The Daubechies Extremal Phase D8 wavelet was used in the computations. The model $M_4$ was marginally superior (in terms of RMSE) to model $M_3$, so it was discarded. The mean residual sums of squares for $M_1$, $M_2$, and $M_3$ are 0.02258, 0.02186, and 0.02177, respectively, for linear estimation. For nonlinear estimation, the figures are 0.02391, 0.02359, and 0.02178, respectively. The final coefficient estimates for model $M_3$ are shown in Figure 14.

For Model $M_3$, Figures 15 and 16 show the 95% confidence regions based on bootstrap, using linear and nonlinear estimates, respectively.

The bootstrap procedure works as follows:

(a) using the real series, we obtain the estimated transfer function coefficients;

(b) we obtain the fitted series $Y_{t,T}$ using these estimated coefficients;

(c) we obtain the residuals from the fitted model;

(d) we resample these residuals and obtain bootstrap samples of the $Y_{t,T}$ series;

(e) we obtain bootstrap estimators of the transfer function coefficients.

We used $B = 300$ bootstrap replications of $\hat{\delta}_1$, $\hat{\delta}_2$, $\hat{\omega}_0$, and $\tilde{\delta}_1$, $\tilde{\delta}_2$, $\tilde{\omega}_0$, to obtain bootstrap standard errors and the 95% confidence regions. We also fitted the model $M_3$ with constant coefficients given an MSE value of 0.02354, indicating that the model with varying coefficients is better. We see that is not possible to plot constant lines totally within these regions, showing that, in fact, the coefficients are time varying.

## 7. Further Remarks

In this paper, we have proposed an estimation procedure for a transfer function model with time-varying coefficients. Basically it is a least squares procedure, with the use of wavelets to expand the function coefficients. Firstly, linear estimators for the time-varying coefficients are obtained, truncating the wavelet expansion at an appropriate scale. Then thresholds are applied to the empirical wavelet coefficients, and inverse wavelet transformation gives the nonlinear smoothed estimators for the function coefficients.

Simulations have shown that this procedure leads to estimators with a good performance. Some statistical properties for the empirical wavelet coefficients in the case of linear estimators were presented.

Further studies are needed on the estimation of the variance and asymptotic normality of the empirical wavelet coefficients in the linear case, on the rate of convergence for the risk of the nonlinear estimator over the smoothness classes $\mathcal{F}_i$ and on issues related to the identification, diagnostics, and forecasting for these nonstationary models.

The issue of forecasting was discussed in the literature for some special and related models. Fryzlewicz et al. [37] addressed the problem of how to forecast locally stationary wavelet processes in the sense of Nason et al. [4] by means of nondecimated wavelets. Van Bellegem and von Sachs [38] consider a special modulated locally stationary process, where a stationary process is modulated by a time-varying variance, so the problem reduces to forecast this variance.

Concerning asymptotic distribution of the empirical wavelet estimators, in the linear case, we believe that this can be obtained under further assumptions, namely on the cumulants of the $\varepsilon_t$, on the thresholds used and on the roots of the polynomials associated with the model (1.1). This will be the subject of further research.

## Appendix

For more details, see Figures 17, 18, 19, 20, 21, and 22.

## Acknowledgments

## References

[1] G. P. Nason and T. Sapatinas, "Wavelet packet transfer function modelling of nonstationary time series," *Statistics and Computing*, vol. 12, no. 1, pp. 45–56, 2002.

[2] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecastingand Control*, Prentice Hall, Englewood Cliffs, NJ, USA, 3rd edition, 1994.

[3] R. Dahlhaus, "Fitting time series models to nonstationary processes," *The Annals of Statistics*, vol. 25, no. 1, pp. 1–37, 1997.

[4] G. P. Nason, R. von Sachs, and G. Kroisandt, "Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum," *Journal of the Royal Statistical Society. Series B*, vol. 62, no. 2, pp. 271–292, 2000.

[5] C. Chiann and P. A. Morettin, "Estimation of time varying linear systems," *Statistical Inference for Stochastic Processes*, vol. 2, no. 3, pp. 253–285, 1999.

[6] C. Chiann and P. A. Morettin, "Time-domain estimation of time-varying linear systems," *Journal of Nonparametric Statistics*, vol. 17, no. 3, pp. 365–383, 2005.

[7] R. Dahlhaus, M. H. Neumann, and R. von Sachs, "Nonlinear wavelet estimation of time-varying autoregressive processes," *Bernoulli*, vol. 5, no. 5, pp. 873–906, 1999.

[8] P. J. Brockwell, R. A. Davis, and H. Salehi, "Maximum likelihood transfer functionmodelling. Computer Science and Statistics," in *Proceedings of the 22nd Symposium on theInterface*, R. Le Page and C. Page, Eds., Springer, New York, NY, USA, 1991.

[9] P. J. Brockwell, R. A. Davis, and H. Salehi, "Transfer-function models with nonstationary input," in *New Directions in Time Series Analysis, Part I*, vol. 45 of *The IMA Volumes in Mathematics and Its Applications*, pp. 65–74, Springer, New York, NY, USA, 1992.

[10] B. Vidakovic, *Statistical Modelling by Wavelets*, John Wiley & Sons, New York, NY, USA, 1999.

[11] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*, vol. 4 of *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press, Cambridge, UK, 2000.

[12] G. P. Nason, *Wavelet Methods in Statistics with R*, Use R!, Springer, New York, NY, USA, 2008.

[13] I. Daubechies, *Ten Lectures on Wavelets*, vol. 61, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa, USA, 1992.

[14] A. Cohen, I. Daubechies, and P. Vial, "Wavelets on the interval and fast wavelet transforms," *Applied and Computational Harmonic Analysis*, vol. 1, no. 1, pp. 54–81, 1993.

[15] G. G. Walter and L. Cai, "Periodic wavelets from scratch," *Journal of Computational Analysis and Applications*, vol. 1, no. 1, pp. 25–41, 1999.

[16] Y. Meyer, *Ondelettes et Opérateurs. I*, Actualités Mathématiques, Hermann, Paris, France, 1990.

[17] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[18] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200–1224, 1995.

[19] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard, "Wavelet shrinkage: asymptopia?" *Journal of the Royal Statistical Society. Series B*, vol. 57, no. 2, pp. 301–369, 1995.

[20] I. M. Johnstone and B. W. Silverman, "Wavelet threshold estimators for data with correlated noise," *Journal of the Royal Statistical Society. Series B*, vol. 59, no. 2, pp. 319–351, 1997.

[21] J. Opsomer, Y. Wang, and Y. Yang, "Nonparametric regression with correlated errors," *Statistical Science*, vol. 16, no. 2, pp. 134–153, 2001.

[22] V. Delouille, *Nonparametric stochastic regression using design-adapted wavelets*, Ph.D. thesis, Institut de Statistique, Université catholique de Louvain, 2002.

[23] G. Kerkyacharian and D. Picard, "Regression in random design and warped wavelets," *Bernoulli*, vol. 10, no. 6, pp. 1053–1105, 2004.

[24] T. T. Cai and L. D. Brown, "Wavelet shrinkage for nonequispaced samples," *The Annals of Statistics*, vol. 26, no. 5, pp. 1783–1799, 1998.

[25] T. T. Cai and L. D. Brown, "Wavelet estimation for samples with random uniform design," *Statistics & Probability Letters*, vol. 42, no. 3, pp. 313–321, 1999.

[26] W. Sweldens, "The lifting scheme: a construction of second generation wavelets," *SIAM Journal on Mathematical Analysis*, vol. 29, no. 2, pp. 511–546, 1998.

[27] M. Jansen and P. Oonincx, *Second Generation Wavelets and Applications*, Springer, London, UK, 2005.

[28] R. F. Porto, P. A. Morettin, and E. C. Q. Aubin, "Wavelet regression with correlated errors on a piecewise Hölder class," *Statistics & Probability Letters*, vol. 78, no. 16, pp. 2739–2743, 2008.

[29] R. F. Porto, P. A. Morettin, and E. C. Q. Aubin, "Smoothed design-adapted Haar wavelet regression with autocorrelated errors," *Journal of Time Series Econometrics*, to appear.

[30] P. A. Morettin and C. Chiann, "Estimation of functional-coefficient autoregressive models by wavelet methods," *Current Development in Theory and Applications of Wavelets*, vol. 1, no. 3, pp. 309–340, 2007.

[31] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B*, vol. 58, no. 1, pp. 267–288, 1996.

[32] L. Breiman, "Better subset regression using the nonnegative garrote," *Technometrics*, vol. 37, no. 4, pp. 373–384, 1995.

[33] H. Triebel, *Theory of Function Spaces II*, Birkhäuser, Basel, Switzerland, 1992.

[34] R. D. Martin, V. J. Yohai, and R. H. Zamar, "Min-max bias robust regression," *The Annals of Statistics*, vol. 17, no. 4, pp. 1608–1630, 1989.

[35] W. Härdle, G. Kerkyacharian, D. Picard, and A. Tsybakov, *Wavelets, Approximation, and Statistical Applications*, vol. 129 of *Lecture Notes in Statistics*, Springer, New York, NY, USA, 1998.

[36] D. B. Percival and A. T. Walden, *Spectral Analysis for Physical Applications*, Cambridge University Press, Cambridge, UK, 1993.

[37] P. Fryzlewicz, S. Van Bellegem, and R. von Sachs, "Forecasting non-stationary time series by wavelet process modelling," *Annals of the Institute of Statistical Mathematics*, vol. 55, no. 4, pp. 737–764, 2003.

[38] S. Van Bellegem and R. von Sachs, "Forecasting economic time series with unconditional time-varying variance," *International Journal of Forecasting*, vol. 20, no. 4, pp. 611–627, 2004.

Advances in
Operations Research

Advances in
Decision Sciences

Journal of
Applied Mathematics

Algebra

Journal of
Probability and Statistics

The Scientific
World Journal

International Journal of
Differential Equations

International Journal of
Combinatorics

Advances in
Mathematical Physics

Journal of
Complex Analysis

Journal of
Mathematics

Mathematical Problems
in Engineering

Abstract and
Applied Analysis

Discrete Dynamics in
Nature and Society

International
Journal of
Mathematics and
Mathematical
Sciences

Journal of
Discrete Mathematics

Journal of
Function Spaces

International Journal of
Stochastic Analysis

Journal of
Optimization