# Wavelet Estimation of Functional Coefficient Regression Models

Michel H. Montoril[1], Pedro A. Morettin[2], and Chang Chiann[2]

[1]Department of Statistics, Institute of Exact Sciences , Federal University of Juiz de Fora, Brazil
[2]Department of Statistics, Institute of Mathematics and Statistics, University of São Paulo, Brazil

# Introduction

Nonlinear time series models started to be strongly developed from the 1980s (see e.g. Tong, 1993) with parametric models.

## Nonlinear Models:

- Exponential autoregressive model (EXPAR)

  Haggan and Ozaki (1981)

- Threshold autoregressive model (TAR)

  Tong (1983)

- Functional-coefficient autoregressive model (FAR)

  Chen and Tsay (1993)

- Functional coefficient regression model (FCR)

  Cai, Z., Fan, J. and Yao, Q. (2000)
  Huang and Shen (2004); Montoril et al. (2014)

# Models

## Functional coefficient regression model (FCR)

Let $\{Y_t, U_t, \boldsymbol{X}_t\}$ be a jointly strictly stationary process, where $U_t$ is a real random variable and $\boldsymbol{X}_t$ a random vector in $\mathbb{R}^d$. Suppose that $\mathbb{E}(Y_t^2) < \infty$. Considering the multivariate regression function $m(\boldsymbol{x}, u) = \mathbb{E}(Y_t | \boldsymbol{X}_t = \boldsymbol{x}, U_t = u)$, the FCR model has the form

$$m(\boldsymbol{x}, u) = \sum_{i=1}^{d} f_i(u) x_i, \tag{1}$$

where the $f_i(\cdot)$s are measurable functions from $\mathbb{R}$ to $\mathbb{R}$ and $\boldsymbol{x} = (x_1, \ldots, x_d)^\top$, with $\top$ denoting the transpose of a matrix or vector. Frequently, the coefficient functions are assumed to be compactly supported in some closed interval $\mathcal{C}$. For the sake of simplicity, assume that $\mathcal{C} = [0, 1]$. In the nonparametric framework, when $U_t$ and $\boldsymbol{X}_t$ are lagged values of $Y_t$, FCR models correspond to the functional-coefficient autoregressive (FAR) models of Chen and Tsay (1993).

# Wavelet-based estimator

By wavelet basis we know that it has an associated multiresolution analysis (MRA), where a sequence of nested and closed subspaces $\{V_j\}_{j\in\mathbb{Z}}$ of $L_2(\mathbb{R})$ satisfies the following properties:

- $V_j \subset V_{j+1}$;
- $f(\cdot) \in V_j \iff f(2\cdot) \in V_{j+1}$;
- $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$;
- $\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L_2(\mathbb{R})$;
- There exists a function $\varphi \in V_0$ such that $\{\varphi(\cdot - k)\}_{k\in\mathbb{Z}}$ is a Riesz basis for $V_0$.

**$V_J$:** $\varphi_{Jk}(x) = 2^{J/2}\varphi(2^J x-k)$, J: resolution level

**$W_j$ of $V_j$ in $V_{j+1}$:** $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x-k), j,k \in Z.$

# Wavelet-based estimator

Frequently, studies related to wavelet-based estimators require assumptions like equally spaced data and dyadic (power of two) sample sizes.

However, in this work we use the Daubechies-Lagarias algorithm (Daubechies and Lagarias, 1991, 1992).

This algorithm is an iterative method useful for computing values of compactly supported orthonormal wavelet functions (for example, Daublets and Symmlets) in specific points of interest with preassigned precision (more details in Vidakovic, 1999).

Besides, this algorithm has the advantage of not requiring equally spaced data nor dyadic sample sizes.

# Wavelet-based estimator

Warped wavelets (Kerkyacharian and Picard (2004), Cai and Brown (1998, 1999))

In order to explain how the warped wavelets work in our case, let $H$ be a continuous distribution function that is assumed to "spread" the data in the unit interval, and $H^{-1}$ its inverse. For the sake of simplicity, let us suppress the subscript $i$ of the coefficient function $f_i$. Denote $g = f \circ H^{-1}$ and $y = H(u)$, $u \in [0,1]$. Then we have that the orthogonal projection $g^J$ of $g$, for some resolution level $J$, will be

$$g^J(y) = \sum_k \alpha_k \varphi_{Jk}(y), \tag{2}$$

where $y \in [0,1]$. Thus, an approximation to the function $f$ will be

$$f(u) = g(H(u)) \approx g^J(H(u)) = \sum_k \alpha_k \varphi_{Jk}(H(u)).$$

Note that, even in the case where the basis $\{\varphi_{Jk}\}_k$ is orthogonal, $f$ is not approximated by the expansion of an orthogonal basis, unless if $H(u) = u$, $u \in [0,1]$. This happens because $\varphi_{Jk}$ is "warped" by the distribution function $H$. A consequence of this is

$$\alpha_k = \int_0^1 g^J(y)\varphi_{Jk}(y)dy = \int_0^1 f^*(u)\varphi_{Jk}(H(u))du,$$

where $f^* = g^J \circ H$.

# Wavelet-based estimator

The main idea of considering warped wavelets is that, when the data set is too concentrated in some specific region, this region will have more precise estimates, while the region with lower density will have estimates with higher variability.

In our case, since we are using the Daubechies-Lagarias algorithm, it is not necessary to deal with equally spaced data sets.

Our main interest is to be able to apply (if necessary) some transformation so that the data used can be close to a uniform distribution.

In this case, the variability of the coefficient function estimates must be about the same in different regions of the support of these functions. An interesting feature of this approach is that it has the classical wavelets as a particular case, when $H(u) = u$.

# Wavelet-based estimator

It is important to mention that, in this situation, it is interesting to consider wavelets defined in the unit interval, because $H(u) \in [0,1]$, $u \in [0,1]$. Periodized wavelets can be defined in such an interval, and they have the advantage of handling boundary conditions. They are denoted by

$$\varphi^p_{jk}(u) = \sum_l \varphi_{jk}(u-l),\ \psi^p_{jk}(u) = \sum_l \psi_{jk}(u-l),\ u \in [0,1],$$

where $j \in \mathbb{Z}$ and $k = 1, \ldots, 2^j$.

It is possible to see that the basis $\{\varphi^p_{Jk}\}_k$ has an associated MRA in $[0,1]$ and, if $\{\varphi_{Jk}\}_k$ is an orthonormal basis, then $\{\varphi^p_{Jk}\}_k$ is orthonormal as well (more details in Restrepo and Leaf, 1997). Hereafter, the superscript "$p$" will be removed from notation for convenience.

# Linear wavelet-based estimator

By model (1), one can think in expressing the process $\{Y_t, U_t, \boldsymbol{X}_t\}$ according to the stochastic representation

$$Y_t = \sum_{i=1}^{d} f_i(U_t) X_{ti} + \epsilon_t,$$

where $\epsilon_t$ corresponds to the errors of the model. Now, let $\boldsymbol{\Sigma}$ be the covariance matrix of the errors and assume initially that it is known. Thus we can estimate the wavelet coefficients in (3) minimizing the least squares function

$$\ell(\boldsymbol{\alpha}) = (\boldsymbol{Y} - \mathbb{V}\boldsymbol{\alpha})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{Y} - \mathbb{V}\boldsymbol{\alpha}), \tag{4}$$

where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^\top, \ldots, \boldsymbol{\alpha}_d^\top)^\top$, $\boldsymbol{\alpha}_i = (\alpha_{i0}, \ldots, \alpha_{i,2^{J_i}-1})^\top$, $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ and $\mathbb{V}$ is a $n \times \sum_{i=1}^d 2^{J_i}$ matrix such that its $t$-th row corresponds to $\phi_{ik}(H(U_t))X_{ti}$, $i = 1, \ldots, d$, $k = 0, 2, \ldots, 2^{J_i} - 1$. The coefficient vector estimator is

$$\hat{\alpha} = (\mathbb{V}^\top \boldsymbol{\Sigma}^{-1} \mathbb{V})^{-1} \mathbb{V}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{Y}. \tag{5}$$

Note that we can assume $\boldsymbol{\Sigma}$ as an identity matrix when the errors are uncorrelated and homoscedastic.

Based on the wavelet coefficient estimates, it is possible to estimate each coefficient function by

$$\hat{f}_i(u) = \sum_{k=0}^{2^{J_i}-1} \hat{\alpha}_{ik} \phi_{ik}(H(u)),$$

where $\hat{\boldsymbol{\alpha}}_i = (\hat{\alpha}_{i0}, \ldots, \hat{\alpha}_{i,2^{J_i}-1})^\top$ is the estimator of $\boldsymbol{\alpha}_i$, $i = 1, 2, \ldots, d$.

# Theoretical results

Since the wavelet coefficients depend on the distribution function $H$ and are intrinsically related to the functions $f_i \circ H^{-1}$, we study the distance between the estimator and the coefficient function using a norm weighted by the probability density function $h(u) = dH(u)/du$. The weighted norm, with a weight function $w$, of a specific function $f$ can be defined as

$$\|f\|_{L_2(w)} := \left( \int_0^1 f^2(x) w(x) dx \right)^{1/2}.$$

Thus, in our case, the distances will be

$$\begin{aligned} \|\hat{f}_i - f_i\|^2_{L_2(h)} &= \int_0^1 (\hat{f}_i(x) - f_i(x))^2 h(x) dx \\ &= \int_0^1 (\hat{f}_i(H^{-1}(y)) - f_i(H^{-1}(y)))^2 dy \\ &= \|\hat{f}_i \circ H^{-1} - f_i \circ H^{-1}\|_2^2, \end{aligned} \tag{6}$$

where $\|f\|_2 = \left( \int_0^1 f^2(x) dx \right)^{1/2}$. The use of (6), it is worth mentioning, is not new in literature. It is used, for example, in Kulik and Raimondo (2009).

The theoretical results presented in this paper are based on a set of frequently used assumptions. Before exhibiting these assumptions, let us present two symbols that will be used. Let $x_n$ and $y_n$ be two positive sequences. Thus we say that $x_n \lesssim y_n$ if the ratio $x_n/y_n$ is uniformly bounded, and $x_n \asymp y_n$ if $x_n \lesssim y_n$ and $y_n \lesssim x_n$.

# Theoretical results

**Assumptions**

(W0) The eigenvalues of $\mathbf{\Sigma}$ are bounded away from zero and infinity;

(W1) The marginal density of $U_t$ is bounded away from zero and infinity uniformly on $[0,1]$;

(W2) The eigenvalues of $\mathbb{E}(\boldsymbol{X}_t\boldsymbol{X}_t^\top|U_t = u)$ are uniformly bounded away from zero and infinity for all $u \in [0,1]$;

(W3) $2^{j_0} \asymp 2^{J_i} \asymp n^r, 0 < r < 1, i = 1, \ldots, d$;

(W4) The process $\{Y_t, \boldsymbol{X}_t, U_t\}_{t\in\mathbb{Z}}$ is jointly strictly stationary. The $\alpha$-mixing coefficient $\alpha(t)$ of $\{Y_t, \boldsymbol{X}_t, U_t\}_{t\in\mathbb{Z}}$ satisfies $\alpha(t) \lesssim t^{-\alpha}$ for $\alpha > (2+r)/(1-r)$;

(W5) For some sufficient large $m > 0$, $\mathbb{E}|X_{ti}|^m < \infty, i = 1, \ldots, d$;

(W6) The distribution function $H$ used for warping the wavelet basis is continuous and strictly monotone, and its probability density function $h$ is bounded away from zero and infinity uniformly on $[0,1]$.

# Theoretical results

Let us denote by $g_i^{J_i}$ the orthogonal projection of $f_i \circ H^{-1}$ onto $V_{J_i}$ and $\rho_i = \|g_i^{J_i} - g_i\|_2$. Thus we can derive rates of convergence to zero for the distances between the wavelet-based estimators and the real coefficient functions, which are presented below.

**Theorem 2.1** *If the assumptions (W0) – (W6) hold, then*

$$\sum_{i=1}^{d} \mathbb{E}\|\hat{f}_i - f_i\|^2_{L_2(h)} \leq C \sum_{i=1}^{d} \left( \frac{2^{J_i}}{n} + \rho_i^2 \right),$$

*for some $C > 0$. In particular, if $\rho_i = o(1)$, then $\mathbb{E}\|\hat{f}_i - f_i\|^2_{L_2(h)} = o(1)$, $i = 1, \ldots, d$.*

# Theoretical results

In practice we replace in (5) the matrix $\mathbf{\Sigma}$ by some estimator, say $\hat{\mathbf{\Sigma}}$, resulting in

$$\bar{\alpha} = (\mathbb{V}^\top \hat{\mathbf{\Sigma}}^{-1} \mathbb{V})^{-1} \mathbb{V}^\top \hat{\mathbf{\Sigma}}^{-1} \boldsymbol{Y}. \tag{7}$$

Then, based on (7), the wavelet-based estimator of the coefficient functions in model (1) can be written as

$$\bar{f}_i(u) = \sum_{k=0}^{2^{J_i}-1} \bar{\alpha}_{ik} \phi_{ik}(H(u)), \quad i = 1, \ldots, d.$$

One can find rates of convergence analogues for the wavelet-based estimator above, whenever the estimator of the covariance matrix $\hat{\mathbf{\Sigma}}$ is consistent in probability. The result follows in the theorem below.

**Theorem 2.2** *If assumptions (W0) – (W6) hold and $\hat{\mathbf{\Sigma}}$ is consistent in probability estimating $\mathbf{\Sigma}$, then*

$$\sum_{i=1}^{d} \|\bar{f}_i - f_i\|^2_{L_2(h)} = O_p\left(\sum_{i=1}^{d}\left(\frac{2^{J_i}}{n} + \rho_i^2\right)\right).$$

*In particular, if $\rho_i = o(1)$, then $\bar{f}_i$ is consistent in probability in estimating $f_i$, i.e., $\|\bar{f}_i - f_i\|_{L_2(h)} = o_p(1)$, $i = 1, \ldots, d$.*

# Algorithm for estimating the coefficient vector

Assuming that the errors are autoregressive, it is possible to rewrite (4) in terms of backshift notation, aiming to minimize the white noise variance. In other words, denoting by $\boxed{\boldsymbol{\eta} \text{ the vector } (\boldsymbol{\alpha}^\top, \boldsymbol{\theta}^\top)^\top}$ and $\mathbb{v}_t$ as the $t$-th row of $\mathbb{V}$, we estimate the wavelet coefficients $\boldsymbol{\alpha}$ and the autoregressive coefficients $\boldsymbol{\theta}$ jointly, minimizing numerically

$$\ell(\boldsymbol{\eta}) = \sum_{t=1}^{n} \left\{ \theta_p(L) \left( Y_t - \mathbb{v}_t^\top \boldsymbol{\alpha} \right) \right\}^2, \tag{8}$$

where $\theta_p(L) = 1 - \theta_1 L - \ldots - \theta_p L^p$, with the backshift operator satisfying $L^k V_t = V_{t-k}$, $k > 0$. In the following, an algorithm to compute the estimates for $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ is presented.

**Algorithm**

(a1) Estimate the coefficient vector $\boldsymbol{\alpha}$ by ordinary least squares, and denote it by $\hat{\boldsymbol{\alpha}}$;

(a2) Fit an autoregressive model to the residuals of step (a1), i.e., $\hat{\epsilon}_t = Y_t - \mathbb{v}_t^\top \hat{\boldsymbol{\alpha}}$, say,

$$\hat{\theta}_p(L)\hat{\epsilon}_t = \varepsilon_t;$$

(a3) Estimate $\boldsymbol{\eta}$ numerically, minimizing (8), using the estimates in steps (a1) and (a2) as initial values.

# Regularized wavelet-based estimators

Since in a MRA the space $V_J$, for a specific resolution level $J$, can be written as $V_{j_0} \oplus W_{j_0} \oplus \ldots \oplus W_{J-1}$, the orthogonal projection $g^J$ of $g$, previously defined in (2), can be analyzed by father and mother wavelets as

$$g^J(y) = \sum_{k=0}^{2^{j_0}-1} \alpha_{j_0 k} \varphi_{j_0 k}(y) + \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} \beta_{jk} \psi_{jk}(y),$$

where the $\beta_{jk}$'s correspond to the detail coefficients at level $j$ and $j_0$ to the coarsest level. In this case, $J-1$ is usually known as the finest level. The detail coefficients at coarser levels (closer to $j_0$) tend to capture global features of $g^J$, while those at finer levels (closer to $J-1$) are more responsible for local characteristics of the function. This is an advantage because one can reduce the noise in the coefficient function estimates by shrinking or thresholding detail coefficient estimates. These detail coefficients $\bar{\beta}_{jk}$'s can be easily obtained by applying discrete wavelet transform (DWT) to the coefficient estimates $\bar{\alpha}_{ik}$, $k = 0, \ldots, 2^{J_i} - 1$ (see Mallat, 2008, for more details). Once the detail coefficients are shrunk/thresholded, one can apply the inverse DWT and obtain the wavelet coefficient estimates back, say $\bar{\alpha}^h_{Jk}$ (the superscript "h" would indicate that it was applied the hard threshold to the detail coefficients in the wavelet domain).

# Regularized wavelet-based estimators

Basically, the hard threshold method corresponds to the function $\eta_\lambda(x) = \mathbb{1}(x > \lambda)x$, where $\mathbb{1}(A)$ is equal to one, if $A$ occurs, or zero otherwise. The value of $\lambda$ corresponds to the threshold and it can be calculated under different approaches. In this work we consider the universal threshold of Donoho and Johnstone (1994), which in our case will be $\lambda = \sigma\sqrt{2(J-1)\log 2}$. The value of $\sigma$ (standard deviation of the noise) is unknown in practice and is usually estimated by the $\boxed{\text{MAD}}$ (median of absolute deviation), of the wavelet coefficients from the finest level of detail. In other words, the MAD estimator of a vector $x$ is defined by $\boxed{\text{MAD}(x) = 1.4826 \cdot \text{median}(|x - \text{median}(x)|)}$, where 1.4826 is a scale factor useful to ensure consistency for when the in the case of normality.

The regularization is based on the wavelet coefficient estimates of individual coefficient function estimates $\bar{f}_i$. In order to make it clearer, the steps of regularization are summarized below.

# Steps to obtain the regularized coefficient functions

(s1) Apply the DWT to $\bar{\alpha}_{i0}, \ldots, \bar{\alpha}_{i,2^{J_i}-1}$ and obtain $\{\{\bar{\alpha}_{j_0 r}\}, \{\bar{\beta}_{jk}\},\ r = 0, \ldots, 2^{j_0}-1,\ k = 0, \ldots, 2^j - 1,\ j = j_0, \ldots, J_i - 1\}$;

(s2) Calculate $\hat{\sigma} = \mathrm{MAD}(\bar{\beta}_{J_i-1})$, where $\bar{\beta}_{J_i-1} = (\bar{\beta}_{J_i-1,0}, \ldots, \bar{\beta}_{J_i-1,2^{J_i-1}-1})^\top$, and set $\hat{\lambda} = \hat{\sigma}\sqrt{2(J_i-1)\log 2}$;

(s3) Apply the hard thresholding method to each detail coefficient defining $\bar{\beta}^h_{jk} = \eta_{\hat{\lambda}}(\bar{\beta}_{jk})$, $k = 0, \ldots, 2^j - 1$, $j = j_0, \ldots, J_i - 1$;

(s4) Apply the inverse DWT to the thresholded coefficients and obtain $\bar{\alpha}^h_{i0}, \ldots, \bar{\alpha}^h_{i,2^{J_i}-1}$. The regularized coefficient function estimate will be

$$\bar{f}^h_i(u) = \sum_{k=0}^{2^{J_i}-1} \bar{\alpha}^h_{ik}\phi_{ik}(H(u)).$$

The steps above must be applied to each coefficient function estimate, as stated before, and the value of the coarsest level $j_0$ does not need to be the same in each case. Similarly, one could regularize the coefficient function estimates $\hat{f}_i$, $i = 1, \ldots, d$, presented in Section 2.1. In this case, in step (s1) above, the DWT would be applied to $\hat{\alpha}_{i0}, \ldots, \hat{\alpha}_{i,2^{J_i}-1}$. The resulting regularized coefficient functions estimated could be denoted by $\hat{f}^h_i(u)$, $i = 1, \ldots, d$.

# Theoretical results

**Theorem 2.3** *If the assumptions (W0) – (W6) hold, then*

$$\sum_{i=1}^{d} \mathbb{E}\|\hat{f}_i^h - f_i\|^2_{L_2(h)} \leq C \sum_{i=1}^{d} \left( \frac{2^{J_i}}{n} + \rho_i^2 \right),$$

*for some $C > 0$. In particular, if $\rho_i = o(1)$, then $\mathbb{E}\|\hat{f}_i^h - f_i\|^2_{L_2(h)} = o(1)$, $i = 1, \ldots, d$.*

**Theorem 2.4** *If assumptions (W0) – (W6) hold and $\hat{\Sigma}$ is consistent in probability estimating $\Sigma$, then*

$$\sum_{i=1}^{d} \|\bar{f}_i^h - f_i\|^2_{L_2(h)} = O_p\left( \sum_{i=1}^{d} \left( \frac{2^{J_i}}{n} + \rho_i^2 \right) \right).$$

*In particular, if $\rho_i = o(1)$, then $\bar{f}_i^h$ is consistent in probability in estimating $f_i$, i.e., $\|\bar{f}_i^h - f_i\|_{L_2(h)} = o_p(1)$, $i = 1, \ldots, d$.*

# Selecting the resolution level and variables

In practical situations we do not know which value of $J_i$ should be chosen.

The same problem happens to the choice of the coarsest level $j_0$ used during the regularization.

An alternative is to use an automatic method to select the more appropriate values for the finest and coarsest levels ($J_i$ - 1 and $j_0$, respectively). These values can be chosen based on some criterion function.

In this work we evaluate three different criteria, namely: AIC Akaike (1974), AICc Hurvich and Tsai (1989) and BIC Schwarz (1978).

# Selecting the resolution level and variables

Denote the sample size by $n$ and the residual mean square by RMS, which will correspond to $\ell(\bar{\eta})/n$, where $\ell$ is presented in equation (8) and $\bar{\eta}$ is obtained in step (a3) of the algorithm of estimation. Furthermore, for specific values of finest and coarsest levels, let us denote

$$\begin{aligned} p = &(\text{number of autoregressive coefficients assumed for the errors}) \\ &+ (\text{number of wavelet coefficients}) \\ &- (\text{number of detail coefficients zeroed during the regularization process}). \end{aligned}$$

The criteria functions are then defined as

$$\begin{aligned} \text{AIC} &= \log(\text{RMS}) + \frac{2p}{n}, \\ \text{AICc} &= \text{AIC} + \frac{2(p+1)(p+2)}{n(n-p-2)}, \\ \text{BIC} &= \log(\text{RMS}) + \frac{p}{2}\log(n). \end{aligned}$$

# Selecting the resolution level and variables

Moreover, the criteria above can be used in the stepwise method for selecting the variables. This is important specially when there is no knowledge on the physical background of the data. Doing similarly to Huang and Shen (2004), in the case of the FAR model (the procedure is straightforward for situations with exogenous variables), fixed a constant $q > 0$, it is possible to select a threshold lag $1 \leq r \leq q$ and a set of significant lags $S_r \subseteq \{1, \ldots, q\}$, which compose the class of candidate models

$$Y_t = \sum_{i \in S_r} f_i(Y_{t-r}) Y_{t-i} + \epsilon.$$

Fixing a threshold lag $r$, it is possible to choose an optimal set of significant lags $S_r^*$ by adding and removing variables. During the addition stage, we can choose one significant lag at a time among the candidate lags, selecting the one that was not selected yet and that minimizes the RMS. This procedure stops when the specified number of significant lags $q$ is reached. The deletion stage is similar to the addition stage, where we start selecting the set with maximum number of significant lags and then we remove one at a time, choosing the one that minimizes the RMS. Then we can choose the optimum set of significant lags $S_r^*$ as the one which minimizes a criterion function (e.g. AIC) among the sequence of subsets of lags obtained during the addition and deletion stages. The model to be chosen corresponds to the one in which the pair $\{r, S_r^*\}$ is minimum, for $1 \leq r \leq q$.

# Simulation study

In order to evaluate how close is an estimate to the real function, we use an approximation of the mean integrated squared error. This approximation is based on the average squared error related to each coefficient function ($\mathrm{ASE}_i$), and it is defined as

$$\mathrm{ASE}^2 = \sum_{i=1}^{d} \mathrm{ASE}_i^2, \quad \text{with}$$

$$\mathrm{ASE}_i = \left\{ n_{\text{grid}}^{-1} \sum_{k=1}^{n_{\text{grid}}} \left[ \bar{f}_i(u_k) - f_i(u_k) \right]^2 \right\}^{1/2},$$

where $\{u_k, k = 1, \ldots, n_{\text{grid}}\}$ is a grid of points equally spaced in an interval that belongs to the range of the data set. Following Huang and Shen (2004), we selected the maximum of the 2.5 percentiles of the simulated data sets as the left boundary ($u_1$) and the minimum of the 97.5 percentiles of the data sets as the right boundary ($u_{n_{\text{grid}}}$). We consider $n_{\text{grid}} = 250$.

Thus, small values of the $\mathrm{ASE}^2$ indicate a good performance of the estimates. These values can be summarized by location/dispersion measures. Although it is not presented here, a few outliers were observed from the $\mathrm{ASE}^2$'s obtained from the simulations. For this reason, we consider robust measures such as the median and the MAD for location and dispersion measures, respectively.

# Simulation study

- 10,000 samples of size 400 are simulated;
- autoregressive errors generated with order 1, 2 and 3, with AR coefficients and standard deviation of white noise presented in Table 1;
- the white noises are iid and normally distributed;
- the coefficient functions are estimated using different wavelet bases, which are *Daublets* D8, D12, D16, D20, and *Symmlets* S8, S12, S16, S20, where D$N$ corresponds to the $N$-tap of Daubechies' Extremal Phase wavelet filter, and S$N$ represents the $N$-tap Daubechies' Least-Asymmetric wavelet filter;
- FAR models with order 2 are studied (with two coefficient functions). For the sake of simplicity, we use the same coarsest and finest level ($j_0$ and $J-1$, respectively) and the same wavelet basis during the estimation process of both coefficient functions, $f_1$ and $f_2$.

**Table 1:** *Parameters of the autoregressive errors used in the simulation studies.*

| $p$ | $\phi_p(L)$ | $\sigma_\varepsilon$ |
|---|---|---|
| 1 | $1-0.8L$ | 0.1200 |
| 2 | $1-0.8L+0.7L^2$ | 0.1260 |
| 3 | $1-0.6L+0.7L^2-0.6L^3$ | 0.1348 |

# **Simulation study** - First example

The first simulated model corresponds to the EXPAR model (Haggan and Ozaki, 1981; Cai et al., 2000; Huang and Shen, 2004; Montoril et al., 2014)

$$Y_t = f_1(Y_{t-1})Y_{t-1} + f_2(Y_{t-1})Y_{t-2} + \epsilon_t,$$

where $f_1(u) = 0.138 + (0.316 + 0.982u)e^{-3.89u^2}$, $f_2(u) = -0.437 + (0.659 + 1.260u)e^{-3.89u^2}$.

**Table 2:** *Median (MAD) of the $ASE^2$'s for the Classical Wavelet (CW) and Warped Wavelet (WW) according to several different wavelet bases. These values were obtained according to different criteria functions (AIC, AICc and BIC) for selecting the coarsest and finest levels among $0 \leq j_0 \leq J$ and $J \in \{2, 3, 4\}$. The hard threshold method was used to regularize the estimates. Three different AR errors were used in the simulations of the data sets, with parameters presented in Table 1.*

| AR order | Basis | AIC | | AICc | | BIC | |
|---|---|---|---|---|---|---|---|
| | | CW | WW | CW | WW | CW | WW |
| AR(1) | D8 | 0.004 (0.003) | 0.011 (0.010) | 0.004 (0.003) | 0.010 (0.009) | 0.004 (0.003) | 0.006 (0.005) |
| | D12 | 0.004 (0.003) | 0.027 (0.024) | 0.004 (0.003) | 0.025 (0.021) | 0.006 (0.004) | 0.018 (0.011) |
| | D16 | 0.004 (0.003) | 0.014 (0.012) | 0.003 (0.003) | 0.013 (0.011) | 0.004 (0.003) | 0.008 (0.006) |
| | D20 | 0.004 (0.002) | 0.023 (0.019) | 0.004 (0.002) | 0.022 (0.017) | 0.005 (0.004) | 0.015 (0.008) |
| | S8 | 0.005 (0.003) | 0.030 (0.028) | 0.005 (0.003) | 0.028 (0.025) | 0.007 (0.005) | 0.018 (0.012) |
| | S12 | 0.004 (0.003) | 0.025 (0.024) | 0.004 (0.003) | 0.024 (0.022) | 0.007 (0.005) | 0.014 (0.008) |
| | S16 | 0.004 (0.003) | 0.021 (0.020) | 0.004 (0.003) | 0.020 (0.018) | 0.006 (0.004) | 0.011 (0.007) |
| | S20 | 0.004 (0.003) | 0.021 (0.018) | 0.004 (0.003) | 0.019 (0.016) | 0.006 (0.004) | 0.012 (0.007) |

**Table 2:** *Median (MAD) of the ASE$^2$'s for the Classical Wavelet (CW) and Warped Wavelet (WW) according to several different wavelet bases. These values were obtained according to different criteria functions (AIC, AICc and BIC) for selecting the coarsest and finest levels among $0 \leq j_0 \leq J$ and $J \in \{2, 3, 4\}$. The hard threshold method was used to regularize the estimates. Three different AR errors were used in the simulations of the data sets, with parameters presented in Table 1.*

| AR order | Basis | AIC | | AICc | | BIC | |
|---|---|---|---|---|---|---|---|
| | | CW | WW | CW | WW | CW | WW |
| AR(2) | D8 | 0.005 (0.003) | 0.007 (0.005) | 0.005 (0.003) | 0.006 (0.004) | 0.006 (0.003) | 0.004 (0.002) |
| | D12 | 0.004 (0.002) | 0.007 (0.006) | 0.003 (0.002) | 0.006 (0.005) | 0.004 (0.002) | 0.004 (0.003) |
| | D16 | 0.004 (0.002) | 0.004 (0.003) | 0.003 (0.002) | 0.004 (0.003) | 0.004 (0.002) | 0.003 (0.002) |
| | D20 | 0.003 (0.002) | 0.006 (0.005) | 0.003 (0.002) | 0.005 (0.004) | 0.004 (0.002) | 0.004 (0.002) |
| | S8 | 0.005 (0.003) | 0.009 (0.007) | 0.005 (0.002) | 0.009 (0.007) | 0.004 (0.002) | 0.006 (0.003) |
| | S12 | 0.004 (0.002) | 0.006 (0.005) | 0.004 (0.002) | 0.006 (0.004) | 0.004 (0.002) | 0.004 (0.003) |
| | S16 | 0.004 (0.002) | 0.005 (0.004) | 0.003 (0.002) | 0.004 (0.003) | 0.004 (0.002) | 0.003 (0.002) |
| | S20 | 0.003 (0.002) | 0.005 (0.004) | 0.003 (0.002) | 0.005 (0.004) | 0.003 (0.002) | 0.003 (0.002) |

**Table 2:** *Median (MAD) of the ASE$^2$'s for the Classical Wavelet (CW) and Warped Wavelet (WW) according to several different wavelet bases. These values were obtained according to different criteria functions (AIC, AICc and BIC) for selecting the coarsest and finest levels among $0 \leq j_0 \leq J$ and $J \in \{2, 3, 4\}$. The hard threshold method was used to regularize the estimates. Three different AR errors were used in the simulations of the data sets, with parameters presented in Table 1.*

| AR order | Basis | AIC | | AICc | | BIC | |
|---|---|---|---|---|---|---|---|
| | | CW | WW | CW | WW | CW | WW |
| AR(3) | D8 | 0.008 (0.004) | 0.010 (0.007) | 0.008 (0.004) | 0.010 (0.007) | 0.009 (0.006) | 0.010 (0.006) |
| | D12 | 0.007 (0.004) | 0.010 (0.008) | 0.007 (0.004) | 0.009 (0.006) | 0.008 (0.005) | 0.006 (0.004) |
| | D16 | 0.007 (0.004) | 0.009 (0.007) | 0.007 (0.004) | 0.009 (0.006) | 0.009 (0.005) | 0.009 (0.004) |
| | D20 | 0.006 (0.004) | 0.009 (0.006) | 0.006 (0.004) | 0.008 (0.006) | 0.008 (0.005) | 0.007 (0.004) |
| | S8 | 0.008 (0.004) | 0.013 (0.011) | 0.008 (0.004) | 0.012 (0.009) | 0.009 (0.005) | 0.008 (0.004) |
| | S12 | 0.007 (0.004) | 0.011 (0.009) | 0.007 (0.004) | 0.010 (0.007) | 0.008 (0.005) | 0.008 (0.004) |
| | S16 | 0.007 (0.004) | 0.011 (0.008) | 0.007 (0.004) | 0.010 (0.007) | 0.009 (0.005) | 0.008 (0.005) |
| | S20 | 0.007 (0.004) | 0.010 (0.008) | 0.007 (0.004) | 0.009 (0.007) | 0.008 (0.005) | 0.008 (0.005) |

**Table 3:** *Median (MAD) of the ASE$^2$'s for the (quadratic) spline-based (Spl) estimates. These values were obtained according to different criteria functions (AIC, AICc and BIC) for selecting the number of (equally spaced) knots among $\{2, 3, 4, 5\}$. Three different AR errors were used in the simulations of the data sets, with parameters presented in Table 1. For the sake of comparisons, ASE$^2$ estimates of the Classical case (CW) and Warped case (WW) are presented, for the estimates based on the Daubechies D16.*

| AR | AIC | | | AICc | | | BIC | | |
|---|---|---|---|---|---|---|---|---|---|
| order | Spl. | CW | WW | Spl. | CW | WW | Spl. | CW | WW |
| AR(1) | 0.004 | 0.004 | 0.014 | 0.004 | 0.003 | 0.013 | 0.005 | 0.004 | 0.008 |
| | (0.002) | (0.003) | (0.012) | (0.002) | (0.003) | (0.011) | (0.003) | (0.003) | (0.006) |
| AR(2) | 0.003 | 0.004 | 0.004 | 0.003 | 0.003 | 0.004 | 0.003 | 0.004 | 0.003 |
| | (0.002) | (0.002) | (0.003) | (0.002) | (0.002) | (0.003) | (0.002) | (0.002) | (0.002) |
| AR(3) | 0.022 | 0.007 | 0.009 | 0.022 | 0.007 | 0.009 | 0.022 | 0.009 | 0.009 |
| | (0.008) | (0.004) | (0.007) | (0.008) | (0.004) | (0.006) | (0.009) | (0.005) | (0.004) |

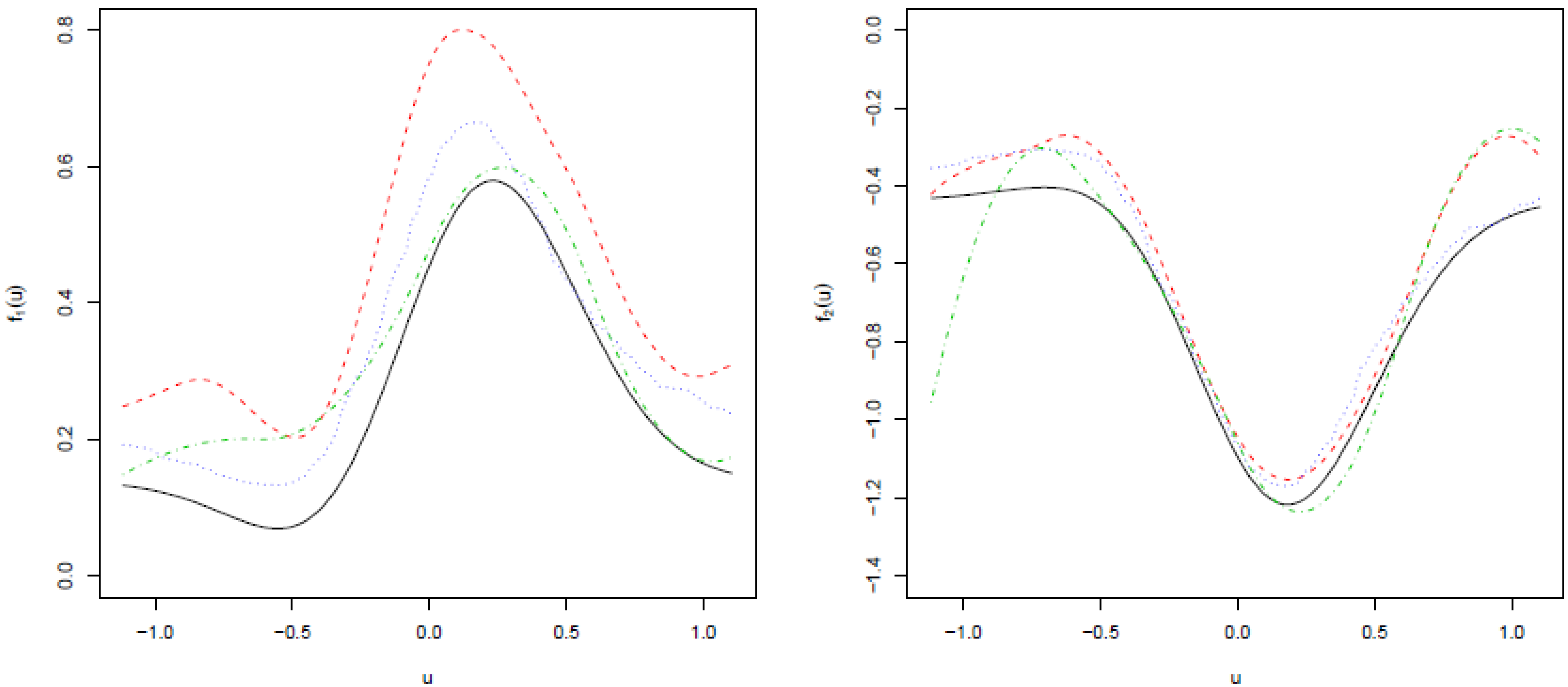


**Figure 1:** *Classical wavelet-based (dashed lines), Warped wavelet-based (dotted lines) and Spline (dot-dashed lines) estimates of the coefficient functions $f_1$ (left side) and $f_2$ (right side) of the EXPAR model generated. The real coefficient functions correspond to the solid lines. The wavelet estimates are based on the 16-tap Daubechies wavelet filter (Daublet D16), and the spline estimates based on the quadratic splines.*

# **Simulation study** - Second example

In this simulation study, we consider the same structure of the EXPAR model in the first simulation example. However, the coefficient functions are a little more irregular. Denoting by $d(x;\mu;\sigma)$ the probability density function of a normal distribution with mean $\mu$ and standard deviation $\sigma$ at the point $x$, the coefficient functions used in this example are

$$f_1(u) = 0.8 + 0.5d(u;0;0.03)$$
$$f_2(u) = -0.4 - 0.25(d(u;-0.2;0.05) + d(u;0.2,.05)).$$

**Table 4:** *Median (MAD) of the $ASE^2$'s for the Classical Wavelet (CW) and Warped Wavelet (WW) according t several different wavelet bases. These values were obtained according to different criteria functions (AIC, AIC and BIC) for selecting the coarsest and finest levels among $0 \leq j_0 \leq J$ and $J \in \{4, 5, 6\}$. The hard threshol method was used to regularize the estimates. Three different AR errors were used in the simulations of the dat sets, with parameters presented in Table 1.*

| AR order | Basis | AIC | | AICc | | BIC | |
|---|---|---|---|---|---|---|---|
| | | CW | WW | CW | WW | CW | WW |
| | D8 | 0.890 | 0.267 | 0.905 | 0.270 | 1.073 | 0.324 |
| | | (0.566) | (0.189) | (0.559) | (0.197) | (0.713) | (0.264) |
| | D12 | 0.828 | 0.263 | 0.834 | 0.289 | 0.907 | 0.452 |
| | | (0.390) | (0.178) | (0.384) | (0.213) | (0.467) | (0.359) |
| | D16 | 0.803 | 0.248 | 0.806 | 0.277 | 0.868 | 0.581 |
| | | (0.291) | (0.185) | (0.284) | (0.232) | (0.307) | (0.433) |
| | D20 | 0.808 | 0.255 | 0.819 | 0.321 | 0.848 | 0.638 |
| | | (0.284) | (0.189) | (0.268) | (0.256) | (0.301) | (0.389) |
| AR(1) | S8 | 0.785 | 0.281 | 0.805 | 0.289 | 0.944 | 0.324 |
| | | (0.521) | (0.193) | (0.511) | (0.198) | (0.576) | (0.236) |
| | S12 | 0.748 | 0.268 | 0.758 | 0.285 | 0.832 | 0.344 |
| | | (0.355) | (0.197) | (0.340) | (0.215) | (0.373) | (0.273) |
| | S16 | 0.735 | 0.257 | 0.747 | 0.269 | 0.794 | 0.392 |
| | | (0.286) | (0.199) | (0.275) | (0.212) | (0.301) | (0.335) |
| | S20 | 0.734 | 0.248 | 0.740 | 0.270 | 0.785 | 0.455 |
| | | (0.260) | (0.185) | (0.248) | (0.210) | (0.260) | (0.391) |

**Table 4:** *Median (MAD) of the ASE$^2$'s for the Classical Wavelet (CW) and Warped Wavelet (WW) according to several different wavelet bases. These values were obtained according to different criteria functions (AIC, AICc and BIC) for selecting the coarsest and finest levels among $0 \leq j_0 \leq J$ and $J \in \{4, 5, 6\}$. The hard threshold method was used to regularize the estimates. Three different AR errors were used in the simulations of the data sets, with parameters presented in Table 1.*

| AR order | Basis | AIC | | AICc | | BIC | |
|---|---|---|---|---|---|---|---|
| | | CW | WW | CW | WW | CW | WW |
| AR(2) | D8 | 1.415 (0.613) | 0.200 (0.153) | 1.409 (0.594) | 0.168 (0.119) | 1.431 (0.532) | 0.180 (0.136) |
| | D12 | 1.337 (0.582) | 0.199 (0.153) | 1.343 (0.568) | 0.157 (0.114) | 1.350 (0.515) | 0.202 (0.165) |
| | D16 | 1.277 (0.591) | 0.181 (0.147) | 1.284 (0.576) | 0.146 (0.113) | 1.306 (0.524) | 0.213 (0.188) |
| | D20 | 1.285 (0.619) | 0.186 (0.143) | 1.287 (0.603) | 0.144 (0.102) | 1.276 (0.516) | 0.209 (0.174) |
| | S8 | 1.356 (0.632) | 0.209 (0.163) | 1.355 (0.619) | 0.165 (0.119) | 1.383 (0.557) | 0.178 (0.132) |
| | S12 | 1.277 (0.610) | 0.182 (0.141) | 1.287 (0.592) | 0.152 (0.107) | 1.300 (0.544) | 0.170 (0.121) |
| | S16 | 1.269 (0.624) | 0.175 (0.136) | 1.254 (0.608) | 0.145 (0.105) | 1.271 (0.527) | 0.180 (0.146) |
| | S20 | 1.269 (0.632) | 0.169 (0.129) | 1.268 (0.609) | 0.143 (0.102) | 1.272 (0.506) | 0.182 (0.151) |

**Table 4:** *Median (MAD) of the ASE$^2$'s for the Classical Wavelet (CW) and Warped Wavelet (WW) according to several different wavelet bases. These values were obtained according to different criteria functions (AIC, AICc and BIC) for selecting the coarsest and finest levels among $0 \leq j_0 \leq J$ and $J \in \{4, 5, 6\}$. The hard threshold method was used to regularize the estimates. Three different AR errors were used in the simulations of the data sets, with parameters presented in Table 1.*

| AR order | Basis | AIC | | AICc | | BIC | |
|---|---|---|---|---|---|---|---|
| | | CW | WW | CW | WW | CW | WW |
| AR(3) | D8 | 1.289 | 0.238 | 1.301 | 0.207 | 1.547 | 0.280 |
| | | (1.159) | (0.187) | (1.168) | (0.144) | (1.198) | (0.224) |
| | D12 | 0.999 | 0.219 | 1.008 | 0.209 | 1.109 | 0.349 |
| | | (0.548) | (0.176) | (0.548) | (0.166) | (0.685) | (0.301) |
| | D16 | 0.893 | 0.219 | 0.892 | 0.211 | 0.914 | 0.449 |
| | | (0.369) | (0.179) | (0.363) | (0.167) | (0.383) | (0.330) |
| | D20 | 0.899 | 0.213 | 0.898 | 0.220 | 0.922 | 0.473 |
| | | (0.364) | (0.173) | (0.351) | (0.176) | (0.373) | (0.315) |
| | S8 | 1.092 | 0.236 | 1.094 | 0.214 | 1.226 | 0.257 |
| | | (0.870) | (0.180) | (0.857) | (0.151) | (1.007) | (0.200) |
| | S12 | 0.892 | 0.225 | 0.894 | 0.208 | 0.919 | 0.300 |
| | | (0.452) | (0.177) | (0.454) | (0.152) | (0.467) | (0.241) |
| | S16 | 0.843 | 0.223 | 0.841 | 0.211 | 0.853 | 0.329 |
| | | (0.356) | (0.175) | (0.350) | (0.163) | (0.356) | (0.270) |
| | S20 | 0.829 | 0.221 | 0.832 | 0.209 | 0.839 | 0.370 |
| | | (0.315) | (0.178) | (0.303) | (0.161) | (0.313) | (0.295) |

**Table 5:** *Median (MAD) of the $ASE^2$'s for the (quadratic) spline-based (Spl) estimates. These values were obtained according to different criteria functions (AIC, AICc and BIC) for selecting the number of (equally spaced) knots among $\{2, \ldots, 30\}$. Three different AR errors were used in the simulations of the data sets, with parameters presented in Table 1. For the sake of comparisons, $ASE^2$ estimates of the Classical case (CW) and Warped case (WW) are presented, for the estimates based on the Daubechies D16.*

| AR | AIC | | | AICc | | | BIC | | |
|---|---|---|---|---|---|---|---|---|---|
| order | Spl. | CW | WW | Spl. | CW | WW | Spl. | CW | WW |
| AR(1) | 1.042 | 0.803 | 0.248 | 1.049 | 0.806 | 0.277 | 1.112 | 0.868 | 0.581 |
| | (0.268) | (0.291) | (0.185) | (0.266) | (0.284) | (0.232) | (0.266) | (0.307) | (0.433) |
| AR(2) | 0.880 | 1.277 | 0.181 | 0.885 | 1.284 | 0.146 | 0.909 | 1.306 | 0.213 |
| | (0.199) | (0.591) | (0.147) | (0.199) | (0.576) | (0.113) | (0.204) | (0.524) | (0.188) |
| AR(3) | 1.031 | 0.893 | 0.219 | 1.041 | 0.892 | 0.211 | 1.084 | 0.914 | 0.449 |
| | (0.195) | (0.369) | (0.179) | (0.196) | (0.363) | (0.167) | (0.217) | (0.383) | (0.330) |

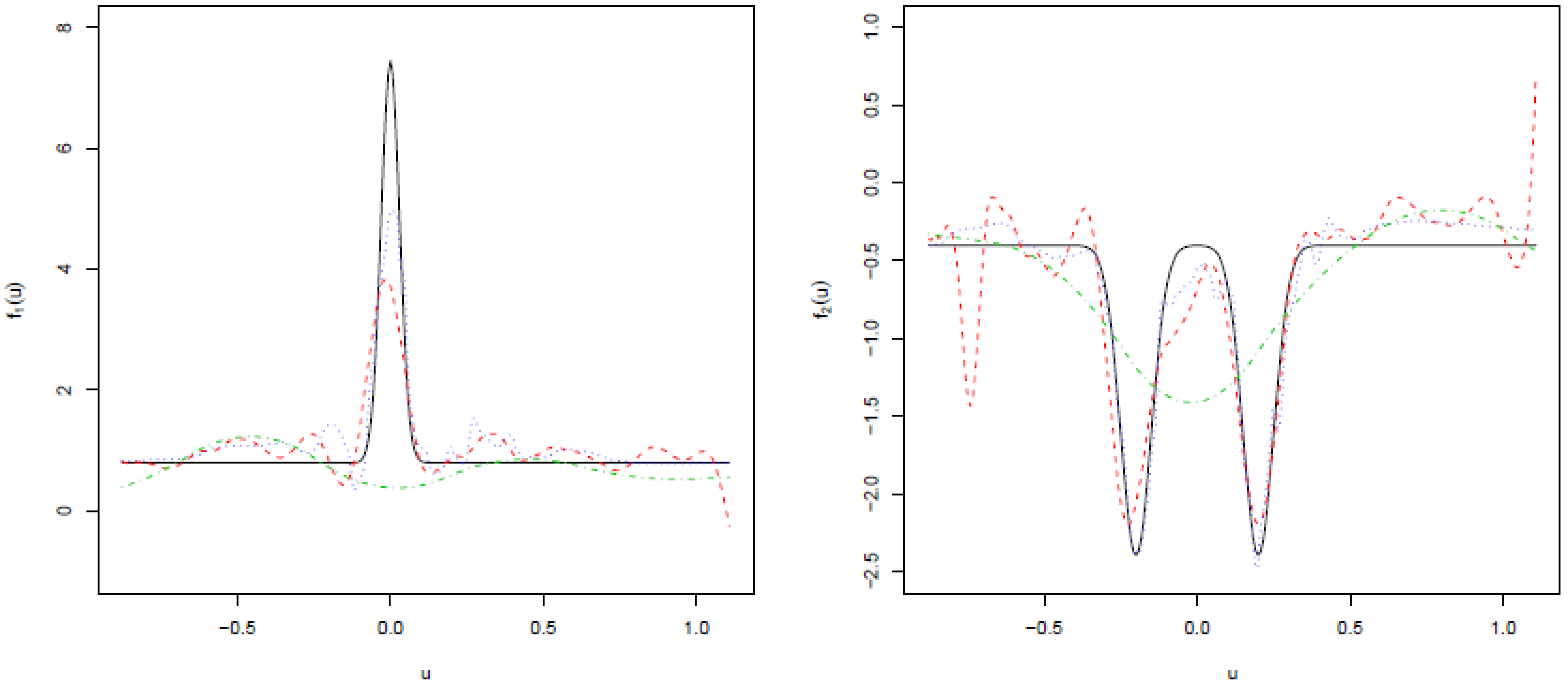


**Figure 2:** *Classical wavelet-based (dashed lines), Warped wavelet-based (dotted lines) and Spline (dot-dashed lines) estimates of the coefficient functions $f_1$ (left side) and $f_2$ (right side) of the data generated in according to the simulation of the second example. The real coefficient functions correspond to the solid lines. The wavelet estimates are based on the 16-tap Daubechies wavelet filter (Daublet D16), and the spline estimates based on the quadratic splines.*

# Application to Industrial Production Index

$$Y_t = 100 \times \log(X_t / X_{t-1})$$

The monthly Seasonally Adjusted Industrial Production index (IPI) of the USA, from December 1980 to December 2007, with 325 observations.

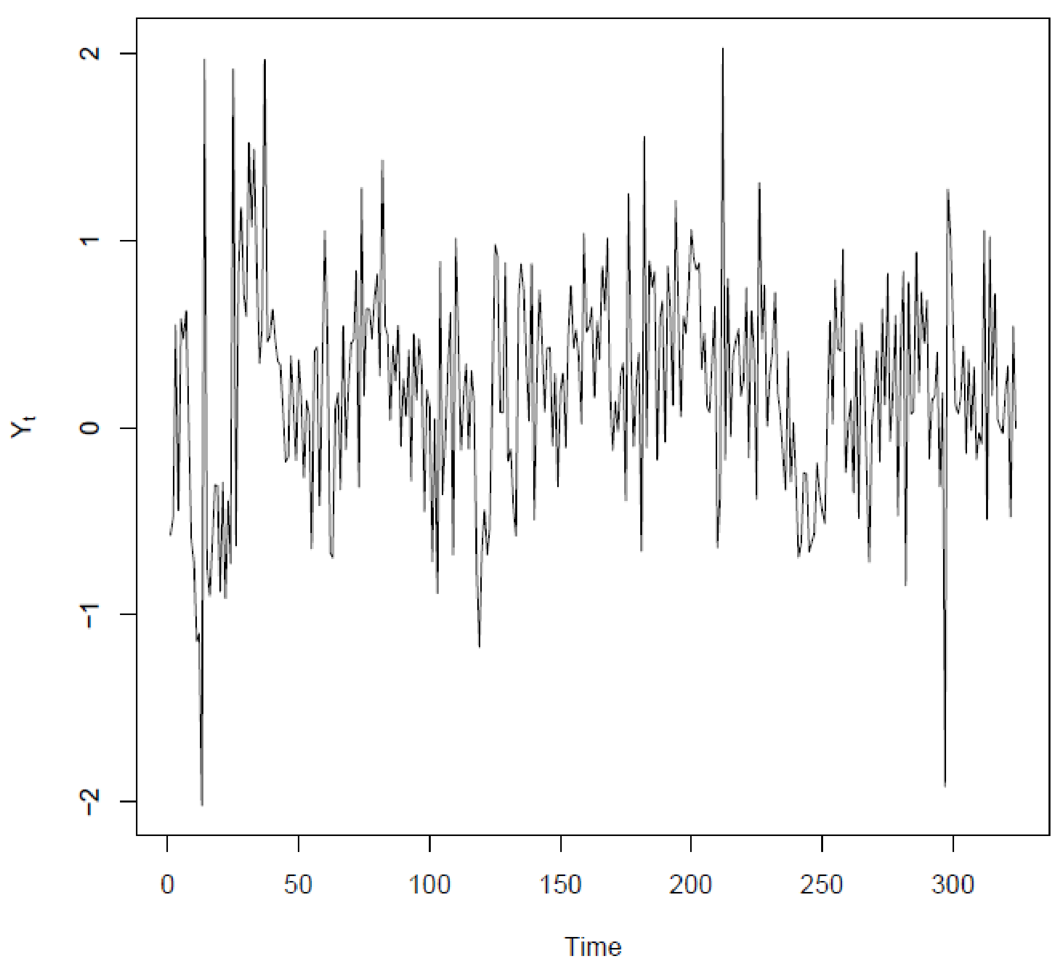


**Figure 3:** *Log-return of the monthly seasonally adjusted industrial production index, from January 1981 to December 2007.*

**Table 6:** *Classical wavelet-based model selections for the industrial production index according to the coarsest ($j_0$) and finest ($J-1$) levels using the 16-tap Daubechies wavelet filter (Daublet D16). AICc$^{(1)}$ corresponds to the criterion function value in the first stage of selection that provided the smallest AICc, $r$ and $S_r^*$ are the resulting threshold and the significant lags, respectively. The two last columns correspond to the autoregressive order suggested to the errors after a residual analysis and the updated value of AICc in the final model.*

| $j_0$ | $J$ | AICc$^{(1)}$ | $r$ | $S_r^*$ | $p$ | AICc$^{(2)}$ |
|---|---|---|---|---|---|---|
| 0 | 2 | -1.210 | 1 | $\{1,2,3\}$ | 1 | -1.219 |
| 1 | 2 | -1.215 | 1 | $\{1,2,3\}$ | 1 | -1.232 |
| 2 | 2 | -1.215 | 3 | $\{2,3\}$ | 0 | -1.215 |
| 0 | 3 | -1.189 | 3 | $\{2,3\}$ | 0 | -1.189 |
| 1 | 3 | -1.189 | 3 | $\{2,3\}$ | 0 | -1.189 |
| 2 | 3 | -1.189 | 3 | $\{2,3\}$ | 0 | -1.189 |
| 3 | 3 | -1.243 | 1 | $\{1,2,3\}$ | 1 | -1.263 |
| 0 | 4 | -1.147 | 2 | $\{2\}$ | 3 | -1.171 |
| 1 | 4 | -1.146 | 2 | $\{2\}$ | 3 | -1.171 |
| 2 | 4 | -1.130 | 2 | $\{2\}$ | 3 | -1.169 |
| 3 | 4 | -1.130 | 2 | $\{2\}$ | 3 | -1.169 |
| 4 | 4 | -1.218 | 1 | $\{2,3\}$ | 7 | -1.273 |

The resulting model selected for the Classical wavelet-based approach is then

$$\begin{aligned} Y_t &= f_2(Y_{t-1})Y_{t-2} + f_3(Y_{t-1})Y_{t-3} + \epsilon_t \\ \epsilon_t &= \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \theta_3\epsilon_{t-3} + \theta_4\epsilon_{t-4} + \theta_5\epsilon_{t-5} + \theta_6\epsilon_{t-6} + \theta_7\epsilon_{t-7} + \varepsilon_t, \end{aligned} \tag{9}$$

where $\varepsilon_t$ is a white noise. The estimated coefficient functions are presented in Figure 4, and the autoregressive estimates are $\hat{\theta}_7(L) = 1 - 0.129L - 0.514L^2 + 0.145L^3 - 0.048L^4 - 0.215L^5 - 0.051L^6 + 0.124L^7$.

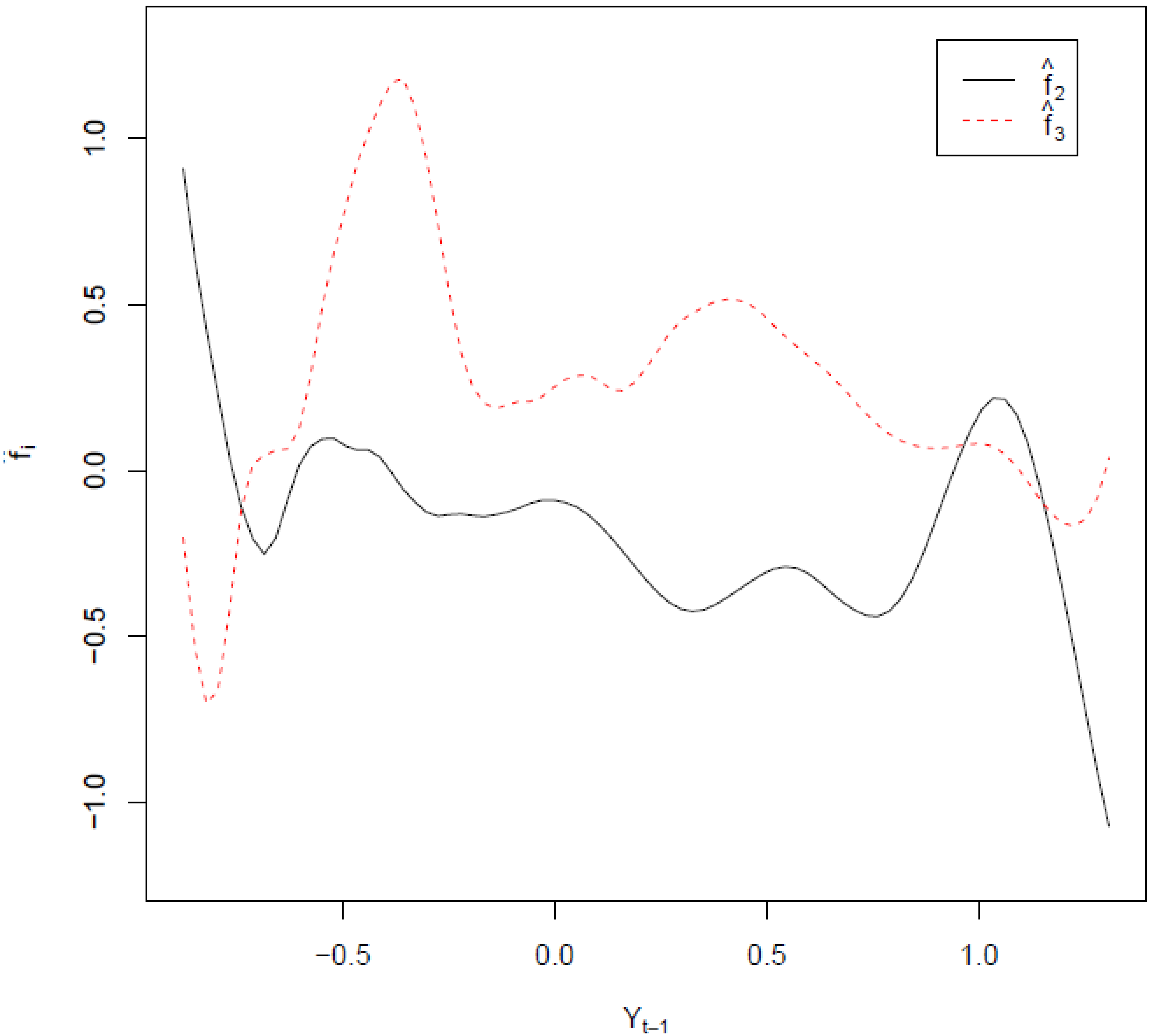


**Figure 4:** *Classical wavelet-based estimates of the coefficient functions of model (9) for the industrial production index series. The coefficient functions are estimated using the 16-tap Daubechies wavelet filter (Daublet D16).*

**Table 7:** *Warped wavelet-based model selections for the industrial production index according to the coarsest ($j_0$) and finest ($J-1$) levels using the 16-tap Daubechies wavelet filter (Daublet D16). AICc$^{(1)}$ corresponds to the criterion function value in the first stage of selection that provided the smallest AICc, $r$ and $S_r^*$ are the resulting threshold and the significant lags, respectively. The two last columns correspond to the autoregressive order suggested to the errors after a residual analysis and the updated value of AICc in the final model.*

| j0 | J | AICc | r | Sr* | $p$ | AICc |
|---|---|---|---|---|---|---|
| 0 | 2 | -1.201 | 3 | $\{2,3\}$ | 0 | -1.201 |
| 1 | 2 | -1.197 | 3 | $\{2,3\}$ | 0 | -1.197 |
| 2 | 2 | -1.213 | 3 | $\{2,3\}$ | 0 | -1.213 |
| 0 | 3 | -1.188 | 3 | $\{2,3\}$ | 0 | -1.188 |
| 1 | 3 | -1.189 | 3 | $\{2,3\}$ | 0 | -1.189 |
| 2 | 3 | -1.185 | 3 | $\{2,3\}$ | 0 | -1.185 |
| 3 | 3 | -1.177 | 3 | $\{2,3\}$ | 0 | -1.177 |
| 0 | 4 | -1.190 | 3 | $\{2,3\}$ | 0 | -1.190 |
| 1 | 4 | -1.201 | 3 | $\{2,3\}$ | 0 | -1.201 |
| 2 | 4 | -1.192 | 3 | $\{2,3\}$ | 0 | -1.192 |
| 3 | 4 | -1.177 | 3 | $\{2,3\}$ | 0 | -1.177 |
| 4 | 4 | -1.131 | 2 | $\{2\}$ | 3 | -1.146 |

$$Y_t = f_2(Y_{t-3})Y_{t-2} + f_3(Y_{t-3})Y_{t-3} + \epsilon_t, \tag{10}$$

where $\epsilon_t$ is white noise.

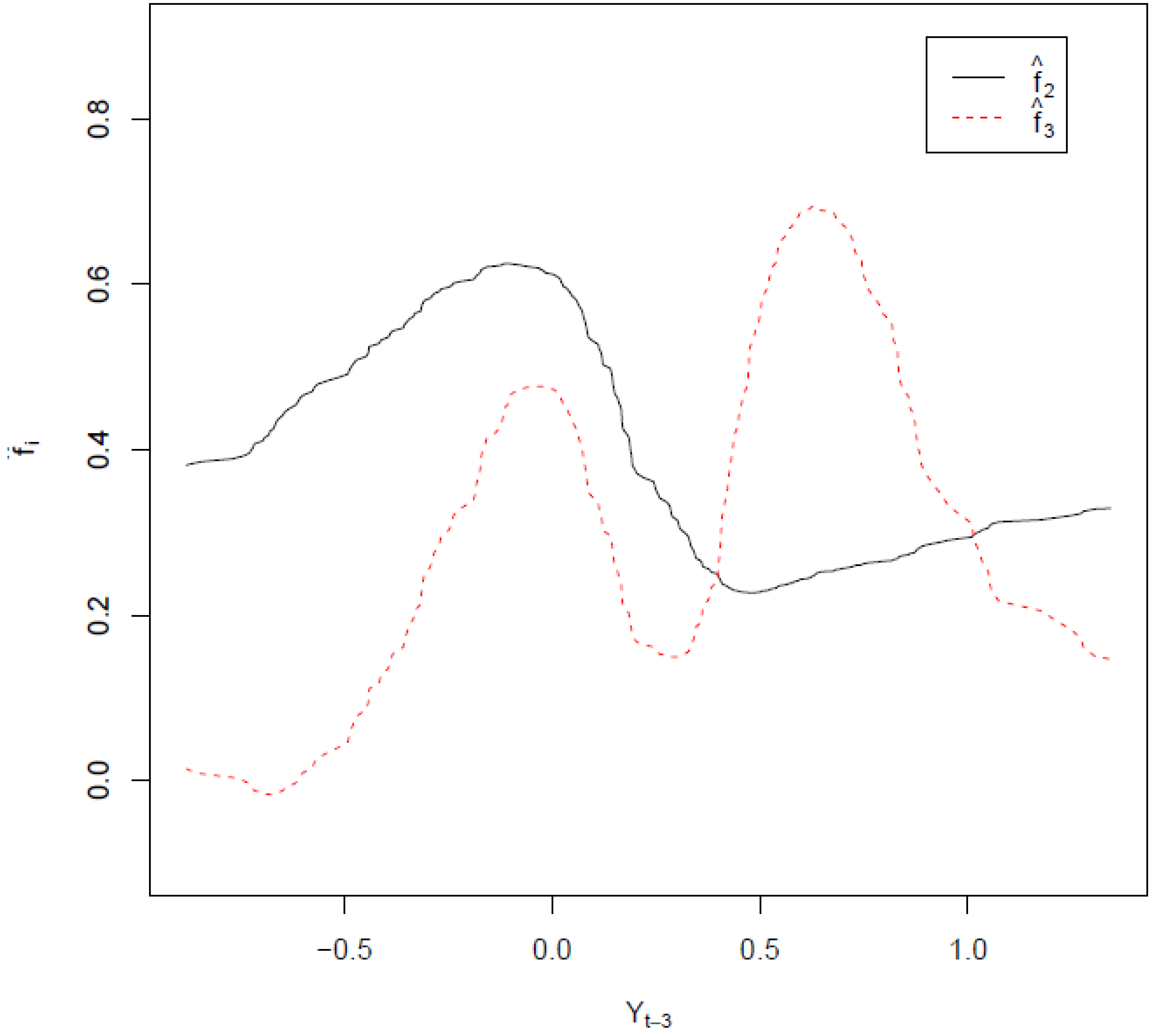


**Figure 5:** *Warped wavelet-based estimates of the coefficient functions of model* (10) *for the industrial production index series. the coefficient functions are estimated using the 16-tap Daubechies wavelet filter (Daublet D16).*

**Table 8:** *MSPEs of multi-step-ahead forecasts, based on 60 subseries of the monthly seasonally adjusted industrial production index. The second column corresponds to AR model results, the third column to the Classical wavelet-based (CW) results, the fourth column to the Warped wavelet-based (WW) forecasts and the fifth column to the (quadratic) spline-based results. The wavelet estimates are based on the 16-tap Daubechies wavelet filter (Daublet D16).*

| h | AR | CW | WW | Spline |
|---|---|---|---|---|
| 1 | 0.289 | 0.354 | 0.328 | 0.305 |
| 2 | 0.301 | 0.372 | 0.331 | 0.345 |
| 3 | 0.305 | 0.378 | 0.317 | 0.345 |
| 4 | 0.294 | 0.318 | 0.288 | 0.307 |
| 5 | 0.288 | 0.325 | 0.305 | 0.310 |
| 6 | 0.287 | 0.315 | 0.283 | 0.298 |
| 7 | 0.279 | 0.326 | 0.285 | 0.307 |
| 8 | 0.284 | 0.298 | 0.265 | 0.290 |
| 9 | 0.289 | 0.301 | 0.288 | 0.298 |
| 10 | 0.288 | 0.280 | 0.277 | 0.288 |
| 11 | 0.289 | 0.274 | 0.278 | 0.289 |
| 12 | 0.273 | 0.262 | 0.271 | 0.289 |

# References


✓Chen, R. and Tsay, R.S. (1993). Functional-coefficient autoregressive models. JASA, 88, 298-308.

✓Haggan, V. and Ozaki, T. (1981). Modelling nonlinear vibrations using an amplitude-dependent autoregressive time series model. Biometrika, 68, 189-196.

✓Tong, H. (1983). Threshold Models in Non-Linear Time Series Analysis. Lecture Notes in Statistics, 21. Heidelberg: Springer.

✓Cai, Z., Fan, J. and Yao, Q. (2000). Functional-coefficient regression models for non-linear time series. JASA, 95, 941-956.

✓Montoril, M. H., P. A. Morettin, and C. Chiann (2014). Spline estimation of functional coefficient regression models for time series with correlated errors. Statistics & Probability Letters 92, 226-231.

# References


- Daubechies, I. and J. C. Lagarias (1991). Two-scale difference equations I: existence and global regularity of solutions. SIAM Journal on Mathematical Analysis 22, 1388–1410.
- Daubechies, I. and J. C. Lagarias (1992). Two-scale difference equations II. local regularity, infinite products of matrices and fractals. SIAM Journal on Mathematical Analysis 23, 1031–1079
- Huang J. and Shen, H.(2004). Functional coefficient regression models for non-linear time series: A polynomial spline approach. Scandinavian Journal of Statistics, 31, 515-534.
- Kerkyacharian, G. and Picard, D. (2004). Regression in random design and warped wavelets. Bernoulli, 10, 1053-1105.

# References


- Vidakovic, B. (1999). Statistical Modeling by Wavelets. Wiley Series in Probability and Statistics. New York: Wiley-Interscience.