

Wavelet Estimation of Functional Coefficient Regression Models

Michel H. Montoril¹, Pedro A. Morettin², and Chang Chiann²

¹Department of Statistics, Institute of Exact Sciences , Federal University of Juiz de Fora,
Brazil

²Department of Statistics, Institute of Mathematics and Statistics, University of São Paulo,
Brazil

April 3, 2017

Abstract

The area of nonlinear time series models has experienced a great development since the 1980s. Although there is a wide range of parametric nonlinear time series models, in general we do not know if the postulated model is the most appropriated for a specific data set. This situation highlights the importance of nonparametric models. An interesting nonparametric model to fit nonlinear time series is the well-known functional coefficient regression model. Nonparametric estimations by, e.g., local linear regression and splines, are developed in the literature. In this work we study the estimation of such a model using wavelets. It is a proposal that takes into account both, classical and warped wavelets. We present the rates of convergence of the proposed estimators and carry out simulation studies to evaluate automatic procedures (among AIC, AICc and BIC) for selecting the coarsest and finest scales to be used during the estimation process. Moreover, we illustrate the methodology with an application to a real data set, where we also calculate multi-step-ahead forecasts and compare the results with other methods known in the literature.

Keywords: Daubechies-Lagarias algorithm; varying coefficient models; functional autoregressive models; father wavelets; warped wavelets

1 Introduction

Nonlinear time series models started to be strongly developed from the 1980s (see e.g. Tong, 1993) with parametric models. Good reasons for these developments are features of nonlinearity that frequently exceed the capacity of linear models such as the ARMA model (Box et al., 1994), the fractional ARMA model (Granger and Joyeux, 1980) and vector ARMA and vector ARMA model with exogenous variables (Hannan and Deistler, 1988). Examples of nonlinear parametric models that have been proposed are the exponential autoregressive (EXPAR) model of Haggan and Ozaki (1981), the threshold autoregressive (TAR) model of Tong (1983) and the smooth transition autoregressive (STAR) model of Granger and Teräsvirta (1993).

Although nonlinear parametric models (as the ones mentioned above) can deal with such features of nonlinearity, nonparametric models have gained more and more attention of researchers with the technological advances in the recent years. One can highlight, among other features, that these models are appealing because of their flexibility, which makes them useful to validate parametric models or suggest new ones. In time series analysis, mathematical tools such as splines, kernels and Fourier series have played an important role in the development of new nonparametric models.

In the field of regression models, the functional coefficient regression (FCR) model arises as an interesting alternative in modeling data sets because of its flexibility. Such a model is defined as follows. Let $\{Y_t, U_t, \mathbf{X}_t\}$ be a jointly strictly stationary process, where U_t is a real random variable and \mathbf{X}_t

a random vector in \mathbb{R}^d . Suppose that $\mathbb{E}(Y_t^2) < \infty$. Considering the multivariate regression function $m(\mathbf{x}, u) = \mathbb{E}(Y_t | \mathbf{X}_t = \mathbf{x}, U_t = u)$, the FCR model has the form

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

where the $f_i(\cdot)$ s are measurable functions from \mathbb{R} to \mathbb{R} and $\mathbf{x} = (x_1, \dots, x_d)^\top$, 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 \mathbf{X}_t are lagged values of Y_t , FCR models correspond to the functional-coefficient autoregressive (FAR) models of Chen and Tsay (1993).

FCR models, also known as varying-coefficient models, were initially studied by Hastie and Tibshirani (1993) in the case of independent data. In the context of time series analysis, we can highlight different approaches. Just to mention a few references, Chen and Tsay (1993) proposed a recursive method using arranged local regression to estimate the coefficient functions; Cai et al. (2000) and Chen and Liu (2001) considered local linear regression to estimate the regression function, using the same bandwidth in the estimation of the coefficient functions; Huang and Shen (2004) proposed a more flexible approach with polynomial splines, where the coefficient functions are allowed to be estimated with different number of knots. An advantage of this last paper is that polynomial splines estimate the coefficient functions globally.

A common feature of the papers mentioned above is that their authors assumed independent errors. However, in practical situations, it may be interesting to relax this assumption. In this sense, Montoril et al. (2014) generalized the spline approach of Huang and Shen (2004), where the errors were allowed to be correlated. We intend to keep using such a flexibility, but now focusing on the use of wavelets in the estimation of the functional coefficients of the FCR model.

The reason why we perform studies using wavelet bases is because this kind of approach has not been deeply studied for FCR models, although wavelets have been successfully used in many fields of Statistics. Some examples are density estimation (Donoho, 1993; Hall and Patil, 1995a,b), spectral density estimation (Gao, 1993, 1997; Moulin, 1994; Neumann, 1996) and nonparametric regression estimation (Donoho and Johnstone, 1995; Donoho et al., 1995; Donoho and Johnstone, 1998; Hall and Patil, 1996; Nason, 1996; Amato and Antoniadis, 2001; Brown et al., 2001; Amato et al., 2006; Kulik and Raimondo, 2009; Zhao et al., 2012). Also, see Ogden (1997) and Vidakovic (1999) for comprehensive survey of wavelet applications in Statistics.

Thus, in this work we study wavelet-based estimators for the functional coefficients of FCR models. The methodology proposed follows part of the idea of warped wavelets of Kerkycharian and Picard (2004), where we apply some transformation to U_t , in order to give more accurate estimates of the coefficient functions in regions where there is a smaller density of data. However, when the transformation is the identity, the estimator becomes based on the classical wavelets. Also, differently from the usual, in our study it is not necessary to assume that the errors are independent. This is important because, depending on the data set, residual analyses may indicate that this assumption is not being satisfied.

We organize this work as follows. In Section 2, we propose a wavelet-based method to estimate the coefficient functions of the model (1). Consistency and rates of convergence are presented and an algorithm for estimation is discussed. In Section 3, we present the criteria functions used in this work, in order to select the coarsest and finest scales in the wavelet bases, as well as the stepwise method for selecting variables for the wavelet-based method. In Section 4, some results of simulation studies are reported. The industrial production index data set is studied in Section 5, where we compute forecasts and compare the results with other models. In Section 6, we give some concluding remarks.

2 Wavelet-based estimator

In this section we present a general wavelet method for estimating the coefficient functions of the model (1). Such a method is based on warped wavelets, which we briefly explain.

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 .

Usually φ is called father wavelet (or scaling function) and it generates a basis $\{\varphi_{Jk}\}_k$ for the space V_J , where J is called resolution level and $\varphi_{Jk}(\cdot) = 2^{J/2}\varphi(2^J \cdot - k)$. Based on the MRA above it is possible to develop infinitely many different kind of scaling functions. Using these scaling functions, one can calculate the mother wavelets ψ (or simply wavelets), which generate a basis $\{\psi_{jk}\}_k$ for the orthogonal complement W_j of V_j in V_{j+1} . Two of the most famous families of compactly supported wavelets are the Daubechies and the Symmlets, whose mother wavelets have specific numbers of vanishing moments (see e.g. Daubechies, 1992; Härdle et al., 1998).

Frequently, studies related to wavelet-based estimators require assumptions like equally spaced data and dyadic (power of two) sample sizes. These assumptions are frequent because the wavelet-based estimators tend to be computed by discrete wavelet transforms (DWT), which are easy and fast to be calculated, requiring just $O(n)$ operations, where n corresponds to the sample size (see e.g. Mallat, 2008, for more details). 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, Daubechies and Symmlets) in specific points of interest with preassigned precision (more details in Vidakovic, 1999). Moreover, although the Daubechies-Lagarias algorithm is not as fast as the DWT, with the computational advances this issue is no longer a problem. Besides, this algorithm has the advantage of not requiring equally spaced data nor dyadic sample sizes.

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

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

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 φ_{Jk} is “warped” by the distribution function H . A consequence of this is

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

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

Warped wavelets were proposed by Kerkyacharian and Picard (2004) and are also related to the papers of Cai and Brown (1998, 1999). The interest of these authors was to keep using the DWT in situations where the data are no longer equally spaced. This was done by applying a transformation to the data set such that it becomes equally spaced, and then the usual wavelet techniques can be applied. The transformation applied in the afore mentioned papers was the empirical distribution function of the data.

For us, 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. Thus, we can use any distribution function that is able to achieve this aim. 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 $H(u) = u$.

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

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

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

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

2.1 Linear wavelet-based estimator

Since the functions in (1) may have different degrees of smoothness, it may be important to be able to consider different scaling functions and resolution levels on the estimation process of each coefficient function. Thus, let $\varphi_{(i)}$ be the father wavelet and J_i the resolution level associated to each f_i , $i = 1, \dots, d$. For the sake of simplicity, denote $\phi_{ik}(\cdot) = 2^{J_i/2} \varphi_{(i)}(2^{J_i} \cdot -k)$. Therefore, it is possible to use the orthogonal projection of each coefficient function onto a multiresolution space V_{J_i} , and then approximate the regression function (1) by

$$m(\mathbf{x}, u) \approx \sum_{i=1}^d \sum_{k=1}^{2^{J_i}} \alpha_{ik} \phi_{ik}(H(u)) x_i, \quad (3)$$

where $\mathbf{x} = (x_1, \dots, x_d)^\top$.

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

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

where ϵ_t corresponds to the errors of the model. Now, let Σ 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

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

where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^\top, \dots, \boldsymbol{\alpha}_d^\top)^\top$, $\boldsymbol{\alpha}_i = (\alpha_{i1}, \dots, \alpha_{i2^{J_i}})^\top$, $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$ and \mathbb{V} is a $n \times \sum_{i=1}^d 2^{J_i}$ matrix such that its t -th row corresponds to $\phi_{ik}(H(U_t)) X_{ti}$, $i = 1, \dots, d$, $k = 1, 2, \dots, 2^{J_i}$. The coefficient vector estimator is

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

Note that we can assume Σ 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

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

where $\hat{\alpha}_i = (\hat{\alpha}_{i1}, \dots, \hat{\alpha}_{i2^{J_i}})^\top$ is the estimator of α_i , $i = 1, 2, \dots, d$.

2.1.1 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

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

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 $y_n \lesssim x_n$.

Assumptions

- (W0) The eigenvalues of Σ 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 $\mathbb{E}(\mathbf{X}_t \mathbf{X}_t^\top | U_t = u)$ are uniformly bounded away from zero and infinity for all $u \in [0, 1]$;
- (W3) $2^{j_0} \asymp 2^{J_i} \asymp n^r$, $0 < r < 1$, $i = 1, \dots, d$;
- (W4) The process $\{Y_t, \mathbf{X}_t, U_t\}_{t \in \mathbb{Z}}$ is jointly strictly stationary. The α -mixing coefficient $\alpha(t)$ of $\{Y_t, \mathbf{X}_t, U_t\}_{t \in \mathbb{Z}}$ satisfies $\alpha(t) \lesssim t^{-\alpha}$ for $\alpha > (2+r)/(1-r)$;
- (W5) For some sufficient large $m > 0$, $\mathbb{E}|X_{ti}|^m < \infty$, $i = 1, \dots, d$;
- (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 $[0, 1]$.

Most of the assumptions (W0) – (W5) above are frequently used in the literature, and basically the same as the assumptions used in Montoril et al. (2014) (only (W3) differs, but it is analogous to another used). The assumption (W6) is quite general, because there are many different distribution functions (as in the cases of the Normal distribution, t -Student and Cauchy) satisfying it.

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*

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

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

The result of the theorem above is valid in both situations, when the errors of the model are independent (Σ is an identity matrix) and when they are correlated. It is easy to see that the assumption (W6) ensures that $\mathbb{E} \|\hat{f}_i - f_i\|_2^2$ and $\mathbb{E} \|\hat{f}_i - f_i\|_{L_2(h)}^2$ have the same rate of convergence, because the probability density function is bounded away from zero and infinity on the support of the coefficient functions. Moreover, denote $\rho_n = \max_{1 \leq i \leq d} \rho_i$ and $J_n = \max_{1 \leq i \leq d} J_i$. It is straightforward that $\mathbb{E} \|\hat{f}_i - f_i\|_2^2 = O\left(\frac{2^{J_n}}{n} + \rho_n^2\right)$, $i = 1, \dots, d$, which is similar to the rates of convergence of the spline-based estimators in Huang and Shen (2004) and Montoril et al. (2014). The quantities ρ_i , $i = 1, \dots, d$, measure the size of the approximation error and their magnitudes are determined according to the smoothness of the coefficient functions, as discussed by Huang and Shen (2004).

It is frequently assumed that the function of interest belongs to Besov spaces in works involving wavelet-based estimators (some examples are Kerkycharian and Picard, 1992; Donoho et al., 1996; Kulik and Raimondo, 2009). In this case, based on Theorem 9.3 of Härdle et al. (1998), it is possible to see that $\rho_i = O(2^{-J_i s})$, where s corresponds to the regularity parameter of the Besov space. This ensures that the rate of convergence of \hat{f}_i to f_i is of order $\frac{2^{J_n}}{n} + 2^{-2s J_n}$. Such a rate of convergence will be minimized when 2^{J_n} has the same rate of convergence as the sequence $n^{1/(2s+1)}$. In this case the rate of convergence of \hat{f}_i to f_i is of order $n^{-\frac{2s}{2s+1}}$.

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

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

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

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

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

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

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

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

Since the results of the theorem above are similar to the results in Theorem 2.1, their conclusions are analogous.

2.1.2 Algorithm for estimating the coefficient vector

In the Section 2 we discuss the estimation of the coefficient functions of the FCR model using a wavelet-based approach. The estimator of the wavelet coefficients is presented in matrix notation, as in (7). The estimation process could be done iteratively, fixing previous estimates of the vector α in order to estimate θ (the autoregressive coefficients of the errors) and *vice-versa*, until some convergence criterion be reached, e.g. the convergence of the residual mean square. This procedure tends to provide estimates that are close to minimize the least squares function. The drawback is that the calculations under matrix notation may be computationally expensive, since it is necessary to handle big matrices.

The estimation under backshift notation, applying autoregressive filters in the least squares function, provides an analogous and more efficient method.

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 η the vector $(\alpha^\top, \theta^\top)^\top$ and \mathbf{v}_t as the t -th row of \mathbb{V} , we estimate the wavelet coefficients α and the autoregressive coefficients θ jointly, minimizing numerically

$$\ell(\eta) = \sum_{t=1}^n \left\{ \theta_p(L) \left(Y_t - \mathbf{x}_t^\top \alpha \right) \right\}^2, \quad (8)$$

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

Algorithm

- (a1) Estimate the coefficient vector α by ordinary least squares, and denote it by $\hat{\alpha}$;
- (a2) Fit an autoregressive model to the residuals of step (a1), i.e., $\hat{\epsilon}_t = Y_t - \mathbf{x}_t^\top \hat{\alpha}$, say,

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

- (a3) Estimate η numerically, minimizing (8), using the estimates in steps (a1) and (a2) as initial values.

2.2 Regularized wavelet-based estimators

An interesting advantage of wavelet-based estimators is their ability to adapt more satisfactory with irregular functions. The MRA can provide expansions based on scaling functions and wavelets in different scales. As a result, one can reduce the noise in the estimates shrinking or thresholding detail coefficients.

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

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

where the β_{jk} 's correspond to the detail coefficients at scale j and j_0 to the coarsest level (or scale). In this case, $J - 1$ is usually known as the finest level (or scale). 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 $\bar{\alpha}_{ik}$, $k = 1, \dots, 2^{J_i}$ (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 $\bar{\alpha}_{jk}^h$ (the superscript "h" would indicate that it was applied the hard threshold to the detail coefficients in the wavelet domain).

For the sake of simplicity, we will consider in this work only the hard thresholding approach to regularize the linear wavelet-based estimators. However, the procedure is similar to other methodologies and the theoretical results also hold in other shrinking/thresholding cases. See e.g. Nason (2008) for other examples of shrinking and thresholding methods.

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 λ 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 σ (standard deviation of the noise) is unknown in practice and is usually estimated by the 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 $\text{MAD}(x) = 1.4826 \cdot \text{median}(|x - \text{median}(x)|)$, 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 \hat{f}_i . In order to make it clearer, the steps of regularization are summarized below.

2.2.1 Steps to obtain the regularized coefficient functions

- (s1) Apply the DWT to $\bar{\alpha}_{i1}, \dots, \bar{\alpha}_{i2^{J_i}}$ and obtain $\{\{\bar{\alpha}_{j_0 r}\}, \{\bar{\beta}_{jk}\}, r = 1, \dots, 2^{j_0}, k = 1, \dots, 2^j, j = j_0, \dots, J_i - 1\}$;
- (s2) Calculate $\hat{\sigma} = \text{MAD}(\bar{\beta}_{J_i-1})$, where $\bar{\beta}_{J_i-1} = (\bar{\beta}_{J_i-1,1}, \dots, \bar{\beta}_{J_i-1,2^{J_i-1}})^\top$, and set $\hat{\lambda} = \hat{\sigma}\sqrt{2(J_i - 1)\log 2}$;
- (s3) Apply the hard thresholding method to each detail coefficient defining $\bar{\beta}_{jk}^h = \eta_{\hat{\lambda}}(\bar{\beta}_{jk})$, $k = 1, \dots, 2^j, j = j_0, \dots, J_i - 1$;
- (s4) Apply the inverse DWT to the thresholded coefficients and obtain $\bar{\alpha}_{i1}^h, \dots, \bar{\alpha}_{i2^{J_i}}^h$. The regularized coefficient function estimate will be

$$\bar{f}_i^h(u) = \sum_{k=1}^{2^{J_i}} \bar{\alpha}_{ik}^h \phi_{ik}(H(u)).$$

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 $\hat{f}_i, i = 1, \dots, d$, presented in Section 2.1. In this case, in step (s1) above, the DWT would be applied to $\hat{\alpha}_{i1}, \dots, \hat{\alpha}_{i2^{J_i}}$. The resulting regularized coefficient functions estimated could be denoted by $\hat{f}_i^h(u), i = 1, \dots, d$.

2.2.2 Theoretical results

Once the linear wavelet-based estimators presented in Section 2.1 are regularized, it is possible to show that their asymptotic properties do not change. The results are presented in the following theorems.

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

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

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

As in the case of the linear estimators, the theorem above is useful to ensure the consistence of the (now regularized) wavelet-based estimators for the case where the covariance matrix of the errors are unknown, as it happens in practice. Since the rates of convergence are the same, all the conclusions with respect to Theorem 2.1 are analogous in the theorems of the regularized case.

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

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

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

3 Selecting the resolution level and variables

The assumption (W3) in Section 2, used in the presented theoretical results, is related to the choice of the multiresolution space V_{J_i} where each coefficient function will be projected so that we can calculate its estimate. However, 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 scale used during the regularization (see Section 2.2.1). An alternative is to use an automatic method to select the more appropriate values for the finest and coarsest scales ($J_i - 1$ 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).

Denote the sample size by n and the residual mean square by RMS, which will correspond to $\ell(\bar{\eta})/n$, where ℓ 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

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

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 $S_r \subseteq \{1, \dots, q\}$, 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 $1 \leq r \leq q$.

4 Simulation study

In this section we run some simulation studies, in order to compare the performance of the wavelet-based approach proposed in Section 2. We can compare, for example, the estimates using the three criteria functions for selecting the coarsest and the finest scales; different situations for the errors of the model; and the quality of estimates of the proposed methods.

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 (ASE_i), and it is defined as

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

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

where $\{u_k, k = 1, \dots, n_{\text{grid}}\}$ 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 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 ASE '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.

We simulate two different models: the EXPAR model and another alternative model, where the coefficient functions are not so smooth as in the EXPAR case. For both situations, similar structures are used:

- 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 *Daubechies* D8, D12, D16, D20, and *Symmlets* S8, S12, S16, S20, where DN corresponds to the N -tap of Daubechies' Extremal Phase wavelet filter, and SN 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 scale (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 about here.]

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

We considered two different wavelet-based estimators, which are called Classical and Warped. In the first case, the warping function is the identity function (i.e., $H(u) = u$). The Warped estimates are obtained using the warping function H as the empirical distribution function linearly interpolated. Based on these two wavelet-estimates, we observed which criterion function (AIC, AICc or BIC) is selecting the most appropriate values for the coarsest and finest scales. For this simulation study, we consider as candidates to the finest and coarsest levels the values based on $J = 2, 3, 4$, with $j_0 = 2, \dots, J$. The case where $j_0 = J$ is a convention to indicate the linear wavelet-based estimator.

The results are summarized in Table 2, and different conclusions can be taken from there. For each criterion function, one can see that the estimates based on the Classical estimates tend to be better than the Warped estimates for all the AR order used for the errors. With respect to the criteria functions, all of them seem to be selecting quite well the coarsest and the finest scales (among their candidates), although the estimates with the AICc are a little better for the Classical wavelet-based case, and the estimates using the BIC criterion show themselves more interesting for the Warped wavelet-based proposal.

[Table 2 about here.]

Since the errors of the EXPAR models generated are correlated in our simulation study, we believe that an interesting approach to be used for comparisons with the wavelet-based estimates corresponds to the spline-based method proposed by Montoril et al. (2014). Thus, using quadratic splines, we computed the ASE²'s based on the three criteria functions (AIC, AICc and BIC) to select the best number of knots (equally spaced). The values used as candidates for the number of knots are varying from 2 to 5. The results are presented in Table 3. In order to make easier the comparisons, we also present in this table the results of the Classical and Warped wavelet-based estimates using the D16 wavelet basis. Observe, initially, that the AIC and AICc are providing the best estimates, as it happened in Montoril et al. (2014). Furthermore, one can observe that the results presented in Table 2 are very competitive, when compared to the results based on splines, specially the Classical estimates. This indicates that the wavelet-based method is able to provide estimates as good as the spline approach.

[Table 3 about here.]

In order to illustrate the estimates, we use one generated data set of the EXPAR model and present in Figure 1 the estimates of f_1 and f_2 based on the Classical and Warped approaches, and the spline method. For this data, in the Classical case, the AIC/AICc selected $j_0 = 3$ and $J = 3$, which corresponds to a linear wavelet-based method. With respect to the Warped case, the AICc function (as well as the BIC) suggested $j_0 = 2$ and $J = 2$. Finally, the spline method was fitted with five knots, selected by the AIC. In this case, the Warped estimates seem to providing values closer to the real coefficient functions. Although the Classical wavelet estimates look biased for f_1 , the shape of its estimates are quite similar to the real functions, which is not happening so well for the Spline estimates, specially for negative values of the coefficient function f_2 in extreme regions.

[Figure 1 about here.]

4.2 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 μ and standard deviation σ at the point x , the coefficient functions used in this example are

$$\begin{aligned} f_1(u) &= 0.8 + 0.5d(u; 0; 0.03) \\ f_2(u) &= -0.4 - 0.25(d(u; -0.2; 0.05) + d(u; 0.2, .05)). \end{aligned}$$

The idea in this simulation study is to evaluate how the estimates behave when the coefficient functions are not as smooth as in the EXPAR case. Actually, the coefficient functions resemble constants with

short “bumps” in small regions. In this study, we consider as candidates for the finest and the coarsest scales the values based on $J = 4, 5, 6$ and $j_0 = 2, \dots, J$. Again, the case where $j_0 = J$ corresponds to the linear wavelet-based estimation.

In Table 4 are presented the results of this simulation study. Observe that for the three criteria functions the Warped wavelet-based estimates overcome the estimates obtained using the Classical wavelet-based method. Furthermore, in the Classical wavelet-based case, the AIC and AICc functions are providing the best selections of coarsest and finest scales. On the other hand, in this simulation the AICc tends to provide better selections of j_0 and J for the Warped wavelet-based method.

[Table 4 about here.]

In this example of simulation we also run estimations by the spline methodology and, as in the first example, we present the Classical and Warped results using the D16 wavelet basis. We considered as candidates to be selected by the criteria functions the number of knots (equally spaced) varying from 2 to 30. The results are presented in Table 5. The good performance of the Warped approach is highlighted when compared to the spline-based estimates. Furthermore, we can observe that the Classical method is providing results very competitive when compared to the spline-based estimates (the Classical estimates do not overcome the spline estimates only for the case where the errors are simulated by an AR(2)). In this simulation, we can also see that the best results of the spline method are obtained by selecting the number of knots with the AIC function, as happened in the first example of simulation.

[Table 5 about here.]

A data set is generated and fitted according to the Classical and Warped wavelet-based methods, and by the spline-based approach. The AIC/AICc selected $j_0 = J = 5$ in the Classical case. The AICc suggested $j_0 = 4$ and $J = 5$ to be used for the Warped wavelet method. In order to fit the spline-based model, the AIC/AICc functions indicated the value 7 to be the number of knots. The estimates are presented in Figure 2. We can observe, for example, how hard it is to the spline estimates to follow the shape of the real coefficient functions. On the other hand, we can see that the Warped estimates are able to follow the “irregularities” of the real coefficient functions more satisfactory than the Classical wavelet-based estimates. Furthermore, the Warped estimates have shown themselves more accurate at more extreme regions.

[Figure 2 about here.]

4.3 Classical Wavelets vs. Warped Wavelets vs. Splines and the criteria functions

The two examples of simulations presented in Sections 4.1 and 4.2 compared two different cases (among infinitely many other possible) of wavelet-based estimation of coefficient functions, where we considered a case that we called Classical and another that we called Warped. In the Classical case, we used as warping function the identity function (or the accumulated distribution function of a Uniform distribution in the unit interval). In the second case, we considered as warping function the empirical cumulated distribution function linearly interpolated.

Using an approximation to the mean integrated square error, denoted here by ASE^2 , we compared the performance of these two wavelet-based estimates, according to several different wavelet bases. Furthermore, using three different criteria functions (AIC, AICc and BIC) we selected the best pair of coarsest and finest scales. An interesting advantage of using this mechanisms is that they are also able to select linear wavelet-based fitting, when it provides the best coefficient function estimates. The results have indicated that, for both wavelet-based cases, the AICc can be considered an appropriate criterion function to select the coarsest and finest scales.

It was observed indications that the Classical method tend to provide better estimates than the Warped case when the coefficient functions are smooth (as in the first example of simulations). On the other hand,

when the coefficient functions have short but abrupt deviations (as in the second example of simulation), the Warped wavelet-based estimates presented the best results.

Furthermore, we compared the wavelet-based estimates with the spline approach, and it was possible to be observed that the wavelet-based results were quite competitive in the first example of simulation, when the coefficient functions are smooth. However, in the second example, when the coefficient functions were no longer too smooth, it became clear the advantage of considering the wavelet approach instead of the spline method, specially the Warped case.

In both examples of simulation, the *Daubechies* D16 presented the good estimates for Classical and Warped cases.

5 Application to Industrial Production Index

In order to illustrate the proposed methodology, we consider the monthly seasonally adjusted industrial production index (IPI) of the USA, from December 1980 to December 2007, with 325 observations. The data set was obtained from the Federal Reserve Bank of St. Louis (<http://research.stlouisfed.org>). We apply the log-return transformation to the series, i.e., $Y_t = 100 \times \log(X_t/X_{t-1})$, resulting in 324 observations that can be observed in Figure 3. This time series is of great interest in economics, frequently fitted by ARMA models. In this work we model its dynamic behavior using wavelet-based methodology of Section 2. Furthermore, we evaluate the performance of the proposed method running out-of-sample forecasts.

[Figure 3 about here.]

As in the simulation study, we take into account the wavelet-based modeling using the identity function as warping function (Classical Wavelets) and also the empirical cumulative distribution function, linearly interpolated, as the warping function (Warped Wavelets).

We consider the set $\{1, 2, 3\}$ to select the candidates for significant lags, as well as for the threshold lag. For the sake of simplicity, the model selection is performed in two stages. In the first stage we focus on the selection of the threshold lag and the significant lags, according to the coarsest and finest scales. Once these quantities are selected, in the second stage we perform residual analyses to each resulting model, in order to detect a more appropriate autoregressive order to the errors, if necessary. The coefficient functions and the AR coefficients of the errors are then jointly estimated and the value of the criterion function is recalculated (AICc for the wavelet-based models). The final model is chosen according to the one that presents the smallest value of the criterion function used. For the residual analysis we used Ljung-Box tests, autocorrelations and partial autocorrelations.

It is important to mention that we regularized the wavelet coefficients using the hard thresholding with the universal policy, as described in Section 2.2. The candidates for the coarsest and finest scales to be used are based on $0 \leq j_0 \leq J$ and $2 \leq J \leq 4$. We consider the Daubechies D16 wavelet basis to estimate the coefficient functions.

The selected models for the Classical wavelet-based method is presented in Table 6. The best model was the case where $j_0 = J = 4$, which corresponds to the linear wavelet-based estimator.

[Table 6 about here.]

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

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

where ε_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 about here.]

The selected models for the Warped wavelet-based method are presented in Table 7. Observe that, with the only exception of pair $j_0 = 4$ and $J = 4$, there is agreement in the choice of threshold lag and significant lags, as well as in the assumption of no correlated errors, which was verified in residual analyses. As in the selection for the Classical case, the smallest AICc value came from a the linear wavelet-based estimates, now with $j_0 = J = 2$. The final model corresponds to

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

where ϵ_t is white noise.

[Table 7 about here.]

The fitted Warped model provided the coefficient function estimates presented in Figure 5.

[Figure 5 about here.]

With respect to the forecasts, it is important to mention that the procedure adopted is similar to Huang and Shen (2004) and Montoril et al. (2014). Briefly, consider a time series $\{Y_1, \dots, Y_T\}$ following the dynamic of a functional coefficient regression model (let us assume with autoregressive errors). In order to predict Y_{T+h} , for a given lead time h , it is possible to generate predictions based on the estimates of the fitted model. Basically, sampling (with replacement) the predicted white noise and applying an autoregressive filter, using the autoregressive estimates of the residuals, it is possible to generate new residuals, say $\{\epsilon_{b,T+1}, \dots, \epsilon_{b,T+h}\}$. Using these generated residuals, we can compute recursively the forecasts $\{\hat{Y}_{b,T+1}, \dots, \hat{Y}_{b,T+h}\}$ using the coefficient function estimates. This procedure can be repeated B times, i.e., $1 \leq b \leq B$.

The optimal predictor that minimizes the mean squared error criterion is $\mathbb{E}(Y_{T+h}|\mathcal{F}_T)$, where \mathcal{F}_T corresponds to the given information set up to the time T . Then we can consider $\hat{Y}_{T+k} = (1/B) \sum_{b=1}^B \hat{Y}_{b,T+k}$, $k = 1, \dots, h$, as the forecasts. It is also possible, using the simulated forecasts, to estimate density forecasts and forecast intervals (see Huang and Shen, 2004, for more details). Based on the arguments of these authors, we consider only forecasts inside of the range of the data set (forecasts out of the range of the historical data will be discarded), because coefficient function estimates out of these range are not reliable. In this paper, the forecasts are calculated based on $B = 5,000$ valid replications.

In order to evaluate the performance of the forecasts, we consider an approximation to the mean squared prediction error (MSPE) for 12 multi-step-ahead forecasts. In this case, we apply a rolling procedure considering 60 subseries (Y_1, \dots, Y_T , with $T = 253, \dots, 312$), making 12 multi-step-ahead forecasts and calculating their square prediction error. Thus, with the 60 results for the 12 multi-step-ahead forecasts, the MSPE is calculated by averaging the squared prediction errors.

For each subseries, we choose the best candidate for the coarsest and finest scales, among $2 \leq j_0 \leq J$, $2 \leq J \leq 4$, and the threshold lag and the significant lags among $\{1, 2, 3\}$ using the Daubechies D16 wavelet basis. In order to select the appropriate AR order for the errors, we use the AIC among the candidates $\{1, \dots, 7\}$. Then, with the chosen model, we make 12 multi-step-ahead forecasts and, applying the procedure described above, we calculate the MSPEs for the forecasts. Both methods were taken into account, the Classical and the Warped wavelet-based approaches.

For the sake of comparisons, we also consider the AR model and spline-based method of Montoril et al. (2014). In these cases, the selection of the order was made by the AIC. For the spline-based approach, we use quadratic splines, with the number of (equally spaced) knots being chosen among the candidates $\{2, \dots, 10\}$. The selection of the FAR model, in the spline case, was made similarly to the wavelet-based methodology.

The forecast results are presented in Table 8. Observe that the MSPEs obtained from the Warped wavelet-based method tend to overcome the others. The forecasts of the Warped method, when compared to the AR model are better in 8 of the 12 horizons (4, 6, 7, 8, 9, 10, 11, 12). With respect to the Classical approach, only in the steps 11 and 12 the Warped method does not provide better forecasts. Finally, when

the forecasts of the Warped wavelet-based model are compared to the Spline-based forecasts, one can see that, in 11 of the 12 forecast horizons the Warped case overcomes the spline approach (for horizons greater than or equal to 2).

[Table 8 about here.]

6 Conclusions and Remarks

In this work we propose a wavelet-based approach for estimating coefficient functions of FCR models. The methodology takes into account the idea of warped wavelets, where the wavelet bases is “warped” by a function H . This can be seen as a general method, because when H is a identity function, the coefficient function estimator is based on the classical wavelets. The proposed method is able to take into account the case where the errors are correlated. Furthermore, the methodology may consider the regularization of the coefficient function estimates. In this work we used the hard threshold method.

The coarsest and finest scales of the wavelet-based estimates are selected according the a similar procedure used by Huang and Shen (2004) and Montoril et al. (2014), based on some criteria function. An advantage of using this procedure is that it can also suggest when it is not necessary to regularize the wavelet-based estimates.

In order to evaluate the potential of the proposed methodology, we performed numerical studies and compared the results with other approaches. In simulation studies, we compared the results of two wavelet-based estimates (Classical and Warped cases) with the spline-based method. We could see improvement of the coefficient function estimates, specially when the coefficient functions are not too smooth.

An application to a real data was also performed. In this case we used the monthly seasonally adjusted industrial production index (IPI) of the USA, from December 1980 to December 2007. We compared forecasts of two wavelet-based methods (as in the simulation study, the Classical and the Warped cases) with the spline-based model and the AR methodology. It was observed that the Warped wavelet-based model overcame the other methods, highlighting the good quality of our proposal.

Acknowledgments

This work was supported by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) under Grants 2008/51097-6, 2009/09588-5, 2013/09035-1 and 2013/21273-5.

Bibliography

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6), 716–723.
- Amato, U. and A. Antoniadis (2001). Adaptive wavelet series estimation in separable nonparametric regression models. *Statistics and Computing* 11(4), 373–394.
- Amato, U., A. Antoniadis, and I. De Feis (2006). Dimension reduction in functional regression with applications. *Computational Statistics & Data Analysis* 50(9), 2422–2446.
- Box, G., G. M. Jenkins, and G. Reinsel (1994). *Time Series Analysis: Forecasting & Control* (3rd ed.). Prentice Hall.
- Brown, P. J., T. Fearn, and M. Vannucci (2001). Bayesian wavelet regression on curves with application to a spectroscopic calibration problem. *Journal of the American Statistical Association* 96(454), 398–408.

- Cai, T. T. and L. D. Brown (1998). Wavelet shrinkage for nonequispaced samples. *The Annals of Statistics* 26(5), 1783–1799.
- Cai, T. T. and L. D. Brown (1999). Wavelet estimation for samples with random uniform design. *Statistics & Probability Letters* 42(3), 313–321.
- Cai, Z., J. Fan, and Q. Yao (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association* 95(451), 941–956.
- Chen, R. and L.-M. Liu (2001). Functional coefficient autoregressive models: Estimation and tests of hypotheses. *Journal of Time Series Analysis* 22(2), 151–173.
- Chen, R. and R. S. Tsay (1993). Functional-Coefficient autoregressive models. *Journal of the American Statistical Association* 88(421), 298–308.
- Daubechies, I. (1992). *Ten Lectures on Wavelets*. Regional Conference Series in Applied Mathematics. Philadelphia: Society for Industrial and Applied Mathematics.
- 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.
- Donoho, D. L. (1993). Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data. In *In Proceedings of Symposia in Applied Mathematics*, Volume 47, pp. 173–205. AMS.
- Donoho, D. L. and I. M. Johnstone (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association* 90(432), 1200–1224.
- Donoho, D. L. and I. M. Johnstone (1998). Minimax estimation via wavelet shrinkage. *The Annals of Statistics* 26(3), 879–921.
- Donoho, D. L., I. M. Johnstone, G. Kerkyacharian, and D. Picard (1995). Wavelet shrinkage: Asymptopia? *Journal of the Royal Statistical Society. Series B (Methodological)* 57(2), 301–369.
- Donoho, D. L., I. M. Johnstone, G. Kerkyacharian, and D. Picard (1996). Density estimation by wavelet thresholding. *The Annals of Statistics* 24(2), 508–539.
- Donoho, D. L. and J. M. Johnstone (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* 81(3), 425–455.
- Gao, H. Y. (1993). *Wavelet Estimation of Spectral Densities in Time Series Analysis*. Ph. D. thesis, University of California Berkeley.
- Gao, H.-Y. (1997). Choice of thresholds for wavelet shrinkage estimate of the spectrum. *Journal of Time Series Analysis* 18(3), 231–251.
- Granger, C. W. J. and R. Joyeux (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* 1(1), 15–29.
- Granger, C. W. J. and T. Teräsvirta (1993). *Modelling nonlinear economic relationships*. Oxford University Press.
- Haggan, V. and T. Ozaki (1981). Modelling nonlinear random vibrations using an Amplitude-Dependent autoregressive time series model. *Biometrika* 68(1), 189–196.

- Hall, P. and P. Patil (1995a). Formulae for mean integrated squared error of nonlinear Wavelet-Based density estimators. *The Annals of Statistics* 23(3).
- Hall, P. and P. Patil (1995b). On wavelet methods for estimating smooth functions. *Bernoulli* 1(1/2), 41–58.
- Hall, P. and P. Patil (1996). On the choice of smoothing parameter, threshold and truncation in nonparametric regression by Non-Linear wavelet methods. *Journal of the Royal Statistical Society. Series B (Methodological)* 58(2), 361–377.
- Hannan, E. J. and M. Deistler (1988). *The Statistical Theory of Linear Systems*. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons.
- Härdle, W., G. Kerkycharian, D. Picard, and A. B. Tsybakov (1998). *Wavelets, Approximation and Statistical Applications*, Volume 129 of *Lecture notes in statistics*. New York: Springer.
- Hastie, T. and R. Tibshirani (1993). Varying-Coefficient models (with discussion). *Journal of the Royal Statistical Society. Series B (Methodological)* 55(4), 757–796.
- Huang, J. Z. and H. Shen (2004). Functional Coefficient Regression Models for Non-linear Time Series: A Polynomial Spline Approach. *Scandinavian Journal of Statistics* 31(4), 515–534.
- Hurvich, C. M. and C.-L. Tsai (1989). Regression and time series model selection in small samples. *Biometrika* 76(2), 297–307.
- Kerkycharian, G. and D. Picard (1992). Density estimation in besov spaces. *Statistics & Probability Letters* 13(1), 15–24.
- Kerkycharian, G. and D. Picard (2004). Regression in random design and warped wavelets. *Bernoulli* 10(6), 1053–1105.
- Kulik, R. and M. Raimondo (2009). Wavelet regression in random design with heteroscedastic dependent errors. *The Annals of Statistics* 37(6A), 3396–3430.
- Mallat, S. (2008). *A Wavelet Tour of Signal Processing: The Sparse Way* (3rd ed.). New York: Academic Press.
- 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.
- Moulin, P. (1994). Wavelet thresholding techniques for power spectrum estimation. *Signal Processing, IEEE Transactions on* 42(11), 3126–3136.
- Nason, G. P. (1996). Wavelet shrinkage using Cross-Validation. *Journal of the Royal Statistical Society. Series B (Methodological)* 58(2), 463–479.
- Nason, G. P. (2008). *Wavelet methods in statistics with R*. Use R! New York: Springer.
- Neumann, M. H. (1996). Spectral density estimation via nonlinear wavelet methods for stationary non-gaussian time series. *Journal of Time Series Analysis* 17(6), 601–633.
- Ogden, R. T. (1997). *Essential wavelets for statistical applications and data analysis*. Birkhäuser; Boston.
- Restrepo, J. M. and G. K. Leaf (1997). Inner product computations using periodized daubechies wavelets. *International Journal for Numerical Methods in Engineering* 40(19), 3557–3578.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* 6(2), 461–464.

- Tong, H. (1983). *Threshold Models in Non-linear Time Series Analysis*. Lecture Notes in Statistics. New York: Springer.
- Tong, H. (1993). *Non-Linear Time Series: A Dynamical System Approach*, Volume 6 of *Oxford Statistical Science Series*. Oxford: Clarendon Press.
- Vidakovic, B. (1999). *Statistical Modeling by Wavelets*. Wiley Series in Probability and Statistics. New York: Wiley-Interscience.
- Zhao, Y., R. T. Ogden, and P. T. Reiss (2012). Wavelet-Based LASSO in functional linear regression. *Journal of Computational and Graphical Statistics* 21(3), 600–617.

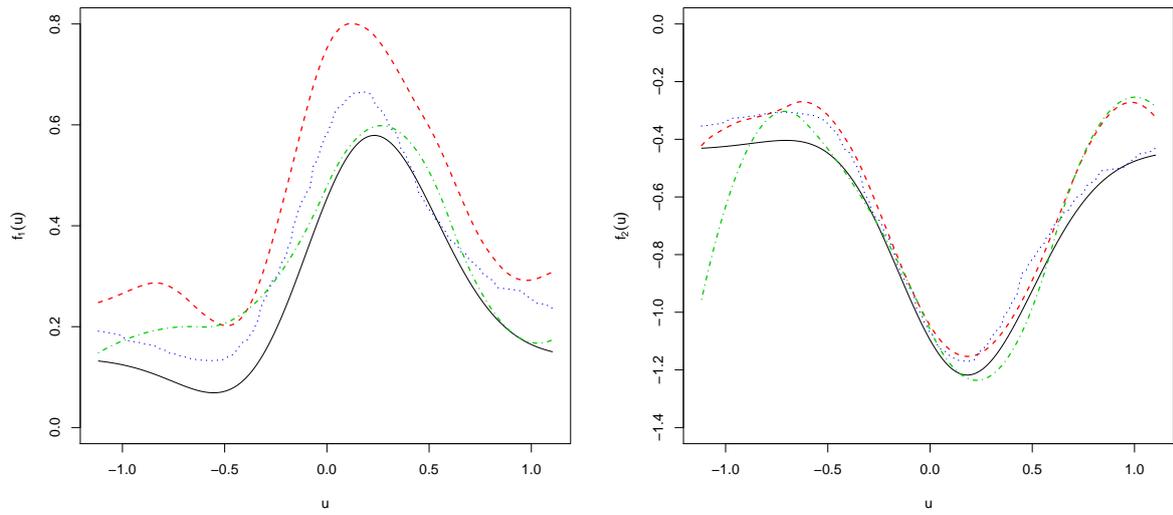


Figure 1: Classical wavelet-based (dashed lines), Warped wavelet-based (dotted lines) and Spline (dot-dashed lines) estimates of the coefficient functions f_1 (left side) and f_2 (right side) of the EXPAR model generated. The real coefficient functions correspond to the solid lines.

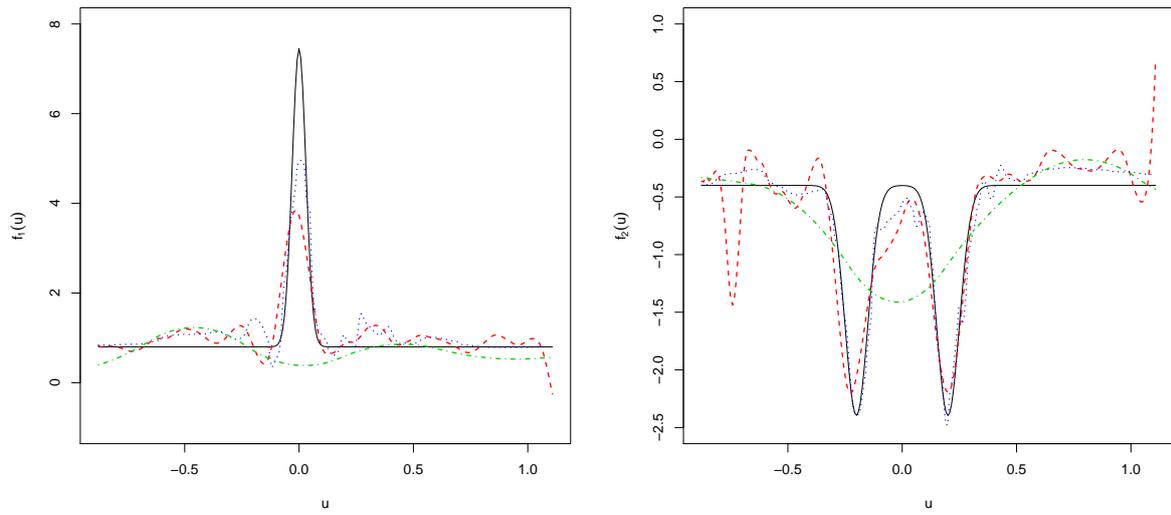


Figure 2: Classical wavelet-based (dashed lines), Warped wavelet-based (dotted lines) and Spline (dot-dashed lines) estimates of the coefficient functions f_1 (left side) and f_2 (right side) of the data generated in according to the simulation of the second example. The real coefficient functions correspond to the solid lines.

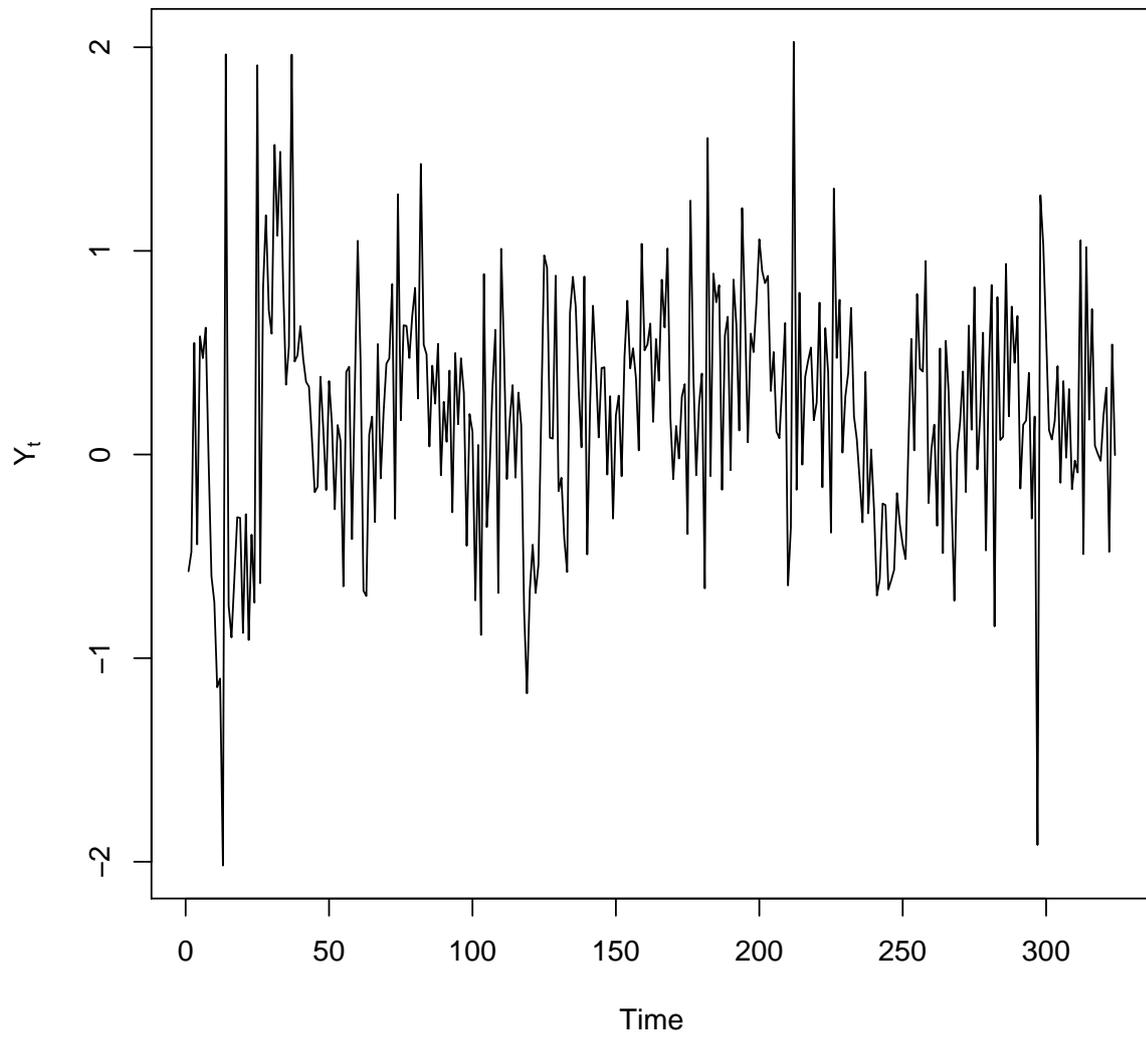


Figure 3: *Log-return of the monthly seasonally adjusted industrial production index, from January 1981 to December 2007 .*

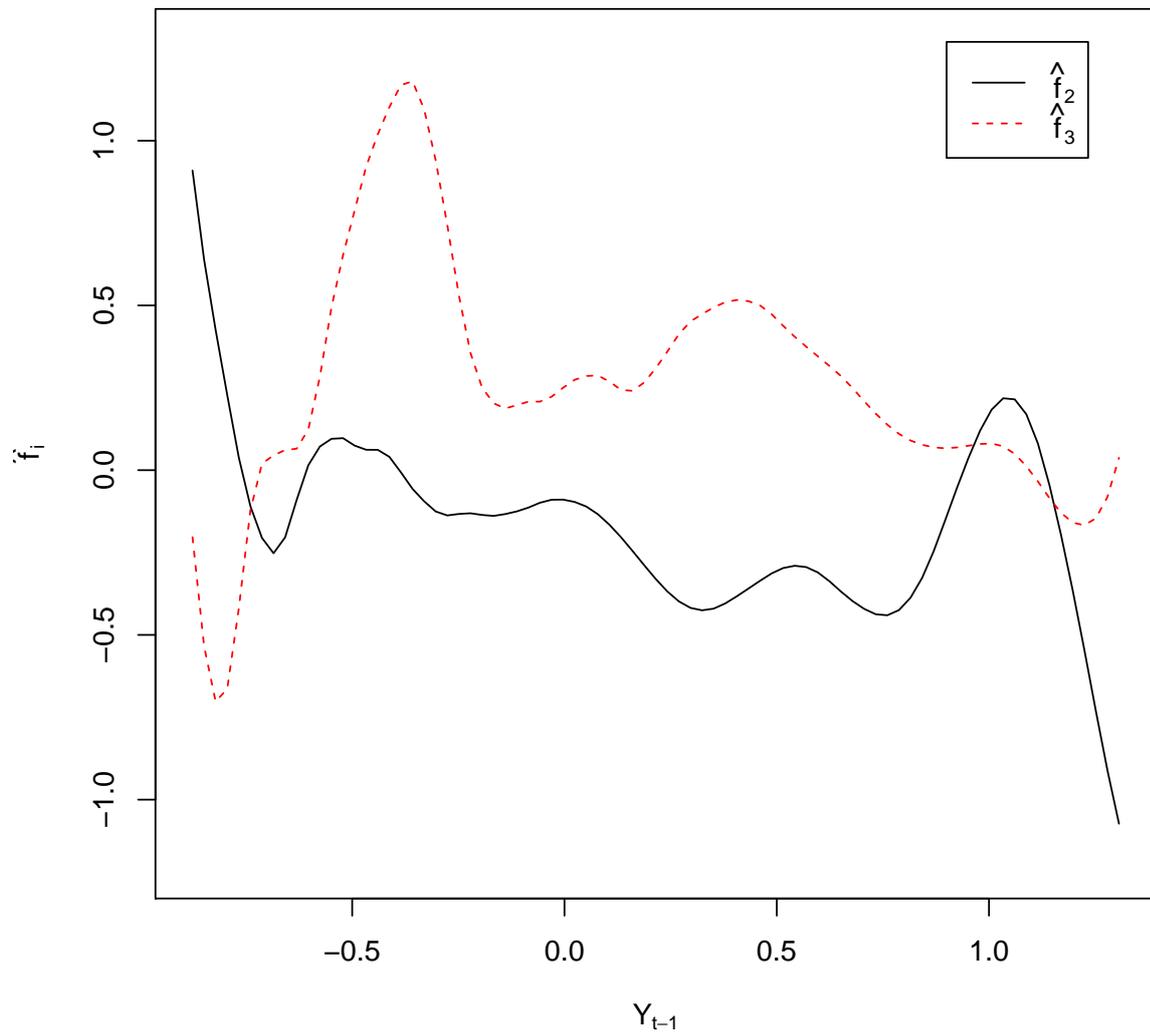


Figure 4: Classical wavelet-based estimates of the coefficient functions of model (9) for the industrial production index series.

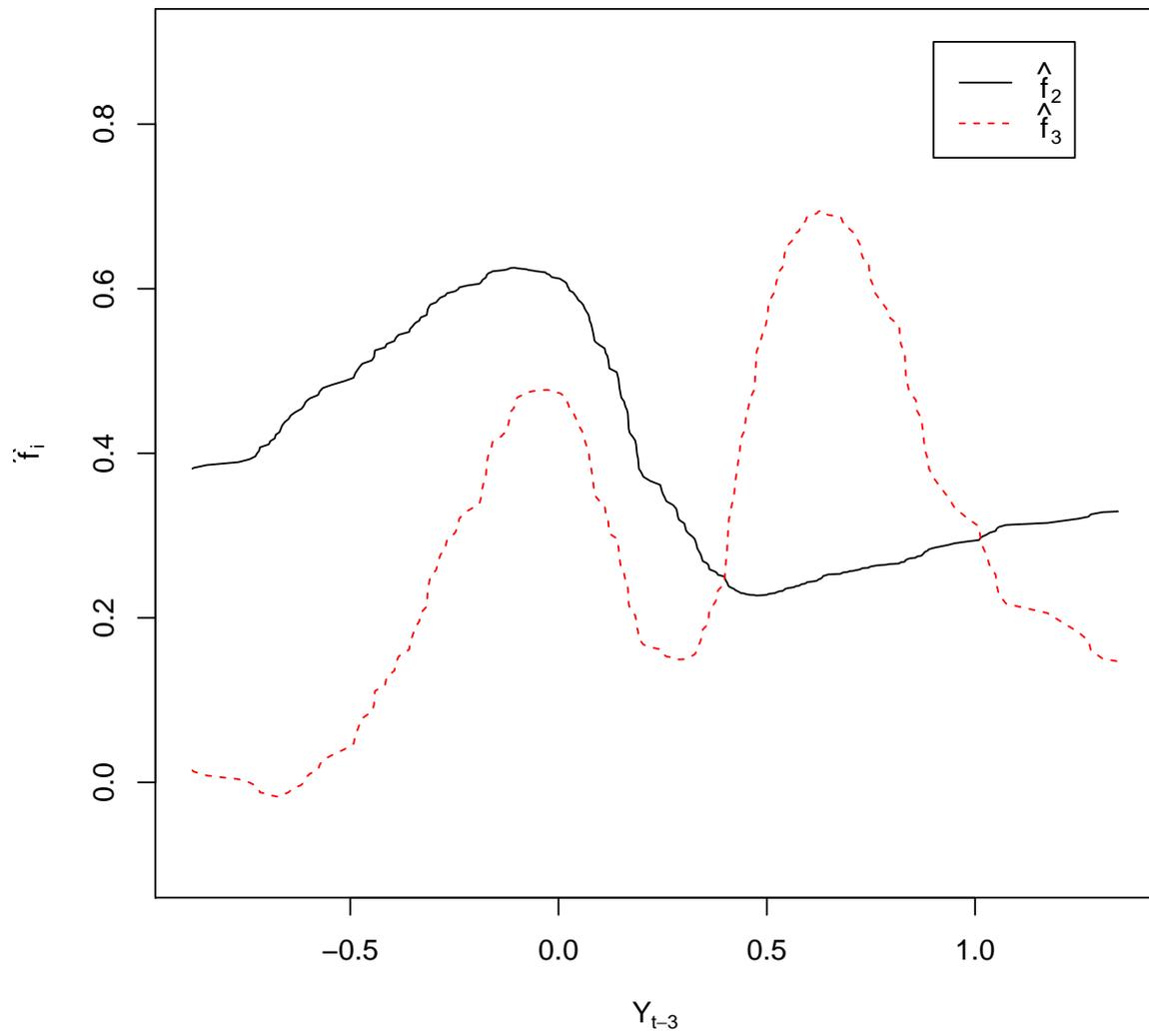


Figure 5: Warped wavelet-based estimates of the coefficient functions of model (10) for the industrial production index series.

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

p	$\phi_p(L)$	σ_ε
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

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 scales 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	0.011	0.004	0.010	0.004	0.006
		(0.003)	(0.010)	(0.003)	(0.009)	(0.003)	(0.005)
	D12	0.004	0.027	0.004	0.025	0.006	0.018
		(0.003)	(0.024)	(0.003)	(0.021)	(0.004)	(0.011)
	D16	0.004	0.014	0.003	0.013	0.004	0.008
		(0.003)	(0.012)	(0.003)	(0.011)	(0.003)	(0.006)
	D20	0.004	0.023	0.004	0.022	0.005	0.015
		(0.002)	(0.019)	(0.002)	(0.017)	(0.004)	(0.008)
S8	0.005	0.030	0.005	0.028	0.007	0.018	
	(0.003)	(0.028)	(0.003)	(0.025)	(0.005)	(0.012)	
S12	0.004	0.025	0.004	0.024	0.007	0.014	
	(0.003)	(0.024)	(0.003)	(0.022)	(0.005)	(0.008)	
S16	0.004	0.021	0.004	0.020	0.006	0.011	
	(0.003)	(0.020)	(0.003)	(0.018)	(0.004)	(0.007)	
S20	0.004	0.021	0.004	0.019	0.006	0.012	
	(0.003)	(0.018)	(0.003)	(0.016)	(0.004)	(0.007)	
AR(2)	D8	0.005	0.007	0.005	0.006	0.006	0.004
		(0.003)	(0.005)	(0.003)	(0.004)	(0.003)	(0.002)
	D12	0.004	0.007	0.003	0.006	0.004	0.004
		(0.002)	(0.006)	(0.002)	(0.005)	(0.002)	(0.003)
	D16	0.004	0.004	0.003	0.004	0.004	0.003
		(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)
	D20	0.003	0.006	0.003	0.005	0.004	0.004
		(0.002)	(0.005)	(0.002)	(0.004)	(0.002)	(0.002)
S8	0.005	0.009	0.005	0.009	0.004	0.006	
	(0.003)	(0.007)	(0.002)	(0.007)	(0.002)	(0.003)	
S12	0.004	0.006	0.004	0.006	0.004	0.004	
	(0.002)	(0.005)	(0.002)	(0.004)	(0.002)	(0.003)	
S16	0.004	0.005	0.003	0.004	0.004	0.003	
	(0.002)	(0.004)	(0.002)	(0.003)	(0.002)	(0.002)	
S20	0.003	0.005	0.003	0.005	0.003	0.003	
	(0.002)	(0.004)	(0.002)	(0.004)	(0.002)	(0.002)	
AR(3)	D8	0.008	0.010	0.008	0.010	0.009	0.010
		(0.004)	(0.007)	(0.004)	(0.007)	(0.006)	(0.006)
	D12	0.007	0.010	0.007	0.009	0.008	0.006
		(0.004)	(0.008)	(0.004)	(0.006)	(0.005)	(0.004)
	D16	0.007	0.009	0.007	0.009	0.009	0.009
		(0.004)	(0.007)	(0.004)	(0.006)	(0.005)	(0.004)
	D20	0.006	0.009	0.006	0.008	0.008	0.007
		(0.004)	(0.006)	(0.004)	(0.006)	(0.005)	(0.004)
S8	0.008	0.013	0.008	0.012	0.009	0.008	
	(0.004)	(0.011)	(0.004)	(0.009)	(0.005)	(0.004)	
S12	0.007	0.011	0.007	0.010	0.008	0.008	
	(0.004)	(0.009)	(0.004)	(0.007)	(0.005)	(0.004)	
S16	0.007	0.011	0.007	0.010	0.009	0.008	
	(0.004)	(0.008)	(0.004)	(0.007)	(0.005)	(0.005)	
S20	0.007	0.010	0.007	0.009	0.008	0.008	
	(0.004)	(0.008)	(0.004)	(0.007)	(0.005)	(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 order	AIC			AICc			BIC		
	Spl.	CW	WW	Spl.	CW	WW	Spl.	CW	WW
AR(1)	0.004 (0.002)	0.004 (0.003)	0.014 (0.012)	0.004 (0.002)	0.003 (0.003)	0.013 (0.011)	0.005 (0.003)	0.004 (0.003)	0.008 (0.006)
AR(2)	0.003 (0.002)	0.004 (0.002)	0.004 (0.003)	0.003 (0.002)	0.003 (0.002)	0.004 (0.003)	0.003 (0.002)	0.004 (0.002)	0.003 (0.002)
AR(3)	0.022 (0.008)	0.007 (0.004)	0.009 (0.007)	0.022 (0.008)	0.007 (0.004)	0.009 (0.006)	0.022 (0.009)	0.009 (0.005)	0.009 (0.004)

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 scales 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(1)	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)
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)	
AR(2)	D8	1.415	0.200	1.409	0.168	1.431	0.180
		(0.613)	(0.153)	(0.594)	(0.119)	(0.532)	(0.136)
	D12	1.337	0.199	1.343	0.157	1.350	0.202
		(0.582)	(0.153)	(0.568)	(0.114)	(0.515)	(0.165)
	D16	1.277	0.181	1.284	0.146	1.306	0.213
		(0.591)	(0.147)	(0.576)	(0.113)	(0.524)	(0.188)
	D20	1.285	0.186	1.287	0.144	1.276	0.209
		(0.619)	(0.143)	(0.603)	(0.102)	(0.516)	(0.174)
S8	1.356	0.209	1.355	0.165	1.383	0.178	
	(0.632)	(0.163)	(0.619)	(0.119)	(0.557)	(0.132)	
S12	1.277	0.182	1.287	0.152	1.300	0.170	
	(0.610)	(0.141)	(0.592)	(0.107)	(0.544)	(0.121)	
S16	1.269	0.175	1.254	0.145	1.271	0.180	
	(0.624)	(0.136)	(0.608)	(0.105)	(0.527)	(0.146)	
S20	1.269	0.169	1.268	0.143	1.272	0.182	
	(0.632)	(0.129)	(0.609)	(0.102)	(0.506)	(0.151)	
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 $\{2, \dots, 30\}$. 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 order	AIC			AICc			BIC		
	Spl.	CW	WW	Spl.	CW	WW	Spl.	CW	WW
AR(1)	1.042 (0.268)	0.803 (0.291)	0.248 (0.185)	1.049 (0.266)	0.806 (0.284)	0.277 (0.232)	1.112 (0.266)	0.868 (0.307)	0.581 (0.433)
AR(2)	0.880 (0.199)	1.277 (0.591)	0.181 (0.147)	0.885 (0.199)	1.284 (0.576)	0.146 (0.113)	0.909 (0.204)	1.306 (0.524)	0.213 (0.188)
AR(3)	1.031 (0.195)	0.893 (0.369)	0.219 (0.179)	1.041 (0.196)	0.892 (0.363)	0.211 (0.167)	1.084 (0.217)	0.914 (0.383)	0.449 (0.330)

Table 6: Classical wavelet-based model selections for the industrial production index according to the coarsest (j_0) and finest (J) scales 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

Table 7: Warped wavelet-based model selections for the industrial production index according to the coarsest (j_0) and finest (J) scales 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	$AICc$	r	S_r^*	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

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 spline-based results.

h	AR	CW	WW	Spline
1	0.289	0.354	0.321	0.305
2	0.301	0.372	0.332	0.345
3	0.305	0.378	0.318	0.345
4	0.294	0.318	0.290	0.307
5	0.288	0.325	0.297	0.310
6	0.287	0.315	0.284	0.298
7	0.279	0.326	0.278	0.307
8	0.284	0.298	0.267	0.290
9	0.289	0.301	0.281	0.298
10	0.288	0.280	0.274	0.288
11	0.289	0.274	0.276	0.289
12	0.273	0.262	0.267	0.289