Wavelet Estimation of Functional Coefficient Regression Models

Michel H. Montoril, Pedro A. Morettin, and Chang Chiann

Department of Statistics
University of São Paulo, Brazil
Corresponding author: Pedro A. Morettin, e-mail: pam@ime.usp.br

Abstract

In this work we study the estimation of functional coefficient regression (FCR) models using wavelets. We will present the convergence rates of the proposed estimator and carry out a simulation study to evaluate which selection criterion is helpful to provide better resolution levels in order to find the more adequate model. Moreover, we will use a real data set to make forecasts and to compare our method with others known in the literature.

Keywords: Regression, functional coefficients, father wavelets, resolution level.

1 Introduction

In this work we study functional coefficient regression (FCR) models using expansion in father wavelets to estimate the coefficient functions. 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 $\mathbb{R}^d$. Let $\mathbb{E}(Y_t^2) < \infty$. 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

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

where the $f_j(\cdot)$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.

Differently from the usual, in our study it is not necessary the assumption of independence for the errors.

2 Estimation

Any wavelet basis has an associated multiresolution analysis (MRA), which is a sequence of nested and closed spaces $\{V_j\}_{j\in\mathbb{Z}}$ of $L_2(\mathbb{R})$ satisfying certain properties. One of them states that 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 is well-known that it generates a basis $\{\varphi_{Jk}\}_k$, where $\varphi_{Jk}(\cdot) = 2^{J/2}\varphi(2^J \cdot -k)$, of the space V_J, where J is called resolution level.

Now, let J_i be a resolution level associated to the coefficient function f_i, and, for sake of simplicity, denote $\phi_{ik}(\cdot) = 2^{J_i/2}\varphi_{(i)}(2^{J_i} \cdot -k)$. Thus, following the idea of Huang and Shen

(2004), it is possible to approximate the each coefficient function by an orthogonal projection in a multiresolution space V_{J_i} and, then, approximate (1) by

$$m(\boldsymbol{x}, u) \approx \sum_{i=1}^{d} \sum_{k} \alpha_{ik} \phi_{ik}(u) x_i, \tag{2}$$

where $\boldsymbol{x} = (x_1, \ldots, x_d)^\top$. We write that k starts from 1 just to simplify the notation.

If the coefficient functions and the father wavelet have compact support, there are just a finite number r_i, $i = 1, \ldots, d$, of wavelet coefficients different from zero. Thus, it is possible to estimate the wavelet coefficients of wavelets and then estimate the functions f_j of the model (1) by

$$\hat{f}_j(u) = \sum_{k=1}^{r_i} \hat{\alpha}_{ik} \phi_{ik}(u),$$

where $\hat{\boldsymbol{\alpha}}_i = (\hat{\alpha}_{i1}, \ldots, \hat{\alpha}_{ir_i})^\top$, $i = 1, 2, \ldots, d$, is the estimator of $\boldsymbol{\alpha}_j$.

Thus, denoting the covariance matrix of the errors by $\boldsymbol{\Sigma}$, and supposing initially that it is known, one can estimate the wavelet coefficients vector minimizing the least squares function

$$\ell(\boldsymbol{\alpha}) = (\boldsymbol{Y} - \quad \boldsymbol{\alpha})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{Y} - \quad \boldsymbol{\alpha}), \tag{3}$$

where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^\top, \ldots, \boldsymbol{\alpha}_d^\top)^\top$, $\boldsymbol{\alpha}_i = (\alpha_{i1}, \ldots, \alpha_{ir_i})^\top$, $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ and the t-th row of corresponds to the vector $\phi_{ik}(U_t) X_{tj}$, $k = 1, 2, \ldots, r_i$, $i = 1, \ldots, d$. Hence, the coefficient vector estimator is given by

$$\hat{\boldsymbol{\alpha}} = (\quad ^\top \boldsymbol{\Sigma}^{-1} \quad)^{-1} \quad ^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{Y}. \tag{4}$$

Note that when the errors are independent, $\boldsymbol{\Sigma}$ is a identity matrix. With assumptions similar to those used by Huang and Shen (2004), we derive rates of convergence for distances between the estimators and the real functions, which are presented bellow as a theorem. It is worth to explain that the notation $\asymp$, used in assumption (O3), means same rate of convergence.

Assumptions.

(O0) The eigenvalues of $\boldsymbol{\Sigma}$ are bounded away from zero and infinity;

(O1) The marginal density of U_t is bounded away from zero and infinity uniformly on $\mathcal{C}$;

(O2) 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 $u \in \mathcal{C}$;

(O3) $2^{J_i} \asymp n^r$, $0 < r < 1$, $j = 1, \ldots, d$;

(O4) The process $\{Y_t, \boldsymbol{X}_t, U_t\}_{t \in \mathbb{Z}}$ is jointly strictly stationary. The α-mixing coefficient $\alpha(t)$ of
$\{Y_t, \boldsymbol{X}_t, U_t\}_{t \in \mathbb{Z}}$ satisfies $\alpha(t) \leq C t^{-\alpha}$ for $\alpha > (2 + r)/(1 - r)$;

(O5) For some sufficient large $m > 0$, $\mathbb{E}|X_{ti}|^m < \infty$, $j = 1, \ldots, d$.

Since $f_i^{J_i}$ is the orthogonal projection of f_i in V_{J_i}, denote $\rho_i = \|f_i^{J_i} - f_i\|$.

Theorem 1 *If the assumptions above hold, then*

$$\sum_{i=1}^{d} \mathbb{E}\|\hat{f}_i - f_i\|_2^2 \leq C \sum_{i=1}^{d} \left(\frac{2^{J_i}}{n} + \rho_i^2 \right),$$

for some $C > 0$. In particular, if $\rho_i = o(1)$, then $\mathbb{E}\|\hat{f}_i - f_i\|_2^2 = o(1)$, $i = 1, \ldots, d$.

As in practical situations the covariance matrix $\boldsymbol{\Sigma}$ is unknown, it has to be estimated (e.g., $\hat{\boldsymbol{\Sigma}}$), and with such estimator, the wavelet coefficients can be computed as

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

If the estimator of the covariance matrix is consistent in probability, in the sense that all eigenvalues of $\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma} - \boldsymbol{I}$ are $o_p(1)$, with $\boldsymbol{I}$ being a identity matrix, it is possible to find that

$$|\tilde{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}}|^2 = o_p(1). \tag{6}$$

Thus, denoting $\tilde{f}_i(u) = \sum_{k=1}^{r_i} \tilde{\alpha}_{ik}\phi_{ik}(u)$, $i = 1, \ldots, d$, based on (6), we can derive the following result.

Proposition 1 *If assumptions (O0) – (O5) hold, with $\hat{\boldsymbol{\Sigma}}$ consistent in probability in estimating $\boldsymbol{\Sigma}$, then*

$$\sum_{i=1}^{d} \|\tilde{f}_i - f_i\|_2^2 = O_p\left(\sum_{i=1}^{d} \left(\frac{2^{J_i}}{n} + \rho_i^2 \right) \right).$$

In particular, if $\rho_i = o(1)$, then $\tilde{f}_i$ is consistent in probability in estimating f_i, i.e., $\|\tilde{f}_i - f_i\|_2 = o_p(1)$, $j = 1, \ldots, d$.

Now, it is possible to find a consistent estimator $\hat{\boldsymbol{\Sigma}}$ supposing that $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$, i.e., the covariance matrix is a function of a parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^\top$, where p is a fixed number. In general, one can suppose that the errors of the model are represented by autoregressive processes AR(p). Thus, once the coefficient function estimators are consistent, the residuals can be considered good predictors of errors. Then it is possible to find a consistent estimator to $\boldsymbol{\theta}$, say $\hat{\boldsymbol{\theta}}$, and hence obtain a consistent estimator for the covariance matrix, $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$.

Borrowing ideas of Cochrane and Orcutt (1949), we can proceed the estimation iteratively. Firstly, the wavelet coefficients vector can be estimated acting as if the errors were independent ($\boldsymbol{\Sigma} = \boldsymbol{I}$) and then computing the residuals. Next, one can fit an autoregressive model to the residuals and by using the estimate of the autoregressive coefficients, the covariance matrix can be estimated. In the following, the wavelet coefficients vector $\tilde{\boldsymbol{\alpha}}$ could be computed by (5), with the estimate of covariance matrix. This double stage procedure (computation of $\hat{\boldsymbol{\Sigma}}$ and $\tilde{\boldsymbol{\alpha}}$) can be repeated until, for example, the convergence of the residual mean square is achieved.

Another procedure, that we will use in this work, is the following. Denoting by $\boldsymbol{\eta}$ the vector $(\boldsymbol{\alpha}^\top, \boldsymbol{\theta}^\top)^\top$ and $\quad_t$ as the t-th row of , we estimate jointly the coefficients of the FCR model $\boldsymbol{\alpha}$ and the autoregressive coefficients $\boldsymbol{\theta}$ minimizing numerically

$$\ell(\boldsymbol{\eta}) = \sum_{t=1}^{n} \left\{ \theta_p(L) \left(Y_t - \quad ^\top_t \boldsymbol{\alpha} \right) \right\}^2, \tag{7}$$

where $\theta_p(L) = 1 - \theta_1 L - \ldots - \theta_p L^p$ and the backshift 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 for estimating the coefficient vector

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

(2) Fit an autoregressive model to residuals of step (1), i.e., $\tilde{\epsilon}_t = Y_t - {}_t^{\top}\tilde{\boldsymbol{\alpha}}$, say,

$$\tilde{\theta}_p(L)\tilde{\epsilon}_t = \tilde{\varepsilon}_t;$$

(3) Estimate $\boldsymbol{\eta}$ numerically, minimizing (7), using the estimates in steps (1) and (2) as initial values.

2.1 Selection of the resolution level J

In this estimation procedure it is important to choose an adequate resolution level J. In this work, we proceed similarly to Huang and Shen (2004). We will use the information criteria AIC, AICc and BIC. Denoting the sample size by n, the number of parameters to be estimated by p and the residual mean square by RMS, these criteria are can be defined as

$$\text{AIC} = \log(\text{RMS}) + \frac{2p}{n}, \quad \text{AICc} = \text{AIC} + \frac{2(p+1)(p+2)}{n(n-p-2)} \quad \text{e} \quad \text{BIC} = \log(\text{RMS}) + \frac{p}{2}\log(n).$$

3 Simulation study

We ran a simulation study to verify which of the three criteria presented above is performing better, in the resolution level selection. The square-root of average squared error (RASE) was used, which is defined as

$$\text{RASE}^2 = \sum_{i=1}^{d} \text{RASE}_i^2, \quad \text{with} \quad \text{RASE}_i = \left\{ n_{\text{grid}}^{-1} \sum_{k=1}^{n_{\text{grid}}} \left[\hat{f}_i(u_k) - f_i(u_k) \right]^2 \right\}^{1/2},$$

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

The model simulated corresponds to the the EXPAR (Cai et al., 2000; Huang and Shen, 2004)

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

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 studied the cases where the ϵ_t's are iid $N(0; 0.2^2)$ and where they can be represented by an AR(1) model, with coefficient $\theta = 0.6$ and white noise iid $N(0; 0.16^2)$. We replicated 10,000 series with length 400. Daublets and Symmlets basis were considered in estimating the coefficient functions f_1 and f_2. For sake of simplicity, we have used the same wavelet basis in the estimation.

The candidates to be resolution levels were selected for each replicate between the values 1, 2 and 3, using the criteria AIC, AICc and BIC. For sake of simplicity, the same resolution level was considered in estimating f_1 and f_2. The mean (standard error) of RASE^2, for each criterion, is shown in Table 1, for independent errors, and Table 2, for correlated errors. In both tables it is possible to see that BIC is providing more interesting results.

Table 1: *Sample mean (SE) of RASE^2, for Daublets and Symmlets basis, under assumption of independence for the errors.*

Wavelet	AIC		AICc		BIC	
	mean	SE	mean	SE	mean	SE
$D4$	0.0247	0.00020	0.0250	0.00017	0.0265	0.00007
$D10$	0.0128	0.00008	0.0126	0.00007	0.0126	0.00007
$D16$	0.0164	0.00011	0.0163	0.00011	0.0163	0.00011
$S5$	0.0104	0.00007	0.0102	0.00007	0.0101	0.00006
$S7$	0.0146	0.00009	0.0145	0.00009	0.0144	0.00009
$S9$	0.0149	0.00010	0.0148	0.00010	0.0147	0.00010

Table 2: *Sample mean (SE) of RASE^2, for Daublets and Symmlets basis, under assumption of autoregressive errors.*

Wavelet	AIC		AICc		BIC	
	mean	SE	mean	SE	mean	SE
$D4$	0.0213	0.00023	0.0199	0.00016	0.0262	0.00011
$D10$	0.0157	0.00013	0.0153	0.00013	0.0152	0.00012
$D16$	0.0191	0.00017	0.0188	0.00016	0.0187	0.00016
$S5$	0.0137	0.00015	0.0130	0.00013	0.0129	0.00012
$S7$	0.0187	0.00016	0.0185	0.00016	0.0185	0.00016
$S9$	0.0179	0.00016	0.0175	0.00015	0.0174	0.00015

3.1 Application to Industrial Production Index

In this section we analyze the returns of the USA monthly industrial production index (IPI), from February 1984 to December 2007. The aim here is compute forecasts and compare the absolute prediction error (APE) to other well-known models. Our forecasts are computed with a similar idea of Huang and Shen (2004). The difference is that the model fitted by wavelets pointed to the existence of correlated errors. Thus, we proceeded bootstrapping the white noise residuals, instead of just the residuals, and then, using the autoregressive estimates, we were able to replicate the residuals of the model.

In order to make comparisons, other two models were fitted: an autoregressive model, by ordinary least squares, and a FAR model by splines (Huang and Shen, 2004). All the models were fitted with data just until December 2006, saving the year of 2007 to compute the APEs of the multi-step-ahead forecasts. Stepwise methods (as in Huang and Shen, 2004) were used to select the order of the model, where the candidates to be the threshold lag and significant lags were chose in a range from 1 to 4.

In the first model, the AIC criterion suggested an AR(3). The fitted model is $\hat{Y}_t = 0.143 + 0.019Y_{t-1} + 0.168Y_{t-2} + 0.202Y_{t-3}$.

The second model was fitted by quadratic splines. The AIC criterion (recommended by Huang and Shen (2004)) used to select the number of knots (between 2 and 6 – the selected number was 3) and the order of the model, suggested the FAR model

$$Y_t = f_2(Y_{t-1})Y_{t-2} + f_3(Y_{t-1})Y_{t-3} + \epsilon_t, \tag{8}$$

where the errors are supposed to be independent by a residual analysis (ACF, PACF and Ljung-Box tests).

Finally, using the $S4$ Symmlet basis, with the BIC criterion being used to select the resolution level (between 1 and 4 – the selected number was 1) and the order of the model, the suggested the FAR is

$$Y_t = f(Y_{t-2})Y_{t-2} + \epsilon_t, \tag{9}$$

with

$$\epsilon_t = -0.070\epsilon_{t-1} - 0.219\epsilon_{t-2} + 0.168\epsilon_{t-3} + \varepsilon_t,$$

in the final fitted model, where ε_t is a white noise.

In order to compare the APEs of these three models, we used the AR model as benchmark. The results are presented in Table 3, where it is possible to see that the nonlinear models (8) and (9) are forecasting more accurately than the AR model. One can also note that the model (9) is providing the best results, specially in the early months.

Table 3: *APE of the American IPI for the year of 2007. The second row is the APE of the AR model. The tird and fourth rows are the APE of the model (8) and (9), respectively, using the APE of the AR model as benchmark.*

Month	1	2	3	4	5	6	7	8	9	10	11	12
APE AR	0.49	0.84	0.27	0.47	0.24	0.41	0.00	0.28	0.16	0.95	0.15	0.15
APE (8)/AR	0.72	1.02	0.67	1.34	0.56	0.59	169.83	0.40	2.15	0.81	2.17	0.28
APE (9)/AR	0.62	0.76	0.04	0.80	0.59	0.94	59.73	0.88	1.31	0.96	1.33	0.64

4 Remarks

It is worth to mention that, in the simulation study, we also compared our results with the results presented in Huang and Shen (2004), in the case of independent errors. It was observed that our approach outperforms the method with splines.

In order to compute the multi-step-ahead forecasts for the models (8) and (9), we replicated such forecasts 10,000 times. The point forecasts were computed with the sample mean of these replicates. Such replicates are useful to compute interval forecasts. All the forecasts belonged to the 95% forecast intervals in both models.

Acknowledgments

The authors are thankful to FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) for the financial support (Proc. 2008/51097-6 and 2009/09588-5).

Bibliography

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.

Cochrane, D. and G. H. Orcutt (1949). Application of least squares regression to relationships containing auto- correlated error terms. *Journal of the American Statistical Association 44*(245), 32–61.

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.