Pak. J. Statist.
2009 Vol. 25(1), 37-46

MULTIVARIATE STOCHASTIC REGRESSION ESTIMATION BY WAVELET METHODS FOR STATIONARY TIME SERIES

Hassan Doosti and **Hosein Ali Niroumand**
Department of Statistics, School of Mathematical Sciences,
Ferdowsi University of Mashhad, Mashhad, Iran
Email: doosti@math.um.ac.ir and nirumand@wali.um.ac.ir

ABSTRACT

The estimation of multivariate stochastic regression $Y = r(X_1, X_2, ...X_d) + \varepsilon;\ d \geq 1$, for a stationary random process $\{X_i\}$ using wavelet methods is considered. Uniform rates of almost sure convergence over compact subsets of $\Re^d$ in the Besov space B_{spq} are established for strongly mixing processes. Also, considering the case $d = 1$, the results given by Doosti et al. (2008) are obtained as special cases.

KEYWORDS

Wavelet method, Besov Spaces; Rates of Strong Convergence; Strongly Mixing Processes; Nonparametric Curve Estimation; Probability Density Estimation

1. INTRODUCTION

Nonparametric regression is a smoothing method for recovering a regression function from data, without having a strong *priori* restriction on its form. There are many interesting examples where applications of regression smoothing methods have yielded analysis essentially unobtainable by other techniques; see for example, the monographs by Eubank (1988) and Müller (1988). Nonparametric curve estimation by wavelets has been treated in numerous articles in various setups. These range from the simple Gaussian *iid* error situation to more complicated data structures that often call for a specific algorithm tailored to the problem at hand. In the fixed design case, and for sample sizes that are a power of 2, wavelet methods offer an appealing method for adaptation of nonparametric curve estimation (Antoniadis, 1994; Antoniadis et al. 1994, 1997; Donoho et al. 1995; Härdle et al. 1998). These methods are prominent because of their computational ease and because they lead to minimax results over very broad classes of function spaces for a variety of loss functions. Delouille et al. (2001, 2004) treat nonparametric stochastic regression using smooth design-adapted wavelets built by means of the lifting scheme. Kohler (2008) introduced a new multivariate regression estimate. It is constructed by hard thresholding of estimates of coefficients of a series expansion of the regression function. Doosti et al. (2008) extended the results of Antoniadis *et al.* (1994, 1997) for mixing sequences of variables. The object of this article is to extend the results of Doosti et al. (2008). This paper consider the estimation of the multivariate stochastic regression $Y = r\left(X_1, X_2, ...X_d\right) + \varepsilon;\ d \geq 1$, for a stationary random process $\{X_i\}$ using wavelet methods. To prove our main results, we follow the methods Doosti *et al.* (2008) and Masry (1997). The organization of the article is as follow: Basic properties of wavelets and multiresolution analysis and Besov space on $\Re^d$, needed in this paper, are presented in section 2. After introducing wavelet multivariate density function estimation given in section 3, we introduce our proposed estimator in section 4 and study its asymptotic properties.

2. WAVELET AND BESOV SPACES

Following Meyer (1992) a multiresolution analysis on the Euclidean space $\Re^d$ is a decomposition of the space $L_2(\Re^d)$ into an increasing sequence of closed subspaces

$$\ldots V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \ldots$$

such that

$$f(2x) \in V_{j+1} \Leftrightarrow f(x) \in V_j \quad \textit{for all } j$$

$$\bigcap_j V_j = 0, \qquad \overline{\bigcup_j V_j} = L_2(\Re^d).$$

V_0 is closed under integer translation. Finally, there exists a scale function $\varphi \in L_2(\Re^d)$ with $\int_{\Re^d} \varphi(\underline{x})\, d\underline{x} = 1$ such that $\{\varphi_{\underline{k}}(\underline{x}) = \varphi(\underline{x} - \underline{k}), \underline{k} \in Z^d\}$ is an orthonormal basis for V_0. It follows that $\{\varphi_{j,\underline{k}}(\underline{x}) = 2^{jd/2}\varphi(2^j \underline{x} - \underline{k}), \underline{k} \in Z^d\}$ is an orthonormal basis for V_j.

Definition. The multiresolution analysis is called r-regular if $\varphi \in L_2(\Re^d)$ and all its partial derivatives up to total order r are rapidly decreasing, i.e., for every integer $i \geq 0$, there exists a constant A_i such that

$$\left|\left(D^{\underline{\beta}}\varphi)(\underline{x}\right)\right| \leq \frac{A_i}{\left(1 + \|\underline{x}\|\right)^i} \quad \textit{for all } |\underline{\beta}| \leq r, \tag{2.1}$$

where

$$\left(D^{\beta}\varphi\right)(\underline{x}) = \frac{\partial^{\underline{\beta}}\varphi(\underline{x})}{\partial x_1^{\beta_1} \ldots \partial x_d^{\beta_d}}, \tag{2.2}$$

and

$$\underline{\beta} = (\beta_1, \ldots, \beta_d), \quad |\underline{\beta}| = \sum_{i=1}^{d} \beta_i. \tag{2.3}$$

Define the detail space W_j by $V_{j+1} = V_j \oplus W_j$.

Then there exist $N = 2^d - 1$ wavelet functions $\{\psi_i(\underline{x}), i = 1, \dots, N\}$ such that

i) $\{\psi_i(\underline{x} - \underline{k}), \underline{k} \in Z^d, \; i = 1, \dots, N\}$ is an orthonormal basis for W_0.

ii) With $\left\{\psi_{i,j,\underline{k}}(\underline{x}) = 2^{jd/2}\psi_i\left(2^j \underline{x} - \underline{k}\right)\right\}$ the functions $\left\{\psi_{i,j,\underline{k}}(\underline{x}), \; i = 1, \dots, N, \underline{k} \in Z^d, \; j \in Z\right\}$ constitute an orthonormal basis for $L_2(\Re^d)$.

iii) ψ_i has the same regularity as φ.

For any $f \in L_2\left(\Re^d\right)$ we have the orthonormal representation

$$f(\underline{x}) = \sum_{\underline{k} \in Z^d} a_{m\underline{k}} \varphi_{m,\underline{k}}(\underline{x}) + \sum_{j \ge m} \sum_{i=1}^{N} \sum_{\underline{k} \in Z^d} b_{ij\underline{k}} \psi_{i,j,\underline{k}}(\underline{x}) \tag{2.4}$$

for any integer *m* where

$$a_{m\underline{k}} = \int_{\Re^d} f(\underline{u}) \varphi_{m,\underline{k}}(\underline{u}) d\underline{u}, \qquad b_{ij\underline{k}} = \int_{\Re^d} f(\underline{u}) \Psi_{i,j,\underline{k}}(\underline{u}) d\underline{u}. \tag{2.5}$$

Note that the orthogonal projection of f on V_l can be written in two equivalent ways:

$$\begin{aligned} \left(P_{V_l} f\right)(\underline{x}) &= \sum_{\underline{k} \in Z^d} a_{l\underline{k}} \varphi_{l,\underline{k}}(\underline{x}) \\ &= \sum_{\underline{k} \in Z^d} a_{m\underline{k}} \varphi_{m,\underline{k}}(\underline{x}) + \sum_{j=m}^{l} \sum_{i=1}^{N} \sum_{\underline{k} \in Z^d} b_{ij\underline{k}} \Psi_{i,j,\underline{k}}(\underline{x}) \end{aligned} \tag{2.6}$$

for any $m \le l$. We suppose that f belongs to the Besov class, B_{spq}. Meyer (1992) provided a characterization of the space B_{spq} in terms of wavelets coefficients. Assume the multiresolution analysis is r- regular and $s < r$, then $f \in B_{spq}$ if and only if

$$J_{spq}(f) = \left\|P_{V_o} f\right\|_{L_p} + \left(\sum_{j \ge 0} \left(2^{js} \left\|P_{W_j} f\right\|_{L_p}\right)^q\right)^{1/q} < \infty.$$

3. MULTIVARIATE DENSITY ESTIMATION

Let $\{X_i\}$ be a stationary random process. Set $\underline{X}_j = \left(X_{j+1}, \dots, X_{j+d}\right)$ and let $f(\underline{x})$ be the joint probability density of the vector $\underline{X}_j$. Assume that $f \in L_2\left(\Re^d\right)$, Then $f(\underline{x})$ admits the wavelet representation (2.4). Given *n* observations $\{\underline{X}_i\}_{i=1}^{n}$ we estimate the coefficients $\{a_{m\underline{k}}\}$ and $\{b_{ij\underline{k}}\}$ by

$$\hat{a}_{m\underline{k}} = \frac{1}{n}\sum_{i=1}^{n}\varphi_{m,\underline{k}}\left(\underline{X}_i\right), \qquad \hat{b}_{ij\underline{k}} = \frac{1}{n}\sum_{t=1}^{n}\psi_{i,j,\underline{k}}\left(\underline{X}_t\right), \tag{3.1}$$

and note that these estimates are unbiased,

$$\mathrm{E}\left[\hat{a}_{m\underline{k}}\right] = a_{m\underline{k}}, \qquad \mathrm{E}\left[\hat{b}_{ij\underline{k}}\right] = b_{ij\underline{k}}.$$

A linear estimate of f can be obtained from (2.4) by

$$\hat{f}_n\left(\underline{x}\right) = \sum_{\underline{k}\in Z^d}\hat{a}_{m\underline{k}}\varphi_{m,\underline{k}}\left(\underline{x}\right) \tag{3.2}$$

or, equivalently, as

$$\hat{f}_n\left(\underline{x}\right) = \sum_{\underline{k}\in Z^d}\hat{a}_{m\underline{k}}\varphi_{m,\underline{k}}\left(\underline{x}\right) + \sum_{j_0\le j\le m}\sum_{i=1}^{N}\sum_{\underline{k}\in Z^d}\hat{b}_{ij\underline{k}}\psi_{i,j,\underline{k}}\left(\underline{x}\right) \tag{3.3}$$

for any $j_0 \le m$. Here the resolution level $m = m(n) \to \infty$ at a rate specified below. We assume that φ and ψ_i have a compact support so that the summations above are finite for each fixed $\underline{x}$ (note that in this case the support of φ and ψ_i is a monotonically increasing function of their degree of differentiability (Daubechies, 1992)). Masry (1997) established rates of strong convergence which are uniform over compact subsets of $\Re^d$.

Let F_i^k be the σ-algebra of events generated by the random variables $\left\{\underline{X}_j, i \le j \le k\right\}$. The stationary process $\left\{\underline{X}_j\right\}$ is called strongly mixing (Rosenblatt, 1956b) if $\sup_{A\in F_{-\infty}^0,\, B\in F_k^\infty}\left|P\left[AB\right] - P\left[A\right]P\left[B\right]\right| = \alpha\left(k\right) \to 0$ *as* $k \to \infty$.

Condition A. Assume that

i) $f(\underline{x}) \le M_1 < \infty$.

ii) $\left|f\left(\underline{x},\underline{y};j\right) - f\left(\underline{x}\right)f\left(\underline{y}\right)\right| \le M_2 < \infty$ for all $j \ge 1$, where $f\left(\underline{x},\underline{y};j\right)$ is the joint probability density of the vectors $\left(\underline{X}_0, \underline{X}_j\right)$ which is assumed to exist.

iii) The strongly mixing coefficient $\alpha(j)$ satisfies

$$\sum_{j=1}^{\infty} j^{\alpha}[\alpha(j)]^{1-2/\nu} < \infty \tag{3.4}$$

for some $\nu > 2$ and $0 < \alpha < 1 - 2/\nu$.

Note that (3.4) is equivalent to $\alpha(j)=O\left(1/j^{c}\right)$ for some c > 2. Also note that for 1 < j < d – 1, the components of $\underline{X}_0$ and of $\underline{X}_j$ overlap; the joint density $f\left(\underline{x},\underline{y};j\right)$ in condition A(ii) is then the density of the random variables $\left(X_{j+1},\ldots,X_{j+d}\right)$.

Lemma 1.

Under Condition A, there exists a constant M (which depends only on f and φ) such that

$$\sup_{\underline{x}\in\Re^{d}} \operatorname{var}[\hat{f}_n(\underline{x})] \le M\,\frac{2^{dm(n)}}{n}. \tag{3.5}$$

Lemma 2.

Assume that the multiresolution analysis is r-regular and density $f\in B_{spq}$, for some $0<s<r,\ 1\le p,\ q\le\infty,$ then for $s>d/p,$

$$E_{\underline{x}\in\Re^{d}}\,|E\hat{f}_n(\underline{x})-f(\underline{x})|=O\left(2^{-(s-d/p)m(n)}\right).$$

for proofs of above lemmas see Masry(1997).

4. MULTIVARIATE REGRESSION ESTIMATION

Consider the nonparametric multivariate regression model which is given as the following. Let $\left(X_{1i},X_{2i},\ldots,X_{di},Y_i\right),\ i=1,2,\ldots,n$ be identically distributed as a (d+1)-dimensional random vector $\left(X_1,\ldots,X_d,Y\right)$ with $E\left(Y^2\right)<\infty$. The proposed here is to estimate the multivariate regression function of Y on $\underline{X}$, denoted by $r\left(\underline{x}\right)=E\left(Y\,|\,\underline{X}=\underline{x}\right);\quad \underline{X}=\left(X_1,\ldots,X_d\right)$ and $\underline{x}=\left(x_1,\ldots,x_d\right)$. An alternative way to write the regression model is the following

$$Y_i=r\left(\underline{X}_i\right)+\varepsilon_i,\quad i=1,\ldots,n,$$

where the error ε_i, conditionally on $\underline{X}_i$, is assumed to be independent with zero expectation and a bounded conditional variance. The above setup corresponds to the random design. Also, we assumed that the common multivariate distribution of $\underline{X}_i$ admitted a multivariate density $f(x_1,\ldots,x_d)$. similar to the setup in earlier literature such as Antoniatis and Pham (1995), Antoniadis *et al.* (1997), Delouille et al. (2004) and Doosti et al. (2008), our estimator of $r(\underline{x})$ will be obtained by taking the ratio of wavelet estimators of $g=rf$ and f.

Recall that,

$$
\begin{aligned}
r(\underline{x}) &= \mathrm{E}(Y \mid \underline{X} = \underline{x}) \\
&= \frac{\int_{\Re} y f(\underline{x}, y)\, dy}{f(\underline{x})} \\
&= \frac{g(\underline{x})}{f(\underline{x})}.
\end{aligned}
$$

Similar to (2.4), if $g(\underline{x}) \in L_2(\Re^d)$, we may write its wavelet expansion as

$$g(\underline{x}) = \sum_{\underline{k}\in Z^d} a'_{m\underline{k}} \varphi_{m,\underline{k}}(\underline{x}) + \sum_{j\geq m} \sum_{i=1}^{N} \sum_{\underline{k}\in Z^d} b'_{ij\underline{k}} \psi_{i,j,\underline{k}}(\underline{x}) \tag{4.1}$$

For any integer m, where

$$a_{m\underline{k}} = \int_{\Re^d} f(\underline{u}) \varphi_{m,\underline{k}}(\underline{u})\, d\underline{u}, \qquad b_{ij\underline{k}} = \int_{\Re^d} f(\underline{u}) \Psi_{i,j,\underline{k}}(\underline{u})\, d\underline{u}. \tag{4.2}$$

Also

$$
\begin{aligned}
(P_{V_l} g)(\underline{x}) &= \sum_{\underline{k}\in Z^d} a'_{l\underline{k}} \varphi_{l,\underline{k}}(\underline{x}) \\
&= \sum_{\underline{k}\in Z^d} a'_{m\underline{k}} \varphi_{m,\underline{k}}(\underline{x}) + \sum_{j=m}^{l} \sum_{i=1}^{N} \sum_{\underline{k}\in Z^d} b'_{ij\underline{k}} \Psi_{i,j,\underline{k}}(\underline{x})
\end{aligned} \tag{4.3}
$$

for any $m \leq l$. Given n observations $\{\underline{X}_i\}_{i=1}^{n}$ we estimate the coefficients $\{a'_{m\underline{k}}\}$ and $\{b'_{ij\underline{k}}\}$ by

$$\hat{a}'_{m\underline{k}} = \frac{1}{n} \sum_{i=1}^{n} Y_i \varphi_{m,\underline{k}}(\underline{X}_i), \qquad \hat{b}'_{ij\underline{k}} = \frac{1}{n} \sum_{t=1}^{n} Y_i \psi_{i,j,\underline{k}}(\underline{X}_t), \tag{4.4}$$

and note that these estimates are unbiased,

$$\mathrm{E}\left[\hat{a}'_{m\underline{k}}\right] = a'_{m\underline{k}}, \qquad \mathrm{E}\left[\hat{b}'_{ij\underline{k}}\right] = b'_{ij\underline{k}}.$$

A linear estimate of g can be obtained by

$$\hat{g}_n(\underline{x}) = \sum_{\underline{k}\in Z^d} \hat{a}'_{m\underline{k}} \varphi_{m,\underline{k}}(\underline{x}) \tag{4.5}$$

or, equivalently, as

$$\hat{g}_n(\underline{x}) = \sum_{\underline{k}\in Z^d} \hat{a}'_{m\underline{k}} \varphi_{m,\underline{k}}(\underline{x}) + \sum_{j_0\leq j\leq m} \sum_{i=1}^{N} \sum_{\underline{k}\in Z^d} \hat{b}'_{ij\underline{k}} \psi_{i,j,\underline{k}}(\underline{x}) \tag{4.6}$$

for any $j_0 \le m$.

In the rest of this section we will prove under some conditions, the results of last section are satisfactory for $g(\underline{x})$ too, and in theorem 3 we obtain the rate of convergence for bias and variance of our proposed estimators.

In following theorem the condition A plays main role.

Theorem 1.

Let condition A hold and regression function $r(\underline{x})$ is locally bounded. Then there exists a constant C (which depends only on g and φ such that

$$\sup_{\underline{x}\in\Re^d} \operatorname{var}\left[\hat{g}_n(\underline{x})\right] \le C\frac{2^{dm(n)}}{n}$$

Proof:

Define the kernel $k(\underline{u},\underline{v})$ by

$$k(\underline{u},\underline{v}) = \sum_{\underline{k}\in Z^d} \varphi(\underline{u}-\underline{k})\varphi(\underline{v}-\underline{k}), \qquad (4.7)$$

the by (4.5) and (4.6) we can write $\hat{g}_n(\underline{x})$ as an extended kernel estimator

$$\hat{g}_n(\underline{x}) = \frac{1}{nh_n^d}\sum_{i=1}^{n} Y_i k\left(\frac{\underline{x}}{h_n}, \frac{\underline{X}_i}{h_n}\right), \quad h_n = 2^{-m(n)}. \qquad (4.8)$$

Now,

$$\begin{aligned} Var(\hat{g}_n(\underline{x}) &= Var\left[\frac{1}{nh_n^d}\sum_{i=1}^{n} Y_i k\left(\frac{\underline{x}}{h_n}, \frac{\underline{X}_i}{h_n}\right)\right] \\ &= \frac{1}{n^2 h_n^{2d}}\sum_{i=1}^{n} Var\left(Y_i k\left(\frac{\underline{x}}{h_n}, \frac{\underline{X}_i}{h_n}\right)\right) \\ &\quad + \frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \operatorname{cov}\left(Y_i k\left(\frac{\underline{x}}{h_n}, \frac{\underline{X}_i}{h_n}\right), Y_j k\left(\frac{\underline{x}}{h_n}, \frac{\underline{X}_j}{h_n}\right)\right) \\ &= T_1 + T_2 \end{aligned} \qquad (4.9)$$

Now, we want to find upper bounds for T_1 and T_2 .

By stationary:

$$T_1 = \frac{Var\left(Y_i k\left(\frac{\underline{x}}{h_n}, \frac{\underline{X}_i}{h_n}\right)\right)}{nh_n^{2d}}$$

$$\leq \frac{1}{nh_n^{2d}}\left[\mathrm{E}\left(Y_i k\left(\frac{\underline{x}}{h_n}, \frac{\underline{X}_i}{h_n}\right)\right)^2 + \mathrm{E}^2\left(Y_i k\left(\frac{\underline{x}}{h_n}, \frac{\underline{X}_i}{h_n}\right)\right)\right]$$

By Masry (1997) Eq. (4.6) and boundedness of $r(\underline{x})$;

$$T_1 \leq \frac{C_1}{nh_n^d}. \tag{4.10}$$

Next, we can write

$$T_2 = \frac{2}{nh_n^d}\sum_{j=1}^{n-1}\left(1-\frac{j}{n}\right)\mathrm{cov}\left(Y_1 k\left(\frac{\underline{x}}{h_n}, \frac{\underline{X}_1}{h_n}\right), Y_j k\left(\frac{\underline{x}}{h_n}, \frac{\underline{X}_j}{h_n}\right)\right)$$

$$\leq \frac{2}{nh_n^d}\sum_{j=1}^{n-1}\mathrm{cov}\left(Y_1 k\left(\frac{\underline{x}}{h_n}, \frac{\underline{X}_1}{h_n}\right), Y_j k\left(\frac{\underline{x}}{h_n}, \frac{\underline{X}_j}{h_n}\right)\right)$$

By Eq. (4.11) in Masry (1997) and the assumption that $r(\underline{x})$ is bounded we see

$$T_2 \leq \frac{C_2}{nh_n^d}. \tag{4.11}$$

and so by (4.8), (4.9) and (4.10), proof will be completed.□

The result of following theorem is similar to results of theorem 3.1 in Doosti et al. (2008), but the conditions of the theorems are quite different so the proof is different too.

Theorem 2. Assume that the multiresolution analysis is r-regular and density $g \in B_{spq}$, for some $0 < s < r$, $1 \leq p$, $q \leq \infty$, then for $s > d/p$,

$$\mathrm{E}_{\underline{x}\in\Re^d} \,|\, \mathrm{E}\,\hat{g}_n(\underline{x}) - g(\underline{x})| = \mathrm{O}\left(2^{-(s-d/p)m(n)}\right).$$

Proof: We have

$$\mathrm{E}\left[\hat{g}_n(\underline{x})\right] = \sum_{\underline{k}\in Z^d} a'_{m\underline{k}}\varphi_{m,\underline{k}}(\underline{x}) = \left(P_{V_0} g\right)(\underline{x})$$

Thus,

$$g(\underline{x}) - \mathrm{E}\left[\hat{g}_n(\underline{x})\right] = \sum_{j\geq m}\left(P_{W_j} g\right)(\underline{x}).$$

As in Kerkyacharian and Picard (1992) we have

$$\left\| g\left(\underline{x}\right) - \mathrm{E}\left[\hat{g}_n\left(\underline{x}\right)\right] \right\|_{L_{p'}} \le C\left(\sum_{j\ge m}\left(\left\|P_{W_j} g\right\|_{L_{p'}}\right)^q\right)^{1/q}$$

$$\le C' 2^{-m(n)s'}\left(\sum_{j\ge m}\left(2^{js'}\left\|P_{W_j} g\right\|_{L_{p'}}\right)^q\right)^{1/q}$$

$$\le C'' 2^{-m(n)s'} J_{spq}\left(g\right) = \mathrm{O}\left(2^{-m(n)s'}\right),$$

since $g \in B_{spq}$, where $s' = s - d\left(1/p - 1/p'\right)$ and the result follows with $p' = \infty$. □

These results allow one to control the convergence rate of estimator $\hat{r} = \frac{\hat{g}_n}{\hat{f}_n}$.

Theorem 3. Under assumption in lemmas 1 and 2 and Theorems 1 and 2 we have,

$$Bias\left(\hat{r}\left(\underline{x}\right)\right) = \mathrm{O}\left(\frac{2^{dm(n)}}{n}\right) + \mathrm{O}\left(2^{-(s-d/p)m(n)}\right)$$

$$Var\left(\hat{r}\left(\underline{x}\right)\right) = \mathrm{O}\left(\frac{2^{dm(n)}}{n}\right).$$

Proof: By using the results of above theorems and following proof of (3.7) and (3.8) in Doosti et al. (2008), proof will be obtained. □

REFERENCES

1. Antoniadis, A. (1994). Smoothing noisy data with coiflets. *Statistica Sinica*, 4, 651-678.
2. Antoniadis, A. and Pham, D.T. (1995). Wavelet regression for random or irregular design. Technical Report RT 148, IMAG-LMC, University of Grenoble, France.
3. Antoniadis, A., Grégoire, G. and McKeague, I. (1994). Wavelet methods for curve estimation. *J. Amer. Statist. Assoc.*, 89, 1340-1353.
4. Antoniadis, A., Gregoire, G. and Vial, P. (1997). Random design wavelet curve smoothing. *Statist. Probab. Lett.*, 35, 225-232.
5. Daubechies, I. (1992). *Ten Lectures on Wavelets*. CBMS-NSF regional conferences series in applied mathematics. Philadelphia: SIAM.
6. Delouille, V., Franke, J. and Von Sachs, R. (2001). Nonparametric stochastic regression with design-adapted wavelets. *Sankhya, Ser. A* 63, 328-366.
7. Delouille, V., Simoens, J. and Von Sachs, R. (2004). Smooth design-adapted wavelets for nonparametric stochastic regression. *J. Amer. Statist. Assoc.*, 99, 643-659.
8. Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: asymptopia (with discussion). *J. Roy. Statist. Soc., Ser. B* 57(2), 301-370.

9. Doosti, H., Afshari, M. and Niroumand, H.A. (2008). Wavelet for nonparametric stochastic regression with mixing stochastic process. *Communication in Statistics-Theory and Methods.* 37, 373-385.
10. Eubank, R.L. (1988). *Spline Smoothing and Nonparametric Regression*. New York: Marcel Dekker.
11. Härdle, W., Kerkyacharian, G., Picard, D. and Tsybabov, A. (1998). *Wavelets: Approximation and Statistical Applications.* New York: Springer-Verlag.
12. Kerkyacharian, G. and Picard, D. (1992). Density estimation in Besov spaces. *Statist. Probab. Lett.* 13, 15-24.
13. Masry, E. (1997). Multivariate probability density estimation by wavelet methods: strong consistency and rates for stationary time series. *Stochastic Processes and Their Application.* 67, 177-193.
14. Meyer, Y. (1992). *wavelets and operators.* Cambridge Univ. Press, Oxford.
15. Müller, H.G. (1988). *Nonparametric Regression Analysis of Longitudinal Data*. Vol. 46. Lecture Notes in Statistics. Berlin: Springer-Verlag.
16. Kohler, M. (2008). Multivariate orthogonal series estimates for random design regression. *J. Statist. Plann. and Inference,* In Press.