TWO-DIMENSIONAL WAVELETS FOR NONLINEAR AUTOREGRESSIVE MODELS WITH AN APPLICATION IN DYNAMICAL SYSTEM

H. Doosti

Islamic Azad University
Mashhad Branch
Department of Statistics
Mashhad
Iran
e-mail: hassandoosti1353@yahoo.com

M.S. Islam

Department of Mathematics and Statistics
University of Prince Edward Island
PE, C1E 4P3
Canada
e-mail: sislam@upei.ca

Y.P. Chaubey

Department of Mathematics and Statistics
Concordia University
1455 de Maisonneuve Blvd. West Montreal
Quebec H3G 1M8
Canada
email: chaubey@alcor.concordia.ca

P. Góra

Department of Mathematics and Statistics
Concordia University
1455 de Maisonneuve Blvd. West Montreal
Quebec H3G 1M8
Canada
e-mail: pgora@vax2.concordia.ca

Abstract. In this note we introduce a new estimator for estimating autoregressive model function based on two-dimensional wavelet expansion of joint density function. We investigate some asymptotic properties of the proposed estimator. We also added the problem of estimating of derivative of autoregressive estimator through new approach. Finally, we apply our method in dynamical systems. In particular, we estimate a chaotic map from a noisy data and filter entropy of the chaotic map.

Keywords: two-dimensional wavelet, multiresolution analysis, Random design, Besov space, wavelets.

1. Introduction

Autoregressive models form important class of processes in time series theory. A nonparametric version of these models was first introduced by Jones [21]. Let $(X, \mathcal{F}, P)$ be a probability space and $\{X_i\}_{i\geq 0}$ be a random process associated with $(X, \mathcal{F}, P)$. We observe the series $\{X_0, X_1, \ldots, X_n\}$ that follow the nonlinear autoregressive model $X_{i+1} = \tau(X_i) + \epsilon_i$, where τ is a transformation and ϵ_i is an error. For theoretical purposes, we consider *iid* perturbations ϵ_i with $\mathbb{E}(\epsilon_i) = 0$, $\mathbb{E}(\epsilon_i^2) = \sigma_i^2$, not necessarily gaussian.

Several authors dealt with the problem of estimating the autoregressive function τ nonparametrically. See Frank et al.[15], Hardle and Tsybakov [19], Robinson [27], Masry and Tjosthein [22], Hafner [18], Tjosthein [31], Buhlmann and McNeil [4], Delouille et al. [7], Delouille and Von Sachs [8] and Delouille et al. [9]. However, very little is known about 'wavelet' estimation for autoregressive designs. The result of Hoffmann [20] that treats autoregressive models using a wavelet estimator is concerned with asymptotical results only, and does not provide an efficient algorithm in practice. Delouille et al. [10] estimated nonlinear autoregressive models using a design-adapted wavelet estimator. We closely follow the work of Doosti et al. [13]. In their note [13], they dealt with two-dimensional wavelets for stochastic regression.

The organization of the paper is as follows. After some preliminaries given in Section 2, we introduce our proposed estimators in Section 3. Asymptotic properties of our proposed estimator are discussed in Section 4. Derivative of of wavelet estimator is found in Section 5. In Section 6, we apply our wavelet estimator to dynamical systems; we estimate a chaotic map from noisy data and entropy for this chaotic map.

2. Preliminaries

In this section, we first introduce one-dimensional wavelet density function estimator and then we introduce multiresolution analysis in two-dimensional case for joint density function estimation.

2.1. Wavelet linear density estimator

Let $\{X_i\}_{i\geq 0}$ be a sequence of real-valued random variables on the probability space $(X, \mathcal{F}, P)$. We suppose that X_i has a bounded and compactly supported marginal density f with respect to Lebesgue measure which does not depend on i. We estimate this density from N observation X_i, $i = 1, ..., N$. For any function $f \in \mathbb{L}_2(\mathbb{R})$, we can write a formal expansion (see Daubechies [5]):

$$f = \sum_{k\in Z} \alpha_{j_0,k}\phi_{j_0,k} + \sum_{j\geq j_0}\sum_{k\in Z} \delta_{j,k}\psi_{j,k} = P_{j_0}f + \sum_{j\geq j_0} D_j f,$$

where the functions

$$\phi_{j0,k}(x) = 2^{j_0/2}\phi(2^{j_0}x - k) \quad \text{and} \quad \psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$$

constitute an (inhomogeneous) orthonormal basis of $\mathbb{L}_2(\mathbb{R})$ and $\phi(x)$ and $\psi(x)$ are the scale function and the orthogonal wavelet respectively. Wavelet coefficients are given by the integrals

$$\alpha_{j_0,k} = \int f(x)\phi_{j_0,k}(x)dx, \quad \delta_{j,k} = \int f(x)\psi_{j,k}dx.$$

We suppose that both ϕ and $\psi \in \mathbb{C}^{r+1}$, $r \in \mathbb{N}$, and have compact supports in $[-\delta, \delta]$. Note that, by Corollary 5.5.2 in (Daubechies [6]), ψ is orthogonal to polynomials of degree $\leq r$, i.e., $\int \psi(x)x^l dx = 0$, $l = 0, 1, ..., r$. We suppose that f belongs to the Besov class (see Meyer [23], section VI.10)

$$F_{s,p,q} = \{f \in B^s_{p,q}, \|f\|_{B^s_{p,q}} \leq M\}$$

for some $0 \leq s \leq r+1$, $p \geq 1$, $q \geq 1$, where $\|f\|_{B^s_{p,q}} = \|P_{j_0} f\|_p + \Big(\sum_{j \geq j_0} (\|D_j f\|_p 2^{js})^q\Big)^{1/q}$.

We say that $f \in B^s_{p,q}$ if and only if

$$\|\alpha_{j_0,.}\|_{l_p} < \infty, \quad \text{and} \quad \left(\sum_{\mathrm{j} \geq \mathrm{j_0}} (\|\delta_{\mathrm{j},.}\|_{\mathrm{l_p}} 2^{\mathrm{j(s+1/2-1/p)}})^{\mathrm{q}}\right)^{1/\mathrm{q}} < \infty, \tag{2.1}$$

where $\|\gamma_{j,.}\|_{l_p} = \Big(\sum_{k \in Z} \gamma^p_{j,k}\Big)^{1/p}$. We consider Besov spaces essentially because of their executional expressive power (see Triebel [33]) and the discussion in Donoho et al. [12]. We construct the density estimator

$$\hat{f}_1 = \sum_{k \in K_{j_0}} \hat{\alpha}_{j_0,k} \phi_{j_0,k}, \quad \text{with} \quad \hat{\alpha}_{\mathrm{j_0,k}} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i=1}}^{\mathrm{n}} \phi_{\mathrm{j_0,k}}(\mathrm{X_i}), \tag{2.2}$$

where K_{j_0} is the set of k such that supp(f) $\cap$ supp $\phi_{\mathrm{j_0,k}} \neq \emptyset$. The fact that ϕ has a compact support implies that K_{j_0} is finite and card $K_{j_0} = O(2^{j_0})$. Wavelet density estimators aroused much interest in the recent literature (see Donoho et al. [1] and Doukhan and Leon [14]). In the case of independent samples, the properties of the linear estimator in (2.2) was studied for a variety of error measures and density classes by Kerkyacharian and Picard [24] and Tribouley [32]). It was shown, for example, that these estimators are minimax when the L_p-risk is concerned and the density belongs to Besov space $B^s_{p,q}$. When the error of estimation is measures in $L_{\acute{p}} - norm$, with $\acute{p} \geq p$, the linear wavelet estimators are not optimal any more, although they are still minimax in the class of linear estimators (see Donoho et al. [11]), Kerkyacharian and Picard [24]). In "weak dependent" cases, Leblanc [16] investigated some asymptotic property of estimator (2.2). The estimator in (2.2) is a special case of a kernel density estimator with kernel $K(x, y) = \sum_k \phi_{j_0,k}(x)\phi_{j_0,k}(y)$. In terms of kernel, (2.2) can be expressed as

$$\hat{f}_1(x) = \frac{1}{n} \sum_{i=1}^{n} K_{j_0}(x, X_i),$$

where the orthogonal projection kernels are $K_{j_0}(x, y) = 2^{j_0} K(2^{j_0}x, 2^{j_0}y)$. Huang [20] studied asymptotic bias and variance of linear wavelet density estimation. Define

$$b_m(x) = x^m - \int_{-\infty}^{\infty} K(x, y)y^m dy.$$

The functions $b_m(x)$ are important in expressing the asymptotic bias of linear estimators and finding their efficiencies with respect to the standard kernel density estimators. Theorem 2.1 below gives the bias for our linear density function estimator 2.2.

Theorem 2.1 (Huang [20]) *Assume that the density f belongs to the Holder space* $\mathbb{C}^{m+\alpha}$, $0 \leq \alpha \leq 1$, *and the wavelet-kernel* $K(x,y)$ *satisfies the following localization property* $\int_{-\infty}^{\infty} K(x,y)(y-x)^{m+\alpha}dy \leq C$, *for some positive C. Let* $j \to \infty$ *and* $n2^{-j} \to \infty$, *as* $n \to \infty$. *Then, for fixed x,*

$$\mathbb{E}\hat{f}_1(x) - f(x) = -\frac{1}{m!}f^{(m)}(x)b_m(2^jx)2^{-mj} + O(2^{-j(m+\alpha)}).$$

The asymptotic variance of $\hat{f}_1$ is given in Theorem 2.2 below. This theorem is a generalization of a theorem proved by Huang [20].

Theorem 2.2 *Let* $f \in \mathbb{C}^1$, $\acute{f}$ *be the first derivative of f, f and* $\acute{f}$ *be uniformly bounded and the mixing rate* α *satisfies* $\sum_{k=1}^{\infty}\alpha(k) < \infty$. *Then, for x fixed,*

$$Var\hat{f}(x) = \frac{2^j}{n}f(x)V(2^jx) + O(n^{-1}),$$

where $V(x) = \int_{-\infty}^{\infty} K^2(x,y)dy = K(x,x)$.

Proof.

$$\begin{aligned} Var\hat{f}(x) &= Var\left\{\frac{1}{n}\sum_{i=1}^{n} K_h(x,X_i)\right\} = \frac{1}{n^2}\sum_{i=1}^{n} Var\{K_h(x,X_i)\} \\ &+ \frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} Cov(K_h(x,X_i),K_h(x,X_j)) = T_1 + T_2. \end{aligned} \tag{2.3}$$

Now, we have

$$\begin{aligned} T_1 &= \frac{1}{n}\int_{-\infty}^{\infty} K_h^2(x,y)f(y)dy - \frac{1}{n}\left(\int_{-\infty}^{\infty} K_h(x,y)f(y)dy\right)^2 \\ &= \frac{1}{n}f(x)\int_{-\infty}^{\infty} K_h^2(x,y)dy + \frac{1}{n}\int_{-\infty}^{\infty} K_h^2(x,y)(f(y)-f(x))dy \\ &\quad -\frac{1}{n}\left(\int_{-\infty}^{\infty} K_h(x,y)f(y)dy\right)^2 \\ &= \frac{1}{nh}f(x)V(x/h) + \frac{1}{n}\int_{-\infty}^{\infty} K_h^2(x,y)(f(y)-f(x))dy \\ &\quad -\frac{1}{n}\left(\int_{-\infty}^{\infty} K_h(x,y)f(y)dy\right)^2. \end{aligned}$$

Below, we show that the second and the third terms in the last equality are of order $O(n^{-1})$

$$|\frac{1}{n}\int_{-\infty}^{\infty} K_h^2(x,y)(f(y)-f(x))dy| \le \frac{1}{n}\sup_x |\acute{f}(x)|\frac{1}{h^2}\int_{-\infty}^{\infty} K^2(x/h,y/h)|y-x|dy$$
$$\le \frac{1}{n}\sup_x |\acute{f}(x)| \sup_{s,t\in\mathbb{R}} |K(s,t)| \int_{-\infty}^{\infty} |K(x/h,t)(t-x/h)|dt = O(n^{-1}).$$

By the uniform boundedness of $f(x)$, it is easy to see that

$$\frac{1}{n}\left(\int_{-\infty}^{\infty} K_h(x,y)f(y)dy\right)^2 = O(\frac{1}{n}).$$

Thus,

$$T_1 = \frac{2^j}{n} f(x)V(2^j x) + O(n^{-1}) \tag{2.4}$$

To complete the proof, it is enough to prove $T_2 = O(n^{-1})$. Now,

$$Cov(K_h(x,X_i),K_h(x,X_j)) = \mathbb{E}(K_h(x,X_i)K_h(x,X_j)) - \mathbb{E}K_h(x,X_i)\mathbb{E}K_h(x,X_j)$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_h(x,y)K_h(x,z)f_{X,Y}(y,z)dydz$$
$$-\left(\int_{-\infty}^{\infty} K_h(x,y)f(y)dy\right)^2 \le \left(\int_{-\infty}^{\infty} K_h(x,y)dy\right)^2 \alpha(j-i).$$

Thus

$$T_2 \le \frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \phi(j-i) = \frac{2}{n}\sum_{k=1}^{n}(1-k/n)\alpha(k) \le \frac{2}{n}\sum_{k=1}^{n}\alpha(k). \tag{2.5}$$

By assumption, the above sum is finite. Hence, (2.4) and (2.5) complete the proof. ■

2.2 Multiresolution analysis in two-dimension

The simplest approach consists in building a 2-D multiresolution analysis by taking the direct (tensor) product of two such structure in 1-D, one for the x direction, one for the y direction. If $V_j \in \mathbb{Z}$ is a multiresolution analysis of $L^2(\mathbb{R})$, then $V_j^{(2)} = V_j \bigotimes V_j, j \in \mathbb{Z}$ is a multiresolution analysis of $L^2(\mathbb{R}^2)$. Writing again $V_j^{(2)} \bigoplus W_j^{(2)} = V_{j+1}^{(2)}$, where $W_j^{(2)}$ is orthogonal complement of $V_j^{(2)}$, it is easy to see that this 2-D analysis requires one scaling function: $\phi(x,y) = \phi(x)\phi(y)$, and three wavelets:

$$\psi^h(x,y) = \phi(x)\psi(y), \quad \psi^v(x,y) = \psi(x)\phi(y), \quad \psi^d(x,y) = \psi(x)\psi(y). \tag{2.6}$$

ψ^h detects preferentially horizontal edges, that is, discontinuities in the vertical direction, whereas ψ^v and ψ^d detect vertical and oblique edges, respectively. Indeed, for j=1, the relation $V_1 = V_0 \bigoplus W_0$ yields:

$$V_1^{(2)} = V_1^{(x)} \bigotimes V_1^{(y)} = (V_0^{(x)} \bigoplus W_0^{(x)}) \bigotimes (V_0^{(y)} \bigoplus W_0^{(y)}),$$

where $V_0^{(2)} = V_0^{(x)} \bigotimes V_0^{(y)} \ni \phi(x)\phi(y)$ is the direct sum of three other products, generated by three wavelets given in (2.6), respectively. Based on the two-multiresolution analysis discussed in Vidakovic [34] and Antoine et al. [3], we introduce two-variate wavelet density estimators. Let f be a density from $\mathbb{L}_2(\mathbb{R}^2)$. The wavelet series is

$$\begin{aligned} f_{X,Y}(x,y) &= \sum_{\underline{k}} \alpha_{j_0,\underline{k}} \phi_{j_0,k_1}(x)\phi_{j_0,k_2}(y) + \sum_{j \geq j_0} \sum_{\underline{k}} (d_{j,\underline{k}}^{(1)} \phi_{j,k_1}(x)\psi_{j,k_2}(y) \\ &\quad + d_{j,\underline{k}}^{(2)} \psi_{j,k_1}(x)\phi_{j,k_2}(y) + d_{j,\underline{k}}^{(3)} \psi_{j,k_1}(x)\psi_{j,k_2}(y)), \end{aligned}$$

where

$$\begin{aligned} \alpha_{j_0,\underline{k}} &= \int\int \phi_{j,k_1}(x)\phi_{j,k_2}(y) f_{X,Y}(x,y)dxdy, \\ d_{j,\underline{k}}^{(1)} &= \int\int \phi_{j,k_1}(x)\psi_{j,k_2}(y) f_{X,Y}(x,y)dxdy, \\ d_{j,\underline{k}}^{(2)} &= \int\int \psi_{j,k_1}(x)\phi_{j,k_2}(y) f_{X,Y}(x,y)dxdy, \\ d_{j,\underline{k}}^{(3)} &= \int\int \psi_{j,k_1}(x)\psi_{j,k_2}(y) f_{X,Y}(x,y)dxdy \end{aligned}$$

and their estimators are

$$\hat{\alpha}_{j_0,\underline{k}} = \frac{1}{n}\sum_{i=1}^{n} \phi_{j,k_1}(X_i)\phi_{j,k_2}(Y_i), \quad \hat{d}_{j,\underline{k}}^{(1)} = \frac{1}{n}\sum_{i=1}^{n} \phi_{j,k_1}(X_i)\psi_{j,k_2}(Y_i)$$
$$\hat{d}_{j,\underline{k}}^{(2)} = \frac{1}{n}\sum_{i=1}^{n} \psi_{j,k_1}(X_i)\phi_{j,k_2}(Y_i), \quad \hat{d}_{j,\underline{k}}^{(3)} = \frac{1}{n}\sum_{i=1}^{n} \psi_{j,k_1}(X_i)\psi_{j,k_2}(Y_i)$$

3. Two-dimensional wavelet for density function estimation

Let the process $\{X_i\}$ be strongly mixing, i.e.,

$$|Pr\{X_i \in A_1, X_{i+k} \in B_1\} - Pr\{X_i \in A_1\}Pr\{X_{i+k} \in B_1\}| \leq \alpha(k),$$

where $\alpha(k) \to 0$ as $k \to \infty$. Let (X_i, X_{i+1}), $i = 0, ..., n$ has unknown joint density function $f_{X_i,X_{i+1}}$ on $\mathbb{R}^2$. We calculate marginal density function by integrating X_{i+1},

$$
\begin{aligned}
f_{X_i}(x) &= \int_0^1 f_{X_i,X_{i+1}}(x,y)dy = \sum_{\underline{k}} \alpha_{j0,\underline{k}} \phi_{j0,k_1}(x) \int_0^1 \phi_{j0,k_2}(y)dy \\
&+ \sum_{j\geq j0} \sum_{\underline{k}} \Bigg[d^{(1)}_{j,\underline{k}} \phi_{j,k_1}(x) \int_0^1 \psi_{j,k_2}(y)dy \\
&+ d^{(2)}_{j,\underline{k}} \psi_{j,k_1}(x) \int_0^1 \phi_{j,k_2}(y)dy + d^{(3)}_{j,\underline{k}} \psi_{j,k_1}(x) \int_0^1 \psi_{j,k_2}(y)dy \Bigg].
\end{aligned}
$$

Now,

$$\int_0^1 \phi_{j,k_2} dy = 2^{j/2} \int_0^1 \phi(2^j y - k_2)dy.$$

Let $t = 2^j y - k_2$. Then

$$\int_0^1 \phi_{j,k_2} dy = 2^{-j/2} \int_{-k_2}^{2^j - k_2} \phi(t)dt.$$

If $0 \leq k2 \leq 2^j - 1$ then we have

$$\int_0^1 \phi_{j,k_2} dy = 2^{-j/2} \int_0^1 \phi(t)dt = 2^{-j/2}.$$

Similarly, for $0 \leq k_2 \leq 2^j - 1$, we have

$$\int_0^1 \psi_{j,k_2}(y)dy = 2^{-j/2} \int_0^1 \psi(t)dt = 0.$$

By our assumption X_i has density function independent of i. Thus,

$$f_X(x) = \sum_{k_1} \beta^1_{j0,k_1} \phi_{j0,k_1}(x) + \sum_{j\geq j0} \sum_{k_1} \gamma^1_{j,k_1} \psi_{j,k_1}(x), \tag{3.1}$$

where

$$\beta^1_{j0,k_1} = 2^{-j0/2} \sum_{k_2=0}^{2^j 0-1} \alpha_{j0,\underline{k}} \text{ and } \gamma^1_{j,k_1} = 2^{-j/2} \sum_{k_2=0}^{2^j 0-1} d^{(2)}_{j,\underline{k}}.$$

Now, using (3.1), we propose the following density function estimator:

$$\hat{f}_2(x) = \sum_{k_1 \in K_{j_0}} \hat{\beta}^1_{j_0,k_1} \phi_{j_0,k_1}(x), \tag{3.2}$$

where

$$\hat{\beta}^1_{j_0,k_1} = 2^{-j_0/2} \sum_{k_2=0}^{2^{j_0}-1} \hat{\alpha}_{j_0,\underline{k}} = \frac{1}{n} \sum_{i=1}^{n} \left[\phi_{j_0,k_1}(X_{i-1}) (2^{-j_0/2} \sum_{k_2=0}^{2^{j_0}-1} \phi_{j_0,k_2}(X_i)) \right].$$

Since ϕ has compact support, $\sum_{k_2=0}^{2^{j_0}-1} 2^{-j_0/2} \phi_{j_0,k_2}(X_i)$ is a finite sum. Thus, $\hat{f}_1$ and $\hat{f}_2$ are close to each other. We can investigate general $L_{\acute{p}}$ convergence rates. The following two lemmas will be useful later.

Lemma 3.3 (Meyer [23]) *Let ϕ be a piecewise continuous function such that for any $i \in \mathbb{N}$ the set of functions $\{\phi_{j,k} = 2^{j/2}\phi(2^j x - k),\ k \in \mathbb{Z}\}$ is an orthogonal family of $L_2(\mathbb{R})$.*

Moreover, suppose that $\theta(x) = \sum_{k\in\mathbb{Z}} |\phi(x-k)| < \infty$. Let $f(x) = \sum_{k\in\mathbb{Z}} \lambda_k \phi_{j,k}$. Then, for $1 \le p \le \infty, \frac{1}{p} + \frac{1}{q} = 1$, we have

$$2^{j(1/2-1/p)}\|\lambda\|_{l_p}\frac{1}{\|\theta\|_1^{1/q}\|\theta\|_\infty^{1/p}} \le \|f\|_p \le 2^{j(1/2-1/p)}\|\lambda\|_{l_p}\|\theta\|_p$$

Lemma 3.4 (Leblanc [16]) *Let $\infty > p \ge 2$ and $\xi_1, ..., \xi_n$ be a sequence of real-valued random variable such that $\mathbb{E}(\xi_i) = 0$, $\|\xi_i\|_\infty < S$, and $\mathbb{E}(\xi_i^2) \le \sigma^2$. Then, there exists C such that*

$$\mathbb{E}\left(\left|\sum_{i=1}^n \xi_i\right|^p\right) \le C\left\{\left(\frac{n}{l}\right)^{p/2}\sigma_l^p + \frac{n}{l}\sigma_l^2(lS)^{p-2} + S^p n^p \alpha(l)\right\},$$

where $l \in \mathbb{N}$, $2 \le l \le \frac{n}{2}$,

$$\sigma_l^2 = \max\left\{\max_{1\le u\le n}\sigma_u^2(l), \max_{1\le u\le n}\sigma_u^2(l-1)\right\} \quad \text{and} \quad \sigma_u^2(l) = \mathbb{E}\left(\sum_{i=u}^{u+l-1}\xi_i\right)^2.$$

The following two theorems are proved for $\hat{f}_1$ by Leblanc [16]. We prove these theorems for our proposed estimator $\hat{f}_2$.

Theorem 3.5 *Let $f_X \in F_{s,p,q}$ with $s \ge \frac{1}{p}$, $p \ge 1$, and $q \ge 1$. Suppose that there exist constants $\alpha > 1$ and c_α such that for any l, $\alpha(l) \le c_\alpha \alpha^{-1}$. Furthermore, suppose that there is a function g with $g(l) \ge G$ (G is a positive constant), such that for any $l = O(\ln(n))$, $\sigma_l^2 \le lg(l)$. Then, for $\acute{p} \ge \max(2,p)$, there exists a constant C such that*

$$\mathbb{E}\|f_X - \hat{f}_2\|_{\acute{p}}^2 \le C\left[\frac{n}{g(\ln(n))}\right]^{-\frac{2\acute{s}}{1+2\acute{s}}},$$

where $\acute{s} = s + \frac{1}{\acute{p}} - \frac{1}{p}$ and $2^{j_0} = \left[\frac{n}{g(\ln(n))}\right]^{\frac{1}{1+2\acute{s}}}$.

Theorem 3.6 *Let $f_X \in F_{s,p,q}$ with $s \ge \frac{1}{p}$, $p \ge 1$, and $q \ge 1$. Suppose that $\alpha(l) \le c_\alpha l^{-\alpha}$, $\alpha \ge \acute{p}(1+\acute{s})/\acute{s}$ for any $l \in \mathbb{N}, 2 \le l \le n/2$. Let us set $\mu = \acute{p}(\acute{s}+1)/[\alpha(1+2\acute{s})]$ and suppose that there is a function g with $g(l) \ge G$ (G is a positive constant), such that for any $l = O(\ln(n))$, $\sigma_l^2 \le lg(l)$. Then, for $\acute{p} \ge \max(2,p)$, there exists a constant C such that*

$$\mathbb{E}\|f_X - \hat{f}_2\|_{\acute{p}}^2 \le C\left[\frac{n}{g(n^\mu)}\right]^{-\frac{2\acute{s}}{1+2\acute{s}}},$$

where $\acute{s} = s + 1/\acute{p} - 1/p$.

Theorem 3.3 and 3.4 are corollaries of the following lemmas:

Lemma 3.7 *Let* $f_X \in F_{s,p,q}$ *with* $s \geq \frac{1}{p}$, $p \geq 1$, $q \geq 1$ *and* $\acute{p} \geq \max(2,p)$. *Then, there exists a constant* C *such that*

$$\mathbb{E}\|f_X - \hat{f}_2\|_{\acute{p}}^2 \leq C\left\{2^{-2j_0\acute{s}} + \frac{2^{j_0}}{n}\frac{\sigma_l^2}{l} + \left(\frac{2^{j_0}}{n}\right)^{2-2/\acute{p}} l^{2/\acute{P}(\acute{p}-3)}\sigma_l^{4/\acute{p}} + 2^{2j_0}\alpha(l)^{2/\acute{p}}\right\},$$

where $l \in \mathbb{N}$, $2 \leq l \leq \frac{n}{2}$, *and* $\acute{s} = \frac{s+1}{\acute{p}} - \frac{1}{p}$.

Proof. First, we decompose $\mathbb{E}\|f_X - \hat{f}_2\|_{\acute{p}}^2$ into a bias term and a stochastic term

$$\mathbb{E}\|f_X - \hat{f}_2\|_{\acute{p}}^2 \leq 2(\|f_X - P_{j_0}f_X\|_{\acute{p}}^2 + \mathbb{E}\|\hat{f}_2 - P_{j_0}f\|_{\acute{p}}^2) = 2(T_1 + T_2) \tag{3.3}$$

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

$$\begin{aligned}\sqrt{T_1} &= \left\|\sum_{j\geq j_0} D_j f_X\right\|_{\acute{p}} \leq \sum_{j\geq j_0}(\|D_j f\|_{\acute{p}}2^{j\acute{s}})2^{-j\acute{s}} \\ &\leq \left\{\sum_{j\geq j_0}(\|Djf\|_{\acute{p}}2^{j\acute{s}})^q\right\}^{1/q}\left\{\sum_{j\geq j_0}2^{-j\acute{s}\acute{q}}\right\}^{1/\acute{q}}.\end{aligned}$$

By Hölder inequality, with $\frac{1}{q} + \frac{1}{\acute{q}} = 1$,

$$C\|f_X\|_{B^{\acute{s}}_{\acute{p},q}}2^{-\acute{s}j_0} \leq C\|f_X\|_{B^s_{p,q}}2^{-\acute{s}j_0}.$$

The last inequality holds, because the continuity of Sobolev injection (see Triebel [33] and the discussion in Donoho et al. [12]), $B^s_{p,q} \subset B^{\acute{s}}_{\acute{p},q}$ implies that $\|f_X\|_{B^{\acute{s}}_{\acute{p},q}} \leq \|f_X\|_{B^s_{p,q}}$. Thus,

$$T_1 \leq K2^{-2\acute{s}j_0}. \tag{3.4}$$

Now,

$$T_2 = \mathbb{E}\|\hat{f}_2 - P_{j_0}f\|_{\acute{p}}^2 = \mathbb{E}\left\|\sum_{k\in K_{j_0}}(\hat{\beta}^1_{j_0,k} - \alpha_{j_0,k})\phi_{j_0,k}(x)\right\|_{\acute{p}}^2.$$

By Lemma 3.1,

$$T_2 \leq C\mathbb{E}\{\|\hat{\beta}^1_{j_0,k} - \alpha_{j_0,k}\|_{l_{\acute{p}}}^2\}2^{2j_0(1/2-1/\acute{p})}.$$

Using Jensen inequality we obtain,

$$T_2 \leq C2^{2j_0(1/2-1/\acute{p})}\left\{\sum_{k\in K_{j_0}}\mathbb{E}|\hat{\beta}^1_{j_0,k} - \alpha_{j_0,k}|^{\acute{p}}\right\}^{2/\acute{p}}. \tag{3.5}$$

To complete the proof, it is enough to estimate $\mathbb{E}|\hat{\beta}^1_{j_0,k} - \alpha_{j_0,k}|^{\acute{p}}$. We know

$$\hat{\beta}^1_{j_0,k} - \alpha_{j_0,k} = \frac{1}{n}\sum_{i=1}^{n}\left\{\left[\phi_{j_0,k}(X_{i-1})\sum_{k_2=0}^{2^{j_0}-1} 2^{-j_0/2}\phi_{j_0,k_2}(X_i) - \alpha_{j_0,k}\right]\right\}.$$

Denote $\xi_i = \left[\phi_{j_0,k}(X_{i-1})\sum_{k_2=0}^{2^{j_0}-1} 2^{-j_0/2}\phi_{j_0,k_2}(X_i) - \alpha_{j_0,k}\right]$. Since ϕ has compact support, $\sum_{k_2=0}^{2^{j_0}-1} 2^{-j_0/2}\phi_{j_0,k_2}(X_i)$ is a finite sum, and we have

$$\|\xi_i\|_\infty \le K.2^{j_0/2}\|\phi\|_\infty, \quad \mathbb{E}\xi_i = 0, \quad \mathbb{E}\xi_i^2 \le \|f_{X_i,X_{i+1}}\|_\infty$$

and

$$|\hat{\beta}^1_{j_0,k} - \alpha_{j_0,k}| = \frac{1}{n}\left|\sum_{i=1}^{n}\xi_i\right|.$$

Hence, applying Lemma 3.2 and using card $K_{j_0} = O(2^{j_0})$ we get,

$$\left\{\sum_{k\in K_{j_0}} \mathbb{E}|\hat{\beta}^1_{j_0,k} - \alpha_{j_0,k}|^{\acute{p}}\right\}^{2/\acute{p}}$$
$$\le \left\{C2^{j_0}\frac{1}{n^{\acute{p}}}\left(\left(\frac{n}{l}\right)^{\acute{p}/2}\sigma_l^{\acute{p}} + \frac{n}{l}\sigma_l^2 l^{\acute{p}-2}2^{j_0/2(\acute{p}-2)} + 2^{j_0\acute{p}/2}n^{\acute{p}}\alpha(l)\right)\right\}^{2/\acute{p}}$$
$$\le K\left\{\frac{\sigma_l^2}{l}\frac{2^{2j_0/\acute{p}}}{n} + \sigma_l^{4/\acute{p}}\frac{2^{j_0}l^{2/\acute{p}(\acute{p}-3)}}{n^{2/\acute{p}(\acute{p}-1)}} + 2^{2j_0/\acute{p}(\acute{p}/2+1)}\alpha^{2/\acute{p}}(l)\right\}.$$

Now, substituting above inequality in (3.5) we get

$$T_2 \le K2^{2j_0(1/2-1/\acute{p})}\left\{\frac{\sigma_l^2}{l}\frac{2^{2j_0/\acute{p}}}{n} + \sigma_l^{4/\acute{p}}\frac{2^{j_0}l^{2/\acute{p}(\acute{p}-3)}}{n^{2/\acute{p}(\acute{p}-1)}} + 2^{2j_0/\acute{p}(\acute{p}/2+1)}\alpha^{2/\acute{p}}(l)\right\}$$

or

$$T_2 \le K\left\{\frac{2^{j_0}}{n}\frac{\sigma_l^2}{l} + \left(\frac{2^{j_0}}{n}\right)^{2-2/\acute{p}} l^{2/\acute{P}(\acute{p}-3)}\sigma_l^{4/\acute{p}} + 2^{2j_0}\alpha(l)^{2/\acute{p}}\right\}. \tag{3.6}$$

By substituting (3.4) and (3.6) in (3.3) completes the proof of the lemma. ∎

In the case of independent variables, $\sigma_l^2 = O(l)$. Moreover, in the dependent case a rough bound $\sigma_l^2 = O(l^2)$ can be easily obtained. If some additional conditions are imposed on the process $\{X_i\}$, the bound $\sigma_l^2 = O(l)$ can be achieved. If $\sigma_l^2 = O(l)$, then the same rate as for the independent case, $n^{-\frac{2\acute{s}}{1+2\acute{s}}}$, is attained using lemma 3.5. If the process is α-mixing, we obtain:

Lemma 3.8 *Let $\{X_n,\ n \geq 1\}$ be a stochastic process on $\mathbb{R}$. Suppose that X_n admits a bounded marginal density which is common for all n. If $\sum_{k=1}^{\infty} \alpha(k) < \infty$, then there exists a constant G such that for any $l \in \mathbb{N}$, $2 \leq l \leq n/2$, $\sigma_l^2 \leq G \cdot l$.*

Proof. First, we define

$$Y_i = \sum_{k_2=0}^{2^{j_0}-1} 2^{-j_0/2} \phi_{j_0,k_2}(X_i).$$

Now, we use the decomposition

$$\begin{aligned}\sigma_{k,u}^2(l) = &\sum_{i=k}^{u+k-1} \mathbb{E}(\phi_{j_0,k}(X_{i-1})Y_i - \alpha_{j_0,k})^2 \\ &+ 2 \sum_{u \leq i < t \leq l+u-1} \mathbb{E}(\phi_{j_0,k}(X_{i-1})Y_i - \alpha_{j_0,k})(\phi_{j_0,k}(X_{t-1})Y_t - \alpha_{j_0,k}) = T_1 + T_2.\end{aligned}$$

Now, we prove T_1 and T_2 are $O(l)$.

$$T_1 \leq l \max_{u \leq i \leq l+u-1} (\phi_{j_0,k}(X_{i-1})Y_i - \alpha_{j_0,k})^2 \leq l||f_{X_{i-1},X_i}||_\infty^2.$$

Proposition 2 of Babu et al. [2] implies that the process $\{X_i, X_{i-1}\}$ is strongly mixing with the same order of speed as $\{X_i\}$. Thus

$$T_2 \leq K \sum_{u \leq i < t \leq l+u-1} \alpha(t-i) = Kl \sum_{k=1}^{l} (1 - k/l)\alpha(k) \leq Kl \sum_{k=1}^{l} \alpha(k).$$

By assumption the above series is finite. Hence, the proof is completed. ■

To compare bias and variance of two density function estimation, $\hat{f}_1$ and $\hat{f}_2$, we prove the following lemma:

Lemma 3.9 *For x fixed, under assumptions of Theorem 2.2:*

(i) $\mathbb{E}\hat{f}_2 = \mathbb{E}\hat{f}_1$;

(ii) $Bias\hat{f}_2 = Bias\hat{f}_1$;

(iii) $Var\hat{f}_2 = O\left(\frac{2^{j_0}}{n}\right)$.

Proof. (i) and (ii) are obvious, because $\hat{f}_1$ and $\hat{f}_2$ are unbiased estimators for $P_{j_0} f_X$. To prove (iii) note that ϕ has compact support. Thus

$$\sum_{k_2=0}^{2^{j_0}-1} 2^{-j_0/2} \phi_{j_0,k_2}(X_i)$$

is a finite sum and hence substituting in $\hat{f}_2$ and using the result of Theorem 2.2, completes the proof. ■

4. Wavelet autoregressive estimators

We consider the nonparametric regression model which is given below. Suppose that we observe the time series $X_0, X_1, ..., X_n$ following the nonlinear autoregressive model

$$X_i = \tau(X_{i-1}) + \epsilon_i, \qquad i = 1, ..., n. \tag{4.1}$$

In this section our main objective is to estimate τ. Observe that

$$\tau(x) = \mathbb{E}(X_i | X_{i-1} = x).$$

We closely follow the method of Delouilie et al. [10]. We will obtain our estimator of τ by taking the ratio of wavelet estimators of g and f, where $g(x) = \tau(x) \cdot f(x)$. One uses the following estimator

$$\hat{g}_1(x) = \sum_{k=-\infty}^{\infty} \left[\frac{1}{n} \sum_{i=1}^{n} X_i \phi_{j,k}(X_{i-1}) \right] \phi_{j,k}(x). \tag{4.2}$$

We propose new wavelet estimator for g as follows: if we have confine our attention to the wavelet basis of $L_2[0,1]$, we know $g(x) = \int_0^1 y f_{X_{i-1},X_i}(x,y)dy$. Now,

$$\begin{aligned} g(x) &= \sum_{\underline{k}} \alpha_{j_0,\underline{k}} \phi_{j_0,k_1}(x) \int_0^1 y \phi_{j_0,k_2}(y)dy \\ &\quad + \sum_{j \ge j_0} \sum_{\underline{k}} d^{(1)}_{j,\underline{k}} \phi_{j,k_1}(x) \int_0^1 y \psi_{j,k_2}(y)dy \\ &\quad + d^{(2)}_{j,\underline{k}} \psi_{j,k_1}(x) \int_0^1 y \phi_{j,k_2}(y)dy + d^{(3)}_{j,\underline{k}} \psi_{j,k_1}(x) \int_0^1 y \psi_{j,k_2}(y)dy. \end{aligned} \tag{4.3}$$

We have

$$\int_0^1 y \phi_{j,k_2} dy = 2^{j/2} \int_0^1 y \phi(2^j y - k_2)dy.$$

Let $t = 2^j y - k_2$. Then

$$\int_0^1 y \phi_{j,k_2} dy = 2^{-3/2j} \int_{-k_2}^{2^j - k_2} (t + k_2)\phi(t)dt.$$

For $0 \le k2 \le 2^j - 1$, we get

$$2^{-3/2j} \int_0^1 (t + k_2)\phi(t)dt = (k_2 + c_0)2^{-3/2j},$$

where $c_0 = \displaystyle\int_0^1 t\phi(t)dt$. Similarly for $0 \le k2 \le 2^j - 1$, we have

$$\int_0^1 y \psi_{j,k_2}(y)dy = 2^{-3/2j} \int_0^1 (t + k_2)\psi(t)dt = \acute{c}_0 2^{-3/2j},$$

where $\acute{c_0} = \int_0^1 t\psi(t)dt$. For simplicity we can choose other wavelet such that $\acute{c_0} = 0$. Then the expansion of g is as follow:

$$g(x) = \sum_{k_1} \beta_{j_0,k_1}\phi_{j_0,k_1}(x) + \sum_{j\geq j_0}\sum_{k_1} \gamma_{j,k_1}\psi_{j,k_1}(x), \tag{4.4}$$

where

$$\beta_{j_0,k_1} = 2^{-3/2j_0}\sum_{k_2=0}^{2^{j_0}-1}(k_2+c_0)\alpha_{j_0,\underline{k}}$$

and

$$\gamma_{j,k_1} = 2^{-3/2j}\sum_{k_2=0}^{2^{j_0}-1}(k_2+c_0)d^{(2)}_{j,\underline{k}}.$$

Now, using estimators of coefficient (2.7) we obtain our estimator:

$$\hat{g}(x) = \sum_{k_1} \hat{\beta}_{j0,k_1}\phi_{j0,k_1}(x), \tag{4.5}$$

where

$$\begin{aligned}\hat{\beta}_{j0,k_1} &= 2^{-3/2j0}\sum_{k_2=0}^{2^j0-1}(k_2+c_0)\hat{\alpha}_{j0,\underline{k}} \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[\phi_{j_0,k_1}(X_{i-1})(2^{-3j_0/2}\sum_{k_2=0}^{2^{j_0}-1}(k_2+c_0)\phi_{j_0,k_2}(X_i))\right].\end{aligned}$$

Wavelet estimator of autoregressive model. We propose the following estimator for our autoregressive model (4.1)

$$\hat{\tau} = \frac{\hat{g}}{\hat{f}_2}, \tag{4.6}$$

where $\hat{g}$ and $\hat{f}_2$ are given by the equations (4.5) and (3.2) respectively.

Below, we study some properties of our proposed estimator (4.6). First, we prove the following useful lemma.

Lemma 4.10 *If marginal density* $f_X \in F_{s,p,q}$*, then* $g \in F_{s,p,q}$.

Proof. By (2.1), we need to prove the following two inequalities:

$$\|\beta_{j0,.}\|_{l_p} < \infty, \qquad \left[\sum_{j\geq j0}(\|\gamma_{j,.}\|_{l_p}2^{j(s+1/2-1/p)})q\right]^{1/q} < \infty. \tag{4.7}$$

We know $f_X \in F_{s,p,q}$. Thus, (2.1) holds for f_X and hence

$$\|\beta^1_{j0,.}\|^p_{l_p} = \sum_{k_1}\left[2^{-j0/2}\sum_{k_2=0}^{2^{j0}-1}\alpha_{j0,\underline{k}}\right]^p < \infty.$$

Now, since $0 \leq k_2 + c_0 < 2^j$, $\sum_{k_1}\left[2^{-3/2j0}\sum_{k_2=0}^{2^{j0}-1}(k_2+c_0)\alpha_{j0,\underline{k}}\right]^p < \infty$ or $\|\beta_{j0,.}\|_{l_p} < \infty$. Similarly, we prove $\left[\sum_{j\geq j0}(\|\gamma_{j,.}\|_{l_p}2^{j(s+1/2-1/p)^q})\right]^{1/q} < \infty.$ ∎

Using the following lemma, we can apply Theorems 3.3 and 3.4 to find similar convergence rate for $\hat{g}_1(x)$.

Lemma 4.11 *For every $\acute{p} \geq 2$ we have:*

$$\|\hat{g} - g\|^2_{\acute{p}} \leq \|\hat{f}_2 - f_X\|^2_{\acute{p}}. \tag{4.8}$$

Proof. We have

$$\|\hat{g} - g\|^2_{\acute{p}} \leq 2(\|g - P_{j_0}g\|^2_{\acute{p}} + \|\hat{g} - P_{j_0}g\|^2_{\acute{p}})$$

and

$$\|\hat{f}_2 - f_X\|^2_{\acute{p}} \leq 2(\|f_X - P_{j_0}f_x\|^2_{\acute{p}} + \|\hat{f}_2 - P_{j_0}f_X\|^2_{\acute{p}}).$$

Now, by (3.7), we have:

$$\begin{aligned}\|f_X - P_{j_0}f_X\|^2_{\acute{p}} &= \left\|\sum_{j\geq j_0}\sum_{k_1}\gamma^1_{j,k_1}\psi_{j,k_1}(x)\right\|^2_{\acute{p}}\\ &= \left\|\sum_{j\geq j_0}\sum_{k_1}2^{-j/2}\sum_{k_2=0}^{2^j-1}d^{(2)}_{j,\underline{k}}\psi_{j,k_1}(x)\right\|^2_{\acute{p}}.\end{aligned}$$

Since $0 \leq k_2 + c_0 < 2^j$, we have

$$\begin{aligned}\|f_X - P_{j_0}f_X\|^2_{\acute{p}} &\geq \left\|\sum_{j\geq j_0}\sum_{k_1}2^{-3/2j}\sum_{k_2=0}^{2^j-1}(k_2+c_0)\psi_{j,k_1}(x)\right\|^2_{\acute{p}}\\ &= \left\|\sum_{j\geq j_0}\sum_{k_1}\gamma_{j,k_1}\psi_{j,k_1}(x)\right\|^2_{\acute{p}} = \|g - P_{j_0}g\|^2_{\acute{p}}.\end{aligned}$$

Similarly, we can prove $\|\hat{f}_2 - P_{j_0}f_X\|^2_{\acute{p}}) \geq \|\hat{g} - P_{j_0}g\|^2_{\acute{p}}$. This proves (4.7). ∎

Using the following lemma, we compare bias and variance of $\hat{g}$ and $\hat{f}_2$. The proof of the following lemma is similar to the proof of Lemma 4.2.

Lemma 4.12 $|\mathbb{E}\hat{g}(x) - g(x)| \leq |\mathbb{E}\hat{f}_2(x) - f_X(x)|$ *and* $Var\hat{g}(x) \leq Var\hat{f}_2(x)$.

These results allow us to control the convergence rate of the estimator $\hat{\tau} = \frac{\hat{g}}{\hat{f}_2}$. Using Rosenblatt's expansion (2.6) ([28], P13), we have

$$\begin{aligned}\hat{\tau}(x) &= \frac{\mathbb{E}\hat{g}}{\mathbb{E}\hat{f}_2} + \frac{\hat{g}(x) - \mathbb{E}\hat{g}(x)}{\mathbb{E}\hat{f}_2(x)} - \frac{\hat{f}_2(x) - \mathbb{E}\hat{f}_2(x)}{[\mathbb{E}\hat{f}_2(x)]^2} \\ &\quad + O_p([\hat{g}(x) - \mathbb{E}\hat{g}(x)]^2) + O_p([\hat{f}_2(x) - \mathbb{E}\hat{f}_2(x)]^2).\end{aligned}$$

Then it follows that

$$\mathbb{E}\hat{\tau}(x) = \frac{\mathbb{E}\hat{g}}{\mathbb{E}\hat{f}_2} + O(Var\hat{g}(x)) + O(Var\hat{f}_2(x)) \leq \frac{\mathbb{E}\hat{g}}{\mathbb{E}\hat{f}_2} + O\left(\frac{2^{j_0}}{n}\right),$$

using Theorem 2.1 and Lemmas 3.4, 4.3.

Now, by equation (2.7) of Rosenblatt [28],

$$\begin{aligned}\frac{\mathbb{E}\hat{g}}{\mathbb{E}\hat{f}_2} &= \tau(x) + \frac{\mathbb{E}\hat{g}(x) - g(x)}{f_X(x)} - \frac{\mathbb{E}\hat{f}_2(x) - f_X(x)}{f_X(x)}\tau(x) \\ &\quad + O_p([g(x) - \mathbb{E}\hat{g}(x)]^2) + O_p([f_X(x) - \mathbb{E}\hat{f}_2(x)]^2).\end{aligned}$$

By Lemmas 3.4, 4.3 and Theorem 2.1, it follows that

$$\frac{\mathbb{E}\hat{g}}{\mathbb{E}\hat{f}_2} \leq \tau(x) + O(2^{-j_0 m}).$$

Therefore, the bias of the estimator $\hat{\tau}$, by (3.9), is

$$\text{bias } (\hat{\tau}(\mathrm{x})) = \mathrm{O}(2^{-\mathrm{j_0 m}}) + \mathrm{O}\left(\frac{2^{\mathrm{j_0}}}{\mathrm{n}}\right). \tag{4.9}$$

For variance of $\tau(x)$ we have,

$$\begin{aligned}Var(\hat{\tau}(x)) &\leq \frac{Var\hat{g}(x)}{[\mathbb{E}\hat{f}(x)]^2} + \frac{[\mathbb{E}\hat{g}(x)]^2}{[\mathbb{E}\hat{f}(x)]^4}Var\hat{f}(x) \\ &\quad + O_p(\mathbb{E}[g(x) - \mathbb{E}\hat{g}(x)]^4) + O_p(\mathbb{E}[f_X(x) - \mathbb{E}\hat{f}(x)]^4).\end{aligned}$$

Assuming that $f(x) > 0$ for all x and given the asymptotic biases and variance of $\hat{g}(x)$ and $\hat{f}(x)$ one can easily, using Theorem 2.2 and Lemmas 3.4, 4.3, obtain

$$Var(\hat{\tau}(x)) \leq O\left(\frac{2^{j_0}}{n}\right). \tag{4.10}$$

5. Derivative of wavelet autoregressive estimators

In this section we restrict our attention to the space $X = [0, 1]$. Prakasa Rao [25] studied estimation of a derivative of a density using the method of wavelets. Let

ϕ be a scaling function generating an r-regular multiresoulution analysis and let $f^{(d)} \in \mathbb{L}_2(\mathbb{R})$. Assume that there exist $C_m \geq 0$ and $\beta_m \geq 0$ such that

$$|f^{(m)}(x)| \leq C_m(1+|x|)^{-\beta_m}, 0 \leq m \leq d. \tag{5.1}$$

He showed that projection of $f^{(d)}$ on V_{j_0} is

$$f^{(d)}_{n,d}(x) = \sum_k a_{j_0,k}\phi_{j_0,k}(x),$$

where $a_{j_0,k} = (-1)^d \int \phi^{(d)}_{j_0,k}(x) f_X(x)dx$. So, its estimator is

$$\hat{f}^{(d)}_{n,d}(x) = \sum_k \hat{a}_{j_0,k}\phi_{j_0,k}(x), \tag{5.2}$$

where $\hat{a}_{j_0,k} = \frac{(-1)^d}{n}\sum_{i=1}^{n} \phi^{(d)}_{j_0,k}(X_i)$.

Define the kernel $E(u,v)$ by $E(u,v) = \sum_k \phi(u,k)\acute{\phi}(y-k)$. We rewrite the above estimator, in a special case $d=1$,

$$\hat{\acute{f}}(x) = \frac{-2^{2j_0}}{n}\sum_{i=1}^{n} E(2^{j_0}x, 2^{j_0}X_i).$$

Note that $\frac{\partial}{\partial y}K(u,y) = E(u,y)$. By using results of Prakasa Rao [26], we see that there exist constants G_j such that

$$\int |E(x,y)|^j dy \leq G_j, j \geq 1. \tag{5.3}$$

As above, we can show that the projection of $\acute{g}$ on V_{j_0} is

$$\acute{g}(x) = \sum_k b_{j_0,k}\phi_{j_0,k}(x),$$

where $b_{j_0,k} = (-1)\int \acute{\phi}_{j_0,k}(x)g(x)dx$. Thus, its estimator is

$$\hat{\acute{g}}(x) = \sum_k \hat{b}_{j_0,k}\phi_{j_0,k}(x), \tag{5.4}$$

where $\hat{b}_{j_0,k} = \frac{-1}{n}\sum_{i=1}^{n} X_{i+1}\acute{\phi}_{j_0,k}(X_i)$.

Now, we want to find the derivative of estimated dynamical system τ. We have

$$\acute{\tau}(x) = \frac{g^*(x)}{f(x)},$$

where $g^*(x) = \acute{g}(x) - \tau(x)\acute{f}(x)$. Thus, we propose the derivative as follows

$$\hat{\tau}(x) = \frac{\hat{\acute{g}}(x) - \hat{\tau}(x)\hat{\acute{f}}(x)}{\hat{f}(x)}. \tag{5.5}$$

To control the convergence rate of our proposed estimator, we need bias and variance of $\hat{g}^*$ where $\hat{g}^*(x) = \hat{\acute{g}}(x) - \hat{\tau}(x)\hat{\acute{f}}(x)$.

Theorem 5.13 *Let the mixing rate α satisfy $\sum_{k=1}^{\infty} \alpha(k) < \infty$ and suppose that the density functions f_X and $\acute{f}_X$ are uniformly bounded, $\acute{f} \in \mathbb{L}_2(\mathbb{R})$ and $j_0 \to \infty$ as $n \to \infty$,*

$$\text{Bias } \hat{\text{g}}^*(\text{x}) = O\left(\left(\frac{2^{3j_0}\ln(n)}{n}\right)^{1/2}\right) + O(2^{-j_0(1-1/p)}) + O(2^{-j_0 m}) + O\left(\frac{2^{j_0}}{n}\right)$$

and

$$\text{Var}(\hat{\text{g}}^*(\text{x})) = \text{O}\left(\frac{2^{3\text{j}_0}}{\text{n}}\right).$$

Proof. We write

$$\begin{aligned} \mathbb{E}\hat{g}^*(x) - g^*(x) &= [\mathbb{E}\hat{\acute{g}}(x) - g(x)] - \mathbb{E}\hat{\tau}(x)[\hat{\acute{f}}(x) - \acute{f}(x)] - \acute{f}(x)[\mathbb{E}\hat{\tau}(x) - \tau(x)] \\ &\leq |\text{Bias } \hat{\acute{\text{g}}}(\text{x})| + |\text{ Bias } \hat{\acute{\text{f}}}(\text{x})| + |\acute{\text{f}}(\text{x})||\text{ Bias } \hat{\tau}(\text{x})| = \text{T}_1 + \text{T}_2 + \text{T}_3. \end{aligned}$$

If $\acute{g}(x) \in F_{s,p,q}$, $0 < p < r$, $1 \geq p, q < \infty$, with $s > 1/p$ and multiresolution analysis is r-regular, then it follows by arguments given in Kerkyacharian and Picard [24] that

$$T_1 = O(2^{-j_0(1-1/p)}). \tag{5.6}$$

By Prakasa Rao [26],

$$T_2 = O\left(\left(\frac{2^{3j_0}\log n}{n}\right)^{1/2}\right) + O(2^{-j_0(1-1/p)}). \tag{5.7}$$

By equation (4.9),

$$T_3 = O(2^{-j_0 m}) + O\left(\frac{2^{j_0}}{n}\right). \tag{5.8}$$

Using (5.5), (5.6) and (5.7) the first assertion of theorem is proved. For, the proof of the second assertion of theorem, note that,

$$\hat{g}^*(x) = \frac{-2^{2j_0}}{n}\sum_{i=1}^{n}\{X_{i+1} - \hat{\tau}(x)\}E(2^{j_0}x, 2^{j_0}X_i).$$

Hence,

$$Var(\hat{g}^*(x)) = \frac{2^{4j_0}}{n^2}\left\{\sum_{i=1}^{n} Var(\{X_{i+1}-\hat{\tau}(x)\}E(2^{j_0}x,2^{j_0}X_i))\right.$$
$$\left.+2\sum_{1\le i<t\le n} Cov(\{X_{i+1}-\hat{\tau}(x)\}E(2^{j_0}x,2^{j_0}X_i),\{X_{t+1}-\hat{\tau}(x)\}E(2^{j_0}x,2^{j_0}X_t))\right\}$$
$$= \frac{2^{4j_0}}{n^2}\{T_1+T_2\}$$

Now, we have

$$\begin{aligned}T_1 = & \sum_{i=1}^{n}\{\mathbb{E}(\{X_{i+1}-\hat{\tau}(x)\}E(2^{j_0}x,2^{j_0}X_i))^2-(\mathbb{E}\{X_{i+1}-\hat{\tau}(x)\}E(2^{j_0}x,2^{j_0}X_i))^2\}\\ \le & \; n\int E^2(2^{j_0}x,2^{j_0}y)f(y)dy+n\left(\int E(2^{j_0}x,2^{j_0}y)f(y)dy\right)^2\\ = & \; n2^{j_0}\int E^2(2^{j_0}x,y)f(2^{-j_0}y)dy+\left(2^{-j_0}\int |E(2^{j_0}x,y)|f(2^{-j_0}y)dy\right)^2.\end{aligned}$$

Using equation (5.3), we have

$$T_1 \le Kn2^{-j_0}(1+o(1)). \tag{5.9}$$

On the other hand, we have

$$\begin{aligned}&Cov(\{X_{i+1}-\hat{\tau}(x)\}E(2^{j_0}x,2^{j_0}X_i),\{X_{t+1}-\hat{\tau}(x)\}E(2^{j_0}x,2^{j_0}X_t))\\ &\quad=\mathbb{E}\{X_{i+1}-\hat{\tau}(x)\}E(2^{j_0}x,2^{j_0}X_i)\{X_{t+1}-\hat{\tau}(x)\}E(2^{j_0}x,2^{j_0}X_t)\\ &\quad-(\mathbb{E}\{X_{i+1}-\hat{\tau}(x)\}E(2^{j_0}x,2^{j_0}X_i))^2\\ &\quad\le\int\int |E(2^{j_0}x,2^{j_0}y)||E(2^{j_0}x,2^{j_0}u)|f_{X_i,X_t}(y,u)dydu\\ &\quad+\left(\int |E(2^{j_0}x,2^{j_0}y)|f(y)\right)^2 \le K2^{-2j_0}\alpha(t-1)(1+o(1)).\end{aligned}$$

Hence, $T_2 \le Kn2^{-2j_0}\sum_{1\le i<t\le n}\alpha(t-i) \le Kn2^{-2j_0}\sum_{k=1}^{n}\alpha(k)$, and thus,

$$T_2 \le Kn2^{-2j_0}. \tag{5.10}$$

Finally, using (5.7) and (5.9) the second assertion is proved. ■

The above results allow us to control the convergence rate of estimators $\hat{\tau}=\frac{\hat{g}^*}{\hat{f}}$. Using expansion (2.6) of Rosenblatt ([28], P13), we have

$$\begin{aligned}\hat{\tau}(x) \;=\; & \frac{\mathbb{E}\hat{g}^*}{\mathbb{E}\hat{f}}+\frac{\hat{g}^*(x)-\mathbb{E}\hat{g}^*(x)}{\mathbb{E}\hat{f}(x)}-\frac{\hat{f}(x)-\mathbb{E}\hat{f}(x)}{[\mathbb{E}\hat{f}(x)]^2}\\ & +O_p([\hat{g}^*(x)-\mathbb{E}\hat{g}^*(x)]^2)+O_p([\hat{f}(x)-\mathbb{E}\hat{f}(x)]^2).\end{aligned}$$

Then, it follows that

$$\mathbb{E}\hat{\tau}(x) = \frac{\mathbb{E}\hat{g}^*}{\mathbb{E}\hat{f}} + O(Var\hat{g}^*(x)) + O(Var\hat{f}(x)) \leq \frac{\mathbb{E}\hat{g}^*}{\mathbb{E}\hat{f}} + O\left(\frac{2^{j_0}}{n}\right) + O\left(\frac{2^{3j_0}}{n}\right),$$

by Theorems 2.1, 5.1 and Lemma 3.4. Now, by equation (2.7) of Rosenblatt [28],

$$\begin{aligned}\frac{\mathbb{E}\hat{g}^*}{\mathbb{E}\hat{f}} &= \acute{\tau}(x) + \frac{\mathbb{E}\hat{g}^*(x) - g^*(x)}{f_X(x)} - \frac{\mathbb{E}\hat{f}(x) - f_X(x)}{f_X(x)}\acute{\tau}(x) \\ &\quad + O_p([g^*(x) - \mathbb{E}\hat{g}^*(x)]^2) + O_p([f_X(x) - \mathbb{E}\hat{f}(x)]^2).\end{aligned}$$

By Lemma 3.4, Theorems 2.1 and 5.1, it follows that

$$\frac{\mathbb{E}\hat{g}^*}{\mathbb{E}\hat{f}} \leq \acute{\tau}(x) + O(2^{-j_0 m}) + O\left(\left(\frac{2^{3j_0}\log n}{n}\right)^{1/2}\right) + O(2^{-j_0(1-1/p)}).$$

Therefore, the bias of the estimators $\hat{\acute{\tau}}$, considering (3.9), is

$$\begin{aligned}\text{bias}(\hat{\acute{\tau}}(\text{x})) &= O(2^{-j_0 m}) + O\left(\left(\frac{2^{3j_0}\log n}{n}\right)^{1/2}\right) + O(2^{-j_0(1-1/p)}) \\ &\quad + O\left(\frac{2^{j_0}}{n}\right) + O\left(\frac{2^{j_0}}{n}\right) + O\left(\frac{2^{3j_0}}{n}\right).\end{aligned}$$

For variance of $\hat{\acute{\tau}}$ we have,

$$\begin{aligned}Var(\hat{\acute{\tau}}(x)) &\leq \frac{Var\hat{g}^*(x)}{[\mathbb{E}\hat{f}(x)]^2} + \frac{[\mathbb{E}\hat{g}^*(x)]^2}{[\mathbb{E}\hat{f}(x)]^4}Var\hat{f}(x) \\ &\quad + O_p(\mathbb{E}[g^*(x) - \mathbb{E}\hat{g}^*(x)]^4) + O_p(\mathbb{E}[f_X(x) - \mathbb{E}\hat{f}(x)]^4).\end{aligned}$$

Assuming $f(x) > 0$ for all x, and given the asymptotic biases and variance of $\hat{g}^*(x)$ and $\hat{f}(x)$, using Theorem 2.2, Lemma 3.4 and Theorem 5.1, we easily obtain

$$Var(\hat{\acute{\tau}}(x)) \leq O\left(\frac{2^{3j_0}}{n}\right). \tag{5.11}$$

6. Application in dynamical systems

In this section, we apply our wavelet estimators 4.6 and 5.5 in dynamical systems. We estimate chaotic dynamical system from noisy data and estimate metric entropy of the chaotic dynamical system. In many physical systems what is observed is only data in the form of points $\{x_1, x_2, ..., x_{n+1}\}$ on a set X. The nature of dynamical system producing the data is unknown. Estimating a point transformation $\tau : X \to X$ such that the dynamical system $x_{n+1} = \tau(x_n)$ has f as its invariant probability density function is an important problem in dynamical systems. Estimation of τ has important application in estimating metric entropy of the observed data. Metric entropy is an important measure of chaos in a dynamical system. When dealing with a system modeled by a discrete time, nonlinear difference equation, $x_{n+1} = \tau(x_n)$ the method described by Abarbanel in [1] and implemented by Short [30] provides an algorithm for computing metric entropy. When the system is contaminated by noise , as in $x_{n+1} = \tau(x_n) + \epsilon_n$, a statistical method is described by Babu et al. in [2] for estimating the transformation τ and filtering the metric entropy of τ from the observed data $X^{(n)}_{\text{data}} = \{x_1, x_2, ..., x_{n+1}\}$ of the noisy system.

In the following numerical example, we show the performance of our wavelet method. We assume that the transformation τ admits an absolutely continuous invariant measure. We can extract from $X^{(n)}_{\text{data}}$ the $\tau-$invariant density f_τ. Using Pesin's formula

$$h(\tau) = \int \log|\tau'(x)| f_\tau(x)dx$$

we can estimate the metric entropy $h(\tau)$ of τ.

Now, we present an example of a dynamical system and verify the performance of our wavelet estimators.

Example 6.14 Consider the skew tent map $\tau : [0,1] \to [0,1]$ defined by

$$\tau(x) = \begin{cases} 3x, & 0 \le x < \frac{1}{3}, \\ \frac{3}{2} - \frac{3}{2}x, & \frac{1}{3} \le x \le 1. \end{cases} \tag{6.1}$$

By perturbing τ with $\epsilon-$neighborhood noise with zero mean and using Maple 9.5 we produce the data set $X^{(n)}_{data}$ for $n = 64$ and $\epsilon = .04$. Figure 1 is the graph of the chaotic dynamical system (transformation) τ, Figure 2 is the graph of noisy data and Figure 3 is the graph of the estimated transformation $\hat{\tau}$.

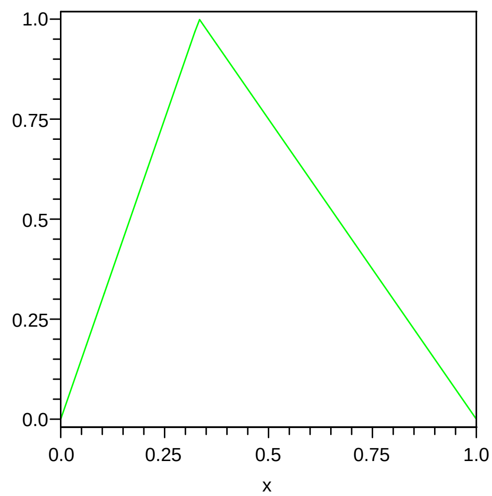

Figure 1: Graph of the transformation τ.

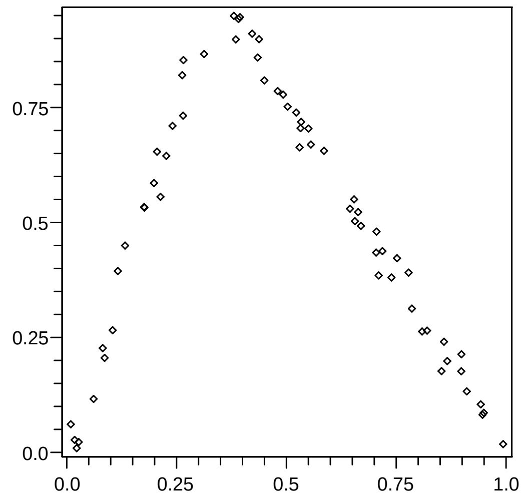

Figure 2: Graph of the noisy data.

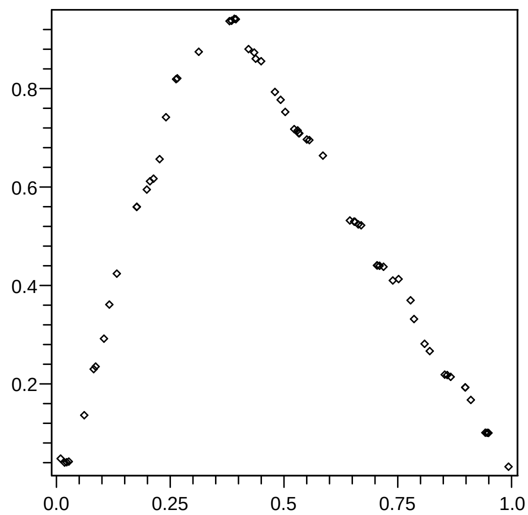

Figure 3: Graph of estimated transformation $\hat{\tau}$.

In this numerical example, we have considered the following scale function

$$\phi(x) = \begin{cases} \dfrac{e^{-\frac{1}{2}(x-\frac{1}{2})^2}}{.382\sqrt{2\pi}} & , \ 0 \leq x < 1, \\ 0 & , \ \text{otherwise}. \end{cases} \tag{6.2}$$

References

[1] ABARBANEL, H., *Analysis of observed chaotic data*, Springer-Verlag, 1995.

[2] BABU, G.J., BOYARSKY, A., CHAUBEY, Y.P. and GÓRA, P., *A new statistical method for filtering and entropy estimation of a chaotic map from noisy data*, International Journal of Bifurcation and Chaos, 14, no. 11 (2004), 3989–3994.

[3] ANTOINE, J.P., MURENZI, R., VANDERGHEYNST, P., and ALI, S.T., *Two-dimensional Wavelets and thier relatives*, Cambrige, United Kingdom, 2004.

[4] BUHLMANN, P. and MCNEIL, A.J., *An algorithm for nonparametric GARCH modelling.* Computational Statistics and Data Analysis, 40 (2002), 665–683.

[5] DAUBECHIES, I., *Ten Lectures on Wavelets*, CBMS-NSF regional conferences series in applied mathematics, SIAM, Philadelphia,1992.

[6] DAUBECHIES, I., *Orthogonal bases of compactly supported wavelets*, Comm. Pure Appl. Math. 41 (1988), 909-996.

[7] DELOUILLE, V., FRANKE, J. and VON SACHS, R., *Nonparametric stochastic regression with design-adapted wavelets*, Sankhya, ser. A 63 (2001), 328–366.

[8] DELOUILLE, V. and VON SACHS, R., *Properties of disign-adapted wavelet transforms of nonlinear autouregression models*, Discussion paper 0225, Institute of Statistics, Catholic University of Louvain, 2002.

[9] DELOUILLE, V., SIMOENS, J. and VON SACHS, R., *Design-adapted wavelet estimator fro two-dimensional tensor product irregular designs*, in Proceedings of Wavelet Applications in Signal and Image Processing X, part of SPIE's International Symposium on Optical Science and Technology, Michael A. Unser, Akram Aldroubi, and Andrew F. Laine (eds.), 2003, 880–891.

[10] DELOUILLE, V., SIMOENS, J. and VON SACHS, R., *Smoth design-adapted wavelets for Nonparametric stochastic regression* , J. Amer. Statist. Assoc., 99 (2004), 643–659.

[11] DONOHO, D. L., JOHNSTONE, I.M. , KERKYACHARIAN, G. and PICARD, D., *Wavelet shrinkage: asymptopia* (*with discussion*), J. Roy. Statist. Soc., Ser. B, 57, (2) (1995), 301–370.

[12] DONOHO, D.L. and JOHNSTONE, I.M., *Adapting to unknown smoothness via Wavelet shrinkage*, J. Amer. Statis. Assoc., B 90 (1995), 1200–1224.

[13] DOOSTI, H., CHAUBEY, Y.P. and NIROUMAND, H.A., *On a Wavelet-Based Method of Estimating a Regression Function*, Communications in Statistics. Theory and Methods, vol. 36, Issue 11 (2007), 20832098.

[14] DOUKHAN, P. and LEON, J.R., *Une note sur la déviation quadratique déstimateurs de densités par projections orthogonales,* C.R. Acad. Sci. Paris, t. 310, série 1, (1990), 425-430.

[15] FRANKE, J. HOLZBERGER, H. and MULLER, M., *Nonparametric estimation of ARMA- and GARCH-processes*, In G. Stahl W. Hardle, Th. Kleinow (eds.), Applied Quantitative Finance, Springer Verlag, 2002.

[16] LEBLANC, F., *Wavelet linear density estimator for a discrete-time stochastic process: L_p-losses* Statistic and Probability Letter, 27 (1996), 71–84.

[17] HOFFMANN, M., DOUKHAN, P. and LEON, J.R., *On nonparametric estimation in nonlinear $AR(1)$-models.*, Statistic and Probability Letter, 44 (1999), 29–45.

[18] HAFNER, C., *Nonlinear Time Series Analysis with Applications to Foreign Exchange Rate Volatility*, Physica-Verlag, 1998.

[19] HARDLE, W. and TSYBAKOV, A., *Local polynomial estimators of the volatility function in nonparametric autoregression*, Journal of Econometrics, 81 (1997), 223–242.

[20] HUANG, S.Y., *Density estimation by wavelet-based reproducing kernels*, Statistica Sinica, 9 (1999), 137–151.

[21] JONES, D., *Nonlinear regressive processes*, Proceeding of the Royal Society London, Series A, 360 (1978), 71–95.

[22] MASRY, E. and TJSTHEIM, D., *Nonparametric estimation and identification of nonlinear ARCH time series*, Econometric Theory, 11 (1995), 258–289.

[23] MEYER, Y., *Ondelettes et Operateurs* , Hermann, Paris, 1990.

[24] KERKYACHARIAN, G. and PICARD, D., *Density estimation in Besov spaces*, Statistic and Probability Letters, 13 (1992), 15–24.

[25] PRAKASA RAO, B.L.S., *Nonparametric estimation of the derivatives of a density by the method of wavelets*, Bull. Inform. Cyb., 28 (1996), 91–100.

[26] PRAKASA RAO, B.L.S., *Nonparametric estimation of partial derivatives of a multivariate probability density by the method of wavelets*, J. Time. Ser. Anal., Asymptotics in Statistics and Probability, Festschrift for G.G. Roussas (Ed. M.L. Puri), VSP, The Netherlands, 2000, 321–330.

[27] ROBINSON, P.M. , *Nonparametric estimators for time series.*, J. Time. Ser. Anal., 4 (1983), 185–207.

[28] ROSENBLATT, M., *Stochastic Curve Estimation*, NSF-CBMS Regional Conference Series in Probability and Statistics, 3, Institute of Mathematical Statistics, Harvard, California, 1991.

[29] SARDY, S., PERCIVAL, D.B., BRUCE, A.G., GAO, H.Y. and STUETZLE, W., *Wavelet De-Noising for unequally spaced data*, Statistics and Computing, 9 (1999), 65–75.

[30] SHORT, K.M., *Direct calculation of metric entropy from time series*, Journal of computational physics, 104 (1993), 162–172.

[31] TJSTHEIM, D., *Non-linear time series: a selective review*, Scand. J. Statist., 21 (1994), 97–130.

[32] TRIBOULEY, K., *Density estimation by cross-validation with wavelet method*, Statistica Neerlandica 45 (1995), 41–62.

[33] TRIBIEL, H., *Theory Of Function Space II.* BirkhaB̈irkhauser Verlag, Berlin, 1992.

[34] VIDAKOVIC, B., *Statistical modeling by wavelets*, Wiley, New York, 1999.

Accepted: 15.08.2008