Ann. Inst. Statist. Math.
Vol. 53, No. 1, 159–178 (2001)

ASYMPTOTICS FOR WAVELET BASED ESTIMATES OF PIECEWISE SMOOTH REGRESSION FOR STATIONARY TIME SERIES

YOUNG K. TRUONG[1]* AND PRAKASH N. PATIL[2]

[1]*Department of Biostatistics, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7400, U.S.A.*, e-mail: truong@bios.unc.edu
and Department of Statistics and Applied Probability, National University of Singapore, 3 Science Drive 2, Singapore 117543, Singapore
[2]*School of Mathematics and Statistics, The University of Birmingham, Birmingham B15 2TT, U.K.*

(Received April 14, 2000; revised August 16, 2000)

Abstract. Wavelet methods are used to estimate density and (auto-) regression functions that are possibly discontinuous. For stationary time series that satisfy appropriate mixing conditions, we derive mean integrated squared errors (MISEs) of wavelet-based estimators. In contrast to the case for kernel methods, the MISEs of wavelet-based estimators are not affected by the presence of discontinuities in the curves. Applications of this approach to problems of identification of nonlinear time series models are discussed.

Key words and phrases: Convergence rate, density estimation, nonparametric regression, piecewise-smoothness, wavelet.

1. Introduction

Many time series encountered in practice do not exhibit characteristics of linear Gaussian processes. Attempts to analyze such series as the sunspots, lynx and blowfly data have led to many important approaches for nonlinear models. See, for example, Priestley (1988), Tong (1990), Auestad and Tjøstheim (1990) and Tjøstheim (1994). The last two papers also discussed various issues for model identification related to the procedures investigated by Truong and Stone (1992, 1994), Tran (1993) and Truong (1994). It has been noted that these approaches are confined to the estimation of the regression function having bounded, continuous derivatives. For the data such as the electromagnectic exposure of a power line worker illustrated in Fig. 1, these methods leave something else to be desired.

To elaborate this, let the exposure data be modeled by

$$y_i = \theta(t_i) + z_i, \quad i = 0, 1, \dots, n$$

with $\theta(\cdot)$ being the mean function, $t_i = (i-1)/n$ (by rescaling the time axis accordingly), n the sample size, and $\{z_i\}$ a sequence of mean zero random variables. The data suggested that the mean function is nonlinear, perhaps piecewise constant with jumps. In the context of independent observations, a number of authors including Donoho and Johnstone (1994*a*, 1994*b*, 1998), Donoho *et al.* (1995), and Hall and Patil 1995*a*, 1995*b*,

*This research was partially suported by NIH R01-CA61937 and a Research Grant RP3982709 from NUS.

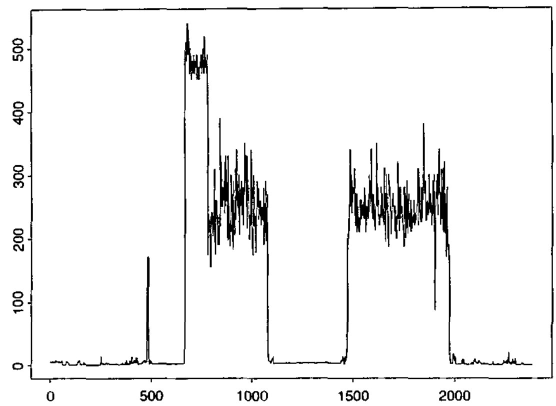

Fig. 1. The electral magnectic exposure of a lineman. (We would like to thank Dr. G. Mihlan, The University of North Carolina at Chapel Hill, for providing the dataset.)

1996*a* have demonstrated explicitly the extraordinary local adaptability of wavelet estimators in handling discontinuities such as those appeared in Fig. 1. Moreover, they observed that nonlinear wavelet estimators possess an important robustness property against oversmoothing—a property that is not shared by kernel estimators described, for example, in Auestad and Tjøstheim (1990).

For correlated noise, Johnstone and Silverman (1997), Johnstone (1999) and Wang (1996) have examined the asymptotic properties of the wavelet-based estimators of the mean function associated with the above model. Neumann and von Sachs (1997) considered the estimation of evolutionary spectra for local stationary time series using tensor products of wavelets. A common technique employed in these papers to derive the asymptotic properties is based on the Gaussian white noise model. Li and Xie (1999) addressed the problem of identifying the thresholds and time delay of threshold autoregressive models (TAR) introduced by Tong (1983).

The wavelet-based approaches to time series are largely focused on regular nonrandom designs, in some cases both sample size and smoothing parameter are dyadic. The main objective of the present paper is to generailize the properties of wavelet-based estimators of the regression functions involving time series. To handle these problems, we remove the above restrictions by generalizing the investigation of Hall and Patil (1996*a*) to time series. Also, in contrast to the aforementioned papers where the asymptotic properties are derived based on the Gaussian white noise model, we consider a different approach by following the framework of Hall and Patil (1995*b*). Specifically, given a stationary time series, we investigate the mean integrated squared error (MISE) properties of nonlinear, thresholded, wavelet-type estimators applied to both density and regression functions that are possibly discontinuous. For the wavelet-based density estimator, we derive an analogue of the classical MISE formula familiar in the context of linear, kernel-type estimators, where MISE admits an expansion with distinct variance and squared bias components. In kernel estimation, this MISE expansion is achieved by assuming the density function f has r continuous derivatives, and the expansion generally fails if such a smoothness condition is absent. By way of contrast, we show that an analogue of the expansion holds in the case of nonlinear wavelet estimators when the underlying density is only piecewise continuous.

More precisely, under reasonable mixing conditions on the dependence structure and a suitably defined smoothness parameter $r > 0$, it is shown that the mean integrated squared error of the nonlinear wavelet estimator is given by $n^{-2r/(2r+1)}$ even if smoothness conditions are imposed only in a piecewise sense. To prove this result, we first apply the MISE property mentioned above to establish a uniform consistency result for the wavelet density estimator. This makes wavelet-based density estimation (involving piecewise continuity) an initial step for estimation of a possibly discontinuous regression function.

The rest of the paper is organized as follows. Section 2 gives a concise description of the wavelets and the associated estimates. Section 3 describes the asymptotic properties of wavelet-based density and regression estimators, including the mean integrated squared error formulae for wavelet estimators. An example is given in Section 4. The last two sections contain proofs of the main results.

2. Wavelet based estimates

In Subsection 2.1, we describe stationary sequences of random variables that satisfy the α-mixing condition, which is followed by a discussion on the basic theory of wavelet methods in Subsection 2.2. Subsection 2.3 is devoted to the estimation of the marginal density function using wavelet methods, this is necessary for estimating the regression function involving bivariate time series. The regression problem is discussed in Subsection 2.4.

2.1 *Stationary time series and strong mixing*

Let (X_i, Y_i), $i = 0, \pm 1, \pm 2, \ldots$, denote a stationary sequence of random vectors. To motivate our later development we can consider the Markov model:

$$x_{i+1} = \mu(x_i) + z_i, \qquad i = 0, \pm 1, \ldots$$

where $\mu(\cdot)$ is a piecewise-smooth function and z_i is a sequence of independent zero-mean random variables. In particular, if $\mu(\cdot)$ is piecewise linear, it is called the threshold autoregressive model; see Tong (1990). In this case, $(X_i, Y_i) = (x_{i-1}, x_i)$ for $i = 0, \pm 1, \pm 2, \ldots$. Alternatively, we can think of (X_i, Y_i) as in the usual regression setting for time series.

Let $\mathcal{F}_j$ and $\mathcal{F}^j$ denote the σ-fields generated respectively by (X_i, Y_i), $-\infty < i \leq j$, and (X_i, Y_i), $j \leq i < \infty$. Given a positive integer u, set

$$\alpha(u) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}_j \text{ and } B \in \mathcal{F}^{j+u}\}.$$

The stationary sequence is said to be α-mixing or strongly mixing if $\alpha(u) \to 0$ as $u \to \infty$.

Among various mixing conditions used in the literature, α-mixing is reasonably weak, and has many practical applications. Sufficient conditions for linear processes to be α-mixing are studied by Gorodetskii (1977) and Withers (1981). Auestad and Tjøstheim (1990) provide an illuminating discussion of the role of α-mixing for model identification in nonlinear time series analysis.

2.2 *Wavelets*

Wavelet methods was introduced to statistics by Donoho (1995). See also Donoho and Johnstone (1994*a*, 1994*b*, 1998), Donoho *et al.* (1995), Kerkyacharian and Picard

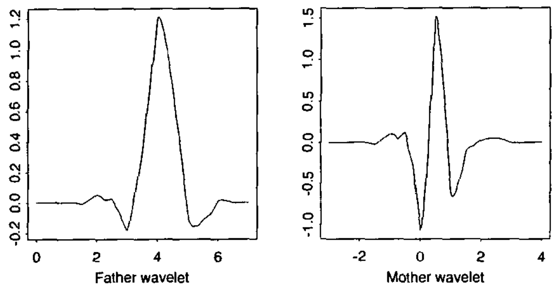

Fig. 2. A pair of father and mother wavelets.

(1992, 1993*a*, 1993*b*, 1993c), and Härdle *et al.* (1998). These authors have demonstrated the virtues of wavelet methods from the viewpoint of adaptive smoothing, typically in the context of the achievability of very good convergence rates uniformly over exceptionally large function classes. In this section, a concise mathematical description of the wavelets will be given.

Let ϕ, the "father wavelet", be a solution of the dilation equation,

$$\phi(x) = \sum_{\ell} c_\ell\, \phi(2x - \ell)\,,$$

where, for some integer $r \geq 1$, the constants c_ℓ satisfy $\sum \ell^{2r} c_\ell^2 < \infty$, $\sum c_\ell = 2$ and

$$\sum_{\ell}(-1)^\ell \ell^k c_\ell = 0\,, \qquad 0 \leq k \leq r-1. \tag{2.1}$$

We normalize ϕ so that $\int \phi = 1$. Assume too that translates of ϕ are orthonormal, i.e.

$$\int \phi(x)\phi(x-\ell)dx = \delta_{0\ell}, \qquad -\infty < \ell < \infty, \tag{2.2}$$

where δ_{ij} is the Kronecker delta. Define the "mother wavelet" by

$$\psi(x) = \sum_{\ell}(-1)^\ell c_{\ell+1}\phi(2x+\ell).$$

It follows from (2.1) that

$$\int x^k \psi(x)dx = 0, \qquad 0 \leq k \leq r-1.$$

We suppose that

(2.3) ϕ and ψ are bounded and compactly supported.

The vast majority of the wavelets used in practice satisfy these conditions; see Daubechies ((1992), Chapter 6). Figure 2 presents such a pair of mother and father wavelets. Conditions (2.1) and (2.2) ensure that for all integers $k \geq 0$ and $-\infty < \ell_1, \ell_2 < \infty$,

$$\int \psi(x-\ell_1)\,\psi(2^k x - \ell_2)\,dx = \delta_{0k}\delta_{\ell_1\ell_2} \quad \text{and} \quad \int \phi(x-\ell_1)\psi(2^k x - \ell_2)dx = 0,$$

and that $\int y^j \psi(y)\,dy = 0$ for $0 \le j \le r-1$. See Strang (1989). Therefore, the functions

$$\phi_\ell(x) = p^{1/2}\phi(px-\ell), \qquad \psi_{k\ell}(x) = p_k^{1/2}\psi(p_k x - \ell),$$

for arbitrary $p > 0$, $-\infty < \ell < \infty$, $k \ge 0$ and $p_k = p2^k$, are orthonormal: $\int \phi_{\ell_1}\phi_{\ell_2} = \delta_{\ell_1\ell_2}$, $\int \psi_{k_1\ell_1}\psi_{k_2\ell_2} = \delta_{k_1k_2}\delta_{\ell_1\ell_2}$, $\int \phi_{\ell_1}\psi_{k\ell_2} = 0$. Furthermore, an arbitrary square-integrable function f may be expanded in a generalized Fourier series, of the form

$$f(x) = \sum_\ell b_\ell^f \phi_\ell(x) + \sum_k \sum_\ell b_{k\ell}^f \psi_{k\ell}(x), \tag{2.4}$$

where $b_\ell^f = \int f\phi_\ell$ and $b_{k\ell}^f = \int f\psi_{k\ell}$. The generalized Fourier series (2.4) converges in L^2.

2.3 *Wavelet-based density estimators*

We need an estimate of the density function in order to address the problem of estimating the regression function involving random predictors, which is more complicated than estimating the mean function (of time) of a univariate stationary time series, as considered by Truong and Patil (1996). (In that paper, we used wavelet-based method to estimate the *highly oscillating continuous* mean function.)

Let f denote the probability density function of X_0. Then $\hat{b}_\ell = n^{-1}\sum_{i=1}^n \phi_\ell(X_i)$ and $\hat{b}_{k\ell} = n^{-1}\sum_{i=1}^n \psi_{k\ell}(X_i)$ are unbiased estimators of b_ℓ^f and $b_{k\ell}^f$, respectively. A nonlinear wavelet estimator of f has the form

$$\hat{f}(x) = \sum_\ell \hat{b}_\ell \phi_\ell(x) + \sum_{k=0}^{q-1}\sum_\ell \hat{b}_{k\ell} I(|\hat{b}_{k\ell}| > \delta)\psi_{k\ell}(x), \tag{2.5}$$

where p, $\delta > 0$ and $q \ge 1$ are tuning parameters. The quantity p^{-1} is the wavelet analogue of the bandwidth, h, for a kernel density estimator. The parameters δ and q control the 'threshold' and level of truncation respectively. The first series in (2.5) represents an unbiased estimator of the first portion of the first series in (2.4), and converges absolutely under condition (2.3). However, the second series would not converge if it were not truncated in the manner suggested here. The more terms are included in the series, the less is the bias but the greater is the variance. The parameters δ and q adjust this trade-off between bias and variance, however, such trade-off does not occur to first-order. For detailed discussion we refer the reader to Hall and Patil (1996*b*).

2.4 *Wavelet-based regression estimators*

Let $\{(X_i, Y_i) : i = 0, \pm1, \pm2, \ldots\}$ denote a stationary bivariate time series. Consider the problem of estimating the regression function $\mu(x) = E(Y_0 \mid X_0 = x)$. Write $h(y,x)$ for the joint density function of (Y_0, X_0), and $f(x)$ for the marginal density function of X_0. Set $g(x) = \int y h(y,x)dy$. Then $\mu(x) = g(x)/f(x)$. If g is square-integrable then its wavelet expansion is given by

$$g(x) = \sum_\ell b_\ell^g \phi_\ell(x) + \sum_{k=0}^{\infty}\sum_\ell b_{k\ell}^g \psi_{k\ell}(x),$$

where $b_\ell^g = \int g\phi_\ell$ and $b_{k\ell}^g = \int g\psi_{k\ell}$. The corresponding wavelet estimator is given by

$$\hat{g}(x) = \sum_\ell \hat{b}_\ell \phi_\ell(x) + \sum_{k=0}^{q-1} \sum_\ell \hat{b}_{k\ell} I(|\hat{b}_{k\ell}| > \delta)\, \psi_{k\ell}(x),$$

where $\hat{b}_\ell = n^{-1}\sum_{i=1}^n Y_i\phi_\ell(X_i)$ and $\hat{b}_{k\ell} = n^{-1}\sum_{i=1}^n Y_i\psi_{k\ell}(X_i)$. Here δ is a threshold as discussed before. A wavelet estimator of $\mu(x)$ is given by $\hat{\mu}(x) = \hat{g}(x)/\hat{f}(x)$, where $\hat{f}(x)$ is the wavelet estimator of $f(x)$ described in Subsection 2.4.

3. Main results

For asymptotic results based on smooth functions, it is assumed that the marginal density function f is r times differentiable. This condition will be weakened to the piecewise-smooth case in the current paper. That is, we consider to estimate density and regression functions exhibiting a finite number of discontinuities. We start with the piecewise smoothness condition on the marginal density function.

CONDITION 3.1. The density f is monotone on $(-\infty, -u)$ and (u, ∞) for sufficiently large u. Moreover, $f^{(r)}$ exists in a piecewise sense. That is, there exist points $x_0 = -\infty < x_1 < \cdots < x_N < \infty = x_{N+1}$ such that the first r derivatives of f exist and are bounded and continuous on (x_i, x_{i+1}) for $0 \le i \le N$, with left- and right-hand limits. In particular, f itself may be only piecewise continuous.

Let U be a nonempty open subset (of $\mathbb{R}$) containing the origin. The following conditions are used to obtain a rate of convergence for the variance of wavelet estimators.

CONDITION 3.2. For $j \ge 1$, the conditional distribution of X_j, given $X_0 = x$, has a density $f_j(\cdot \mid x)$; and there is a positive constant M_2 such that $f_j(x' \mid x) \le M_2$ for all $x, x' \in \mathbb{R}$ and $j \ge 1$.

CONDITION 3.3. For some $0 < \rho < 1$, $\alpha(u) = O(\rho^u)$ as $u \to \infty$.

Careful examination of our proof will reveal that Condition 3.3 is somewhat stronger than actually needed. However, we choose the above form to simplify the presentation. Many interesting time series can be shown to have a geometric α-mixing rate, see Auestad and Tjøstheim (1990) and Tjøstheim (1994). Note that for density estimation, the mixing condition can be defined in terms of the sequence X_i alone.

CONDITION 3.4. The functions $g(\cdot)$ and $\mu(\cdot)$ have piecewise bounded r-th derivative.

The following condition ensures that the variance of the estimator converges to zero at a desirable rate.

CONDITION 3.5. The density function f of X_0 is bounded away from zero and infinity on U; that is, there is a positive constant M_1 such that $M_1^{-1} \le f(x) \le M_1$ for $x \in U$.

For the clarity of presentation, in this paper we consider only bounded stationary time series. Alternatively, this condition can be weakened by using the moment generating function or a sufficiently stringent moment condition depending on the smoothness parameter r.

CONDITION 3.6. Y_0 is bounded.

Define $\kappa = (r!)^{-1}\int y^r\psi(y)dy = (r!)^{-1}(-\frac{1}{2})^{r+1}\sum(-1)^j j^r c_j$. Let Q denote a fixed compact subset of U having nonempty interior.

THEOREM 3.1. *Suppose* (2.1), (2.2) *and Conditions* 3.1–6 *hold. Also, suppose* ϵ *is a sufficiently small positive constant and that* $p = p(n) \to \infty$ *and* $q = q(n) \to \infty$ *in such a manner that* $p_q\delta^2 = O(n^{-\epsilon})$ *and* $p_q^{2r+1}\delta^2 \to \infty$, *and* $\delta \geq C(n^{-1}\log n)^{1/2}$ *for some* $C > 0$.

(i) *If* $p \gg \delta^{-2/(2r+1)}$, *then*

$$E\left|\int(\hat{f}-f)^2 - \left\{n^{-1}p + p^{-2r}\kappa^2(1-2^{-2r})^{-1}\int f^{(r)^2}\right\}\right| = o(n^{-1}p + p^{-2r});$$

(ii) *If* $p = O(\delta^{-2/(2r+1)})$, *then*

$$\delta^{4r/(2r+1)} = O\left\{\int E(\hat{f}-f)^2\right\}.$$

(iii) *If* $p \gg \delta^{-2/(2r+1)}$, *then*

$$E\left|\int_Q(\hat{g}-g)^2 - \left\{\sigma_p^2 n^{-1} + p^{-2r}\kappa^2(1-2^{-2r})^{-1}\int_Q g^{(r)^2}\right\}\right| = o(n^{-1}p + p^{-2r}),$$

where $\sigma_p^2 = \sum_{\ell=0}^{p}\mathrm{var}(Y_1\phi_\ell(X_1))$.

(iv) *If* $p = O(\delta^{-2/(2r+1)})$, *and if* $g^{(r)} > 0$ *over a small interval in* Q, *then*

$$\delta^{4r/(2r+1)} = O\left\{\int_Q E(\hat{g}-g)^2\right\}.$$

(v) *Suppose* (2.1), (2.2) *hold and that* $r > 1$. *Then*

$$\int_Q(\hat{\mu}-\mu)^2 = O_p(n^{-1}p + p^{-2r}).$$

Moreover, if p *is chosen of size* $n^{1/(2r+1)}$ *with* $(r+1)/(2r+1) < \epsilon < 2r/(2r+1)$, *then*

$$\int_Q(\hat{\mu}-\mu)^2 = O_p(n^{-2r/(2r+1)}).$$

A proof of this result is given in Section 5.

Remark 1. (i) Note that by choosing $p \sim \text{const.}\, n^{1/(2r+1)}$ it can be shown that the mean integrated squared error satisfies

$$\int E(\hat{f}-f)^2 \sim n^{-1}p + p^{-2r}\kappa^2(1-2^{-2r})^{-1}\int f^{(r)^2} \sim \text{const.}\, n^{-2r/(2r+1)},$$

which is the usual optimal rate of convergence for estimating an r–times differentiable density function f. See also Remarks 2.1 and 2.2 of Hall and Patil (1995*b*) for an illuminating discussion of the robustness of the wavelet density estimator against oversmoothing. Note that, if $p \sim \text{const.}\, n^{1/(2r+1)}$, then ϵ is chosen so that $0 < \epsilon < 2r/(2r+1)$.

(ii) For a compactly supported density function f, results for wavelet based density estimates may be refined as follows. Suppose that $\operatorname{supp} f = [c,d]$, a compact interval. Also, $f^{(r)}$ restricted to $[c,d]$ is bounded away from zero in neighbourhoods of points of discontinuity, and has only a finite number of zeros, in neighbourhoods of which $f^{(r+1)}$ exists and is bounded away from zero; and that $p^{2r+1}\delta^2 \to \ell$ where $0 \le \ell < \infty$. Then

$$\int E(\hat{f}-f)^2 \sim C(\ell)\delta^{4r/(2r+1)},$$

where

$$C(\ell) = \begin{cases} \ell^{-2r/(2r+1)}\kappa^2 \int f^{(r)^2}\left[\sum_k 2^{-2rk} I\{(\kappa f^{(r)})^2 \le \ell 2^{(2r+1)k}\}\right] & \text{if } \ell > 0 \\ (1-2^{-2r})^{-1}|\kappa|^{2/(2r+1)} \int |f^{(r)}|^{2/(2r+1)} & \text{if } \ell = 0. \end{cases}$$

(iii) It follows from the proof of this theorem that $\sum_{\ell=1}^{p} \operatorname{var}(Y_1\phi_\ell(X_1)) = O(p)$. Hence

$$E\int_Q (\hat{g}-g)^2 = O\left(n^{-1}p + p^{-2r}\kappa^2(1-2^{-2r})^{-1}\int_Q g^{(r)^2}\right).$$

(iv) The presence of discontinuities can have an adverse effect on the performance and asymptotic properties of kernel estimators; see Remark 2.6 of Hall and Patil (1995*b*). Specifically, in view of the above result, the rate of convergence of kernel estimators is inferior to that in the wavelet approach.

(v) From Subsection 2.2, we note that the father wavelet has a behavior similar to the density function, thus it has the same effect as the kernel function in modeling the low frequency component. On the other hand, the mother wavelet has r vanishing moments and is suitable for handling the high frequency components. Consequently, through Taylor expansion, one can easily see that the wavelet coefficients from the mother side will be very small for smooth functions, while they can be large for functions with jumps. This explains heuristically why wavelets are useful for detecting the singularity of a given function. The location of the singularity can also be precisely located using the the translation and dilation.

4. An example and some concluding remarks

To illustrate the usefulness of the wavelet based method described in the previous sections, we use the linemen time series for constructing the regression function. In the absence of the covariate series, the regression problem becomes the one for estimating the mean function of the series. Here the data is viewed as the superposition of the mean and the stationary noise series (Fig. 1). After applying the (discrete) wavelet transform, the coefficients b_ℓ and $b_{k\ell}$ are presented in Fig. 3. Here the wavelet expansion is carried out using 6 levels.

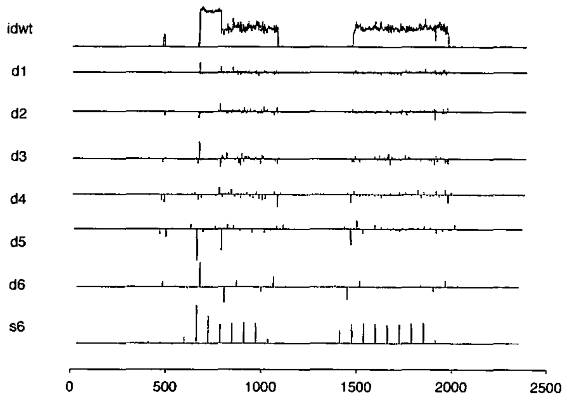

Fig. 3. The coefficients in the discrete wavelet transformation of the linemen time series.

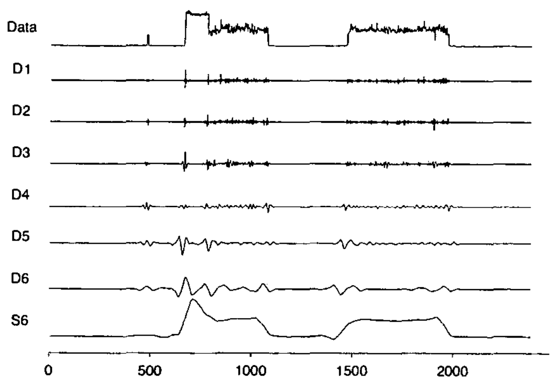

Fig. 4. Decompostion of the data into seven wavelet functions.

Figure 4 shows the multiresolution decompostion of the linemen data into seven wavelet function. Each function is the product of the coefficients and the corresponding scaled-translated wavelets at each level. The estimated mean function is given at the bottom of the figure. The sum of all the seven wavelet functions yields the so called multiresolution approximations of the data. This is illustrated in Fig. 5.

In this paper, we present some asymptotic results for estimating a possibly discontinued function from stationary time series. Under appropriate conditions, it has shown that the optimal rates of convergence of the estimates of the density and the regression functions can be achieved in a manner similar to the usual case in handling random samples. The feasibility in computing these estimates has been demonstrated through the example given above. Software packages that include wavelet analysis are: XploRe (Härdle *et al.* (1995)), SPLUS and MATLAB. There is however more works to be done in the regression problems involving bivariate time series. In particular, it will be important to address the problem in setting the threshold level. Another equally important problem is to use wavelet method to identify the number of thresholds in TAR models.

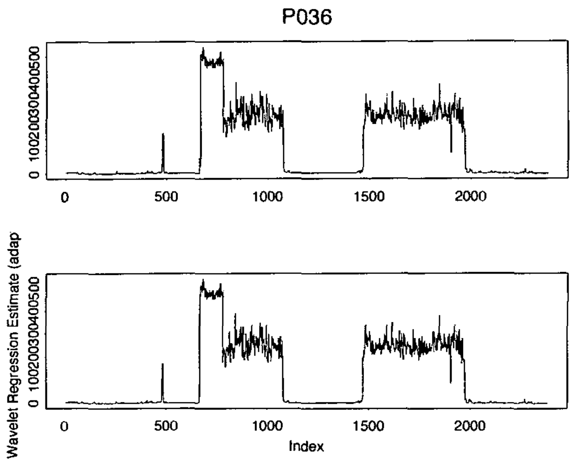

Fig. 5. A wavelet shrinkage analysis of the linemen data.

5. Proof of Theorem 3.1

A sketch of the proof will be given in this section. More details can be found in the Appendix. To simplify the presentation, we may assume, without loss of generality, that ϕ and ψ are compactly supported on $[0,1]$ in this section. Moreover, $Q=[0,1]$.

Proofs of (i) and (ii) will be omitted since they are special cases of the proof of (iii). We first prove the smooth version of (iii), the piecewise smooth version will be given later. Symbols $C_1, C_2, C_3, \ldots$ denote positive constants.

Set $b_\ell = \int g\,\phi_\ell$ and $b_{k\ell} = \int g\,\psi_{k\ell}$. Then

$$g(x) = \sum_\ell b_\ell \phi_\ell(x) + \sum_k \sum_\ell b_{k\ell}\psi_{k\ell}(x).$$

Since ϕ and ψ are compactly supported on $[0,1]$, it follows from the orthogonality of the wavelet basis functions that

$$\int_Q (\hat{g}-g)^2 = S_1 + S_2 + S_3 + S_4, \tag{5.1}$$

where

$$S_1 = \sum_{\ell=0}^{p-1} (\hat{b}_\ell - b_\ell)^2, \qquad S_2 = \sum_{k=0}^{q-1}\sum_{\ell=0}^{p_k-1} b_{k\ell}^2 I(|\hat{b}_{k\ell}| \le \delta),$$

$$S_3 = \sum_{k=0}^{q-1}\sum_{\ell=0}^{p_k-1} (\hat{b}_{k\ell} - b_{k\ell})^2 I(|\hat{b}_{k\ell}| > \delta), \qquad S_4 = \sum_{k=q}^{\infty}\sum_{\ell=0}^{p_k-1} b_{k\ell}^2.$$

It can be shown that

$$E|S_1 - E(S_1)| = E|S_1 - \sigma_p^2 n^{-1}| = o(n^{-1}p). \tag{5.2}$$

$$E\left|S_2 - \kappa^2(1-2^{-2r})^{-1} p^{-2r}\int_Q g^{(r)^2}\right| = o(p^{-2r}), \qquad n\to\infty. \tag{5.3}$$

(5.4) $E(S_3) = o\big(n^{-2r/(2r+1)}\big)$.

(5.5) $S_4 = O(p_q^{-2r}) = o\{\min\big(p^{-2r}, \delta^{4r/(2r+1)}\big)\}, \qquad p_q^{2r+1}\delta^2 \to \infty \quad \text{and} \quad q \to \infty.$

We conclude from (5.1)–(5.5) that

$$E\left|\int_Q (\hat{g}-g)^2 - \left\{\sigma_p^2 n^{-1} + \kappa^2(1-2^{-2r})^{-1}p^{-2r}\int_Q g^{(r)^2}\right\}\right| = o(n^{-1}p + p^{-2r}),$$
$p^{2r+1}\delta^2 \to \infty$.

This completes the proof of Theorem 3.1 (iii).

If $p^{2r+1}\,\delta^2$ is bounded then, by the same sequence of results, we obtain

$$\int_Q E(\hat{g}-g)^2 \ge n^{-1}p + C_1\,\delta^{4r/(2r+1)} + o\big(n^{-2r/(2r+1)} + \delta^{4r/(2r+1)}\big).$$

When $p^{2r+1}\,\delta^2$ is bounded,

$$p = O(\delta^{-2/(2r+1)}) = O\{(n/\log n)^{1/(2r+1)}\},$$

so that $n^{-1}p = o(n^{-2r/(2r+1)})$ and

$$\int_Q E(\hat{g}-g)^2 \ge C_1\,\delta^{4r/(2r+1)} + o\big(n^{-2r/(2r+1)} + \delta^{4r/(2r+1)}\big).$$

It follows that

$$\inf_{p:p^{2r+1}\delta^2 \le C_2} \int_Q E(\hat{g}-g)^2 \ge C_3\delta^{4r/(2r+1)}, \qquad C_2 > 0.$$

This completes the proof of Theorem 3.1 (iv).

To prove (v), we need the following uniform consistency result for wavelet density estimators. (Its proof is given in Subsection A.5)

$$\sup_{x\in Q} |\hat{f}(x) - f(x)| = o_p(1). \tag{5.6}$$

Now write

$$(\hat{\mu}-\mu) - \{f^{-1}(\hat{g}-\mu\hat{f})\} = f^{-1}(\hat{g}-\mu\hat{f})(\hat{f}^{-1}f - 1).$$

By Condition 3.5 and (5.6), the squared integral of the latter term equals $o_p(n^{-2r/(2r+1)})$ under optimal choice of the smoothing parameters in $\hat{f}$ and $\hat{g}$. Therefore,

$$\int_Q (\hat{\mu}-\mu)^2 = \{1 + o_p(1)\}\int_Q f^{-2}(\hat{g}-\mu\hat{f})^2. \tag{5.7}$$

Set $\hat{c}_\ell = \int \mu\hat{f}\phi_\ell$ and $\hat{c}_{k\ell} = \int \mu\hat{f}\psi_{k\ell}$. In this notation,

$$\hat{g} - \mu\hat{f} = \sum_{\ell=0}^{p-1}(\hat{b}_\ell - \hat{c}_\ell)\phi_\ell + \sum_{k=0}^{q-1}\sum_{\ell=0}^{p_k-1}\{\hat{b}_{k\ell}I(|\hat{b}_{k\ell}| > \delta) - \hat{c}_{k\ell}\}\psi_{k\ell} + \sum_{k=0}^{q-1}\sum_{\ell=0}^{p_k-1}\hat{c}_{k\ell}\psi_{k\ell}. \tag{5.8}$$

It follows from (5.7) and (5.8) that

$$\int_Q (\hat{\mu}-\mu)^2 = \{1+o_p(1)\}\left[\sum_{\ell=0}^{p-1}(\hat{b}_\ell - \hat{c}_\ell)^2 f(\ell/p)^{-2} + \sum_{k=0}^{q-1}\sum_{\ell=0}^{p_k-1}\{\hat{b}_{k\ell}I(|\hat{b}_{k\ell}|>\delta) - \hat{c}_{k\ell}\}^2 f(\ell/p_k)^{-2} + \sum_{k=0}^{q-1}\sum_{\ell=0}^{p_k-1}\hat{c}_{k\ell}^2 f(\ell/p_k)^{-2}\right].$$

Let $\sigma^2(X)$ denote the conditional variance of Y given X. If $\hat{f}$ is a wavelet density estimator with $p \sim n^{1/(2r+1)}$, then

$$\sum_\ell(\hat{b}_\ell - \hat{c}_\ell)^2 f(\ell/p)^{-2} = n^{-1}p\int_Q \sigma^2 f + \sum_\ell \left(E\hat{b}_\ell - E\hat{c}_\ell\right)^2 f(\ell/p)^{-2} + o_p(n^{-2r/(2r+1)}).$$

Similarly,

$$\sum_{k=0}^{q-1}\sum_{\ell=0}^{p_k-1}\{\hat{b}_{k\ell}I(|\hat{b}_{k\ell}|>\delta) - \hat{c}_{k\ell}\}^2 f(\ell/p_k)^{-2} + \sum_{k=0}^{q-1}\sum_{\ell=0}^{p_k-1}\hat{c}_{k\ell}^2 f(\ell/p_k)^{-2}$$
$$= \sum_{k=0}^{q-1}\sum_{\ell=0}^{p_k-1}\{E\hat{b}_{k\ell}I(|\hat{b}_{k\ell}|>\delta) - E\hat{c}_{k\ell}\}^2 f(\ell/p_k)^{-2} + \sum_{k=0}^{q-1}\sum_{\ell=0}^{p_k-1}(E\hat{c}_{k\ell})^2 f(\ell/p_k)^{-2} + o_p(n^{-2r/(2r+1)}).$$

Hence, $\int_Q(\hat{\mu}-\mu)^2 = O_p(n^{-2r/(2r+1)})$.

To complete the proof, we need to argue that the above results also hold for the piecewise smooth case. We sketch the argument for the case of density estimate. By the orthogonality properties of ϕ and ψ, $\int(\hat{f}-f)^2 = I_q(\boldsymbol{Z},\boldsymbol{Z},\ldots)$, where $\boldsymbol{Z}$ denotes the set of all integers and

$$I_q(\mathcal{L},\mathcal{L}_0,\ldots) = \sum_{\ell\in\mathcal{L}}(\hat{b}_\ell - b_\ell)^2 + \sum_{k=0}^{q-1}\sum_{\ell\in\mathcal{L}_k}(\hat{b}_{k\ell}-b_{k\ell})^2\, I(|\hat{b}_{k\ell}|>\delta) + \sum_{k=0}^{q-1}\sum_{\ell\in\mathcal{L}_k} b_{k\ell}^2\, I(|\hat{b}_{k\ell}|\le\delta) + \sum_{k=q}^{\infty}\sum_{\ell\in\mathcal{L}_k} b_{k\ell}^2.$$

Let $\mathcal{X}$ denote the finite set of points where $f^{(s)}$ has a discontinuity for some $0\le s\le r$. If $\operatorname{supp}\psi \subseteq (-v,v)$ then, unless

$$\ell\in\mathcal{K}_k = \{j : j\in(p_k\,x - v, p_k\,x+v) \text{ for some } x\in\mathcal{X}\},$$

both $b_{k\ell}$ and $\hat{b}_{k\ell}$ are constructed entirely from an integral over or an average of data values from an interval where $f^{(r)}$ exists and is bounded. Likewise, if $\operatorname{supp}\phi \subseteq (-v,v)$ then, unless

$$\ell\in\mathcal{K} = \{j : j\in(px - v, px+v) \text{ for some } x\in\mathcal{X}\},$$

b_ℓ and $\hat{b}_\ell$ are also constructed solely from such regions. We may write $\int(\hat{f}-f)^2 = I_q(\mathcal{K},\mathcal{K}_1,\ldots) + I_q(\widetilde{\mathcal{K}},\widetilde{\mathcal{K}}_1,\ldots)$, where $\widetilde{\mathcal{K}}_k$ denotes the complement of $\mathcal{K}_k$ in $\boldsymbol{Z}$. The argument used so far may be employed to prove that $I_q(\widetilde{\mathcal{K}},\widetilde{\mathcal{K}}_1,\ldots)$ has precisely the asymptotic properties claimed for $\int(\hat{f}-f)^2$. Furthermore, noting that both $\mathcal{K}$ and $\mathcal{K}_k$ have no more than $(2v+1)(\#\mathcal{X})$ elements, for each k, and that $q = O(\log n)$ and $p_q^{-1} = o(n^{-2r/(2r+1)})$ we may show that $E\{I_q(\mathcal{K},\mathcal{K}_1,\ldots)\} = o(n^{-1}p + p^{-2r})$ in the context of part (i) of Theorem 3.1, or $= o(\delta^{4r/(2r+1)})$ in the context of part (ii). Combining these results we obtain Theorem 3.1 for the piecewise smooth case. (The condition $p_q^{-1} = o(n^{-2r/(2r+1)})$ is used to establish negligibility of the series $\sum_{k\geq q}\sum_\ell b_{k\ell}^2$; note that at discontinuities, $b_{k\ell}^2$ is of size p_k^{-1}.) This completes the proof of Theorem 3.1.

Acknowledgements

We are grateful to Professor Peter Hall for his generous support and comments on our research. We thank the referees for comments that improve the readability of the paper.

Appendix

This section contains more detailed arguments given in the previous section. The justifications of (5.2)–(5.5) are given in Subsections A.1 to A.4, respectively. Subsection A.5 presents a proof of the uniform consistency for wavelet based density estimator, which is needed in showing the rates of convergence of the regression estimate. The last section gives the proof of the variance term (A.2).

A.1 *Proof of* (5.2)

Set $\mu_2(x) = E(Y_1^2 \mid X_1 = x)$. By Conditions 3.5 and 3.6, and the boundedness of $|\phi|$,

$$b_\ell = p^{-1/2}\int g\{p^{-1}(x+\ell)\}\phi(x)\,dx = O(p^{-1/2}), \qquad \ell = 0,1,\ldots,p;$$

$$E\{Y_1\phi_\ell(X_1)\}^2 = \int\{\phi(x)\}^2\mu_2\{p^{-1}(x+\ell)\}f\{p^{-1}(x+\ell)\}dx = O(1),$$
$$\ell = 0,1,\ldots,p.$$

Let f_i denote the density function of (X_1, X_{1+i}) and set $h_i(x_1,x_2) = f_i(x_1,x_2) - f(x_1)f(x_2)$. Then, by Conditions 3.1 and 3.2,

$$\left|\operatorname{cov}\{Y_1\phi_\ell(X_1), Y_{1+i}\phi_\ell(X_{1+i})\}\right| = O(p^{-1}), \qquad \ell = 0,1,\ldots,p.$$

On the other hand, by Corollary A.1 of Hall and Heyde ((1980), p. 277),

$$|\operatorname{cov}\{Y_1\phi_\ell(X_1), Y_{1+i}\phi_\ell(X_{1+i})\}| = O(p\alpha(i)).$$

From $E[Y_1\phi_\ell(X_1)] = b_\ell$ and Condition 3.3, for $M_n \to \infty$ such that $M_n = o(p)$,

$$\begin{aligned} nE(\hat{b}_\ell - b_\ell)^2 &= \operatorname{var}(Y_1\phi_\ell(X_1)) + 2\sum_{i=1}^{n-1}(1-n^{-1}i)\operatorname{cov}\{Y_1\phi_\ell(X_1), Y_{1+i}\phi_\ell(X_{1+i})\} \\ &= \operatorname{var}(Y_1\phi_\ell(X_1)) + C_1\left(\sum_{i\leq M_n} + \sum_{i>M_n}\right)\max\{p^{-1}, p\alpha(i)\}. \end{aligned}$$

Recall that $\sigma_p^2 = \sum_{\ell=0}^{p} \operatorname{var}(Y_1\phi_\ell(X_1))$. Then

$$\sum_{\ell=0}^{p} E(\hat{b}_\ell - b_\ell)^2 = \sigma_p^2 n^{-1} + o(n^{-1}p). \tag{A.1}$$

From $E\{Y_1\phi_\ell(X_1)\}^2 = O(1)$, $\ell = 0, 1, \ldots, p$, we have

$$\sigma_p^2 = \sum_{\ell=0}^{p} \operatorname{var}(Y_1\phi_\ell(X_1)) = O(p).$$

Furthermore (a proof will be given shortly),

$$\operatorname{var}\left\{\sum_{\ell=0}^{p}(\hat{b}_\ell - b_\ell)^2\right\} = o(n^{-2}p^2), \tag{A.2}$$

which together with (A.1) imply that

$$E\,|S_1 - E(S_1)| = E\,|S_1 - \sigma_p^2 n^{-1}| = o(n^{-1}p). \tag{A.3}$$

A.2 *Proof of* (5.3)

Let $\varepsilon > 0$, and define

$$S_{21} = \sum_{k=0}^{q-1}\sum_{\ell=0}^{p_k-1} b_{k\ell}^2 I\{|b_{k\ell}| \le (1+\varepsilon)\delta\}, \qquad S_{22} = \sum_{k=0}^{q-1}\sum_{\ell=0}^{p_k-1} b_{k\ell}^2 I\{|b_{k\ell}| \le (1-\varepsilon)\,\delta\},$$

$$S_{23} = \sum_{k=0}^{q-1}\sum_{\ell\in\mathcal{J}} b_{k\ell} I\big(|b_{k\ell}| \le \tfrac{1}{2}\,\delta\big),$$

$$\Delta_1 = \sum_{k=0}^{q-1}\sum_{\ell=0}^{p_k-1} b_{k\ell}^2 I(|\hat{b}_{k\ell} - b_{k\ell}| > \varepsilon\delta), \qquad \Delta_2 = \sum_{k=0}^{q-1}\sum_{\ell\in\mathcal{J}} b_{k\ell}^2 I(|\hat{b}_{k\ell} - b_{k\ell}| > |b_{k\ell}|),$$

for any set $\mathcal{J}$ of integers $\ell \in \{0, 1, \ldots, p_k - 1\}$. It follows from

$$I(|\hat{b}_{k\ell}| \le \delta) \le I\{|b_{k\ell}| \le (1+\varepsilon)\,\delta\} + I(|\hat{b}_{k\ell} - b_{k\ell}| > \varepsilon\delta)$$

and

$$I\{|b_{k\ell}| \le (1-\varepsilon)\,\delta\} \le I(|\hat{b}_{k\ell}| \le \delta) + I(|\hat{b}_{k\ell} - b_{k\ell}| > \varepsilon\delta)$$

that

$$S_{22} - \Delta_1 \le S_2 \le S_{21} + \Delta_1. \tag{A.4}$$

Since $I\big(|b_{k\ell}| \le \frac{1}{2}\,\delta\big) \le I(|\hat{b}_{k\ell}| \le \delta) + I(|\hat{b}_{k\ell} - b_{k\ell}| > |b_{k\ell}|)$,

$$S_{23} - \Delta_2 \le S_2. \tag{A.5}$$

It follows from the Taylor expansion

$$\int g(p_k^{-1}(y+\ell))\psi(y)dy$$
$$= \int \psi(y)\Bigg[\sum_{v=0}^{r-1}(v!)^{-1}(y/p_k)^v\, g^{(v)}(\ell/p_k)$$
$$+ \{(r-1)\}^{-1}(y/p_k)^r \int_0^1 (1-t)^{r-1} g^{(r)}\{(\ell+ty)/p_k\}dt\Bigg]dy$$

that, for C_0 sufficiently large,

$$b_{k\ell} = \kappa\, p_k^{-\{r+(1/2)\}}(g_{k\ell} + \xi_{k\ell}), \tag{A.6}$$

where $g_{k\ell} = g^{(r)}(\ell/p_k)$ and

$$\sup_{\substack{0\le\ell\le p_k-1\\ 0\le k\le q-1}} |\xi_{k\ell}| \to 0.$$

We shall assume that

$$p^{2r+1}\,\delta^2 \to \zeta, \tag{A.7}$$

where $0 \le \zeta \le \infty$. (The case where such convergence is only along a subsequence may be treated similarly.) Suppose first that $\zeta < \infty$. Let $C_2, C_3 > 0$ be such that the set $\mathcal{J}'$ of integers $\ell \in \{0,1,\dots,p_k-1\}$ with $|g_{k\ell}| > 2|\kappa|^{-1}C_2$, has at least $2C_3p_k$ elements for all $k \ge 0$ and all sufficiently large n. Then a certain subset $\mathcal{J}$ of $\mathcal{J}'$ consists entirely of integers ℓ such that $|b_{k\ell}| > C_2\, p_k^{-(r+(1/2))}$, and has between C_3p_k and $2C_3p_k$ elements for all $k \ge 0$ and all large n. We shall use this $\mathcal{J}$ in the definition of S_{23} and Δ_2. Note too that for some $C_4 > 0$ and all k and ℓ, $|b_{k\ell}| \le \frac{1}{2}C_4^{1/2}p_k^{-(r+(1/2))}$. Thus,

$$\begin{aligned} S_{23} &\ge \sum_{k=0}^{q-1}\sum_{j\in\mathcal{J}} C_2^2\, p_k^{-(2r+1)} I(C_4p_k^{-(2r+1)} \le \delta^2) \\ &\ge C_2^2C_3\sum_{k=0}^{q-1} p_k^{-2r} I\{p_k \ge (C_4/\delta^2)^{1/(2r+1)}\}. \end{aligned}$$

It follows from our assumption $p = O(\delta^{-2/(2r+1)})$ that

$$S_{23} \ge C_5\, \delta^{4r/(2r+1)}. \tag{A.8}$$

Furthermore, by Bernstein's inequality (see Truong (1994)) and (A.6),

$$E(\Delta_2) \le 2\sum_{k=0}^{q-1}\sum_{j\in\mathcal{J}} b_{k\ell}^2 \exp(-C_6nb_{k\ell}^2) = O(n^{-2r/(2r+1)}). \tag{A.9}$$

Combining (A.5), (A.8) and (A.9), and noting that by our assumption, $\delta \ge C(n^{-1} \log n)^{1/2}$, we deduce that for sufficiently large n,

$$E(S_2) \ge S_{23} - E(\Delta_2) \ge C_9\delta^{4r/(2r+1)}. \tag{A.10}$$

Suppose next that in (A.7), $\zeta = \infty$. Then, using (A.6),

$$\sup_{\ell} |b_{k\ell}| \le C_{10}\, p_k^{-(r+(1/2))} \le C_{10}p^{-(r+(1/2))} \ll \delta,$$

whence it follows that for all sufficiently large n,

$$\begin{aligned} S_{21} = S_{22} &= \sum_{k=0}^{q-1}\sum_{\ell=0}^{p_k-1} \kappa^2 p_k^{-(2r+1)}(g_{k\ell} + \xi_{k\ell})^2 \\ &= \kappa^2(1-2^{-2r})^{-1}\, p^{-2r} \int_Q {g^{(r)}}^2 + o(p^{-2r}). \end{aligned}$$

Therefore,

$$S_{21} = S_{22} \sim \kappa^2(1-2^{-2r})^{-1}\, p^{-2r} \int_Q g^{(r)^2}. \tag{A.11}$$

By Bernstein's inequality (see Truong (1994)) and for all sufficiently large n,

$$E(\Delta_1) \le 2\sum_{k=0}^{q-1}\sum_{\ell=0}^{p_k-1} b_{k\ell}^2 \exp(-C_{20} n\delta^2) = o\left(\sum_{k=0}^{q-1}\sum_{\ell=0}^{p_k-1} b_{k\ell}^2\right) = o(S_{21}), \tag{A.12}$$

the second-last identity following since $n\delta^2 \to \infty$. By (A.4), (A.11) and (A.12),

$$E\left|S_2 - \kappa^2(1-2^{-2r})^{-1}\, p^{-2r} \int_Q g^{(r)^2}\right| = o(p^{-2r}), \qquad n \to \infty. \tag{A.13}$$

A.3 *Proof of* (5.4)

Let α, β denote positive numbers satisfying $\alpha + \beta = 1$, and set

$$S_{31} = \sum_{k=0}^{q-1}\sum_{\ell=0}^{p_k-1} E\{(\hat{b}_{k\ell} - b_{k\ell})^2\} I(|b_{k\ell}| > \alpha\delta),$$
$$S_{32} = \sum_{k=0}^{q-1}\sum_{\ell=0}^{p_k-1} E\{(\hat{b}_{k\ell} - b_{k\ell})^2 I(|\hat{b}_{k\ell} - b_{k\ell}| > \beta\delta)\}.$$

Since $I(|\hat{b}_{k\ell}| > \delta) \le I(|b_{k\ell}| > \alpha\delta) + I(|\hat{b}_{k\ell} - b_{k\ell}| > \beta\delta)$ then

$$E(S_3) \le S_{31} + S_{32}. \tag{A.14}$$

We shall bound S_{31} and S_{32}, in turn. By argument similar to that leading to (A.1), we have

$$E(\hat{b}_{k\ell} - b_{k\ell})^2 = O(n^{-1}),$$

and by (A.6) with $n^{1/2}\delta \to \infty$,

$$S_{31} = O(n^{-1})\sum_{k=0}^{q-1} p_k I\{p_k \le (C_2/\delta)^{2/(2r+1)}\} = o\big(n^{-2r/(2r+1)}\big). \tag{A.15}$$

Let a, b denote positive numbers satisfying $a^{-1} + b^{-1} = 1$. By Lemma 9 of Truong and Stone (1992),

$$E|\hat{b}_{k\ell} - b_{k\ell}|^{2a} = n^{-2a} E\left|\sum_i \{Y_i \psi_{k\ell}(X_i) - b_{k\ell}\}\right|^{2a} \le n^{-2a} p_k^a B_n^{2a} (np_k^{-1})^a = n^{-a} B_n^{2a}.$$

By Bernstein's inequality (see Truong (1994)), for n sufficiently large,

$$P(|\hat{b}_{k\ell} - b_{k\ell}| > \beta\delta) \le 2\exp\{-C_{11}\beta^2 n\delta^2\}$$

uniformly in $0 \le k \le q-1$ and ℓ. Hence, by Hölder's inequality,

$$S_{32} = O\left[\sum_{k=0}^{q-1} n^{-1} B_n^2 \sum_{\ell=0}^{p_k-1} \exp\{-C_{11}\beta^2 b^{-1} n\delta^2\}\right]. \tag{A.16}$$

Now select $\delta \geq C(n^{-1}\log n)^{1/2}$ with C chosen such that $CC_{11}b^{-1}\beta^2 = \gamma = 2r/(2r+1)$. Therefore by (A.16), and since $p_q\delta^2 = o(1)$ and $n^{1/2}\delta \to \infty$, we have

$$S_{32} = O\left\{\sum_{k=0}^{q-1} n^{-1}n^{-\gamma}p_k\right\} = o(n^{-2r/(2r+1)}). \tag{A.17}$$

Combining (A.14), (A.15) and (A.17) we deduce that

$$E(S_3) = o(n^{-2r/(2r+1)}). \tag{A.18}$$

A.4 *Proof of* (5.5)

This follows from $p_q^{2r+1}\delta^2 \to \infty$ and $q \to \infty$ as $n \to \infty$, and

$$S_4 = \sum_{i=q}^{\infty}\sum_{\ell=0}^{p_k-1} \kappa^2 p_k^{-(2r+1)}(g_{k\ell}+\xi_{k\ell})^2 \leq 2\kappa^2 \sum_{k=q}^{\infty} p_k^{-(2r+1)} \sum_{\ell=0}^{p_k-1} g_{k\ell}^2. \tag{A.19}$$

A.5 *Proof of* (5.6)

Since $\phi(\cdot)$ and $\psi(\cdot)$ are compactly supported,

$$\sup_Q |\hat{f} - f| \leq s_1 + s_2 + s_3 + s_4,$$

where

$$s_1 = \sum_{\ell=0}^{p-1} |\hat{b}_\ell - b_\ell|\, \|\phi_\ell\|_\infty, \qquad s_2 = \sum_{k=0}^{q-1}\sum_{\ell=0}^{p_k-1} |b_{k\ell}|I(|\hat{b}_{k\ell}| \leq \delta)\mid \|\psi_{k\ell}\|_\infty,$$

$$s_3 = \sum_{k=0}^{q-1}\sum_{\ell=0}^{p_k-1} |\hat{b}_{k\ell} - b_{k\ell}|I(|\hat{b}_{k\ell}| > \delta)\|\psi_{k\ell}\|_\infty, \qquad s_4 = \sum_{k=q}^{\infty}\sum_{\ell=0}^{p_k-1} |b_{k\ell}|\, \|\psi_{k\ell}\|_\infty.$$

Note that $\|\phi_\ell(\cdot)\|_\infty = O(p^{1/2})$ and $\|\psi_{k\ell}(\cdot)\|_\infty = O(p_k^{1/2})$. By Hölder's inequality and (A.2),

$$s_1 \leq p\left\{\sum_{\ell=0}^{p-1}(\hat{b}_\ell - b_\ell)^2\right\}^{1/2} = p\left\{O_p(n^{-1}p)\right\}^{1/2} = o_p(1). \tag{A.20}$$

Similarly, it follows from $p_q\delta^2 = O(n^{-\epsilon})$ with $(r+1)/(2r+1) < \epsilon < 2r/(2r+1)$, (A.13) and (A.17) that

$$s_2 \leq q^{1/2}p_q\left\{O_p(p^{-2r})\right\}^{1/2} = o_p(1), \tag{A.21}$$

$$s_3 \leq q^{1/2}p_q\left\{O_p(p^{-2r})\right\}^{1/2} = o_p(1). \tag{A.22}$$

According to the argument leading to (A.19), and Hölder's inequality,

$$s_4 = o(1). \tag{A.23}$$

The desired result follows from (A.20)–(A.23). This completes the proof of (5.6).

A.6 *Proof of* (A.2)

Put $Z_{\ell i} = p^{1/2} Y_i \phi(pX_i - \ell) - b_\ell$. Then

$$n^2 \sum_\ell (\hat{b}_\ell - b_\ell)^2 = \sum_i \sum_\ell Z_{\ell i}^2 + \sum_{i_1 \neq i_2} \sum_\ell Z_{\ell i_1} Z_{\ell i_2}. \tag{A.24}$$

Now,

$$n^{-1} E\left\{ \sum_i \sum_\ell (Z_{\ell i}^2 - EZ_{\ell i}^2) \right\}^2$$
$$= E\left\{ \sum_\ell (Z_{\ell 1}^2 - EZ_{\ell 1}^2) \right\}^2$$
$$+2 \sum_i (1 - i/n) E\left[\left\{ \sum_\ell (Z_{\ell 1}^2 - EZ_{\ell 1}^2) \right\} \left\{ \sum_\ell (Z_{\ell,1+i}^2 - EZ_{\ell,1+i}^2) \right\} \right],$$

and

$$\tfrac{1}{2} \sum_\ell Z_{\ell 1}^2 \le p \sum_\ell Y_1^2 \phi(pX_1 - \ell)^2 + \sum_\ell b_\ell^2 \le p(\sup \phi^2)(\mathrm{supp}\phi + 2) + \int_Q g^2.$$

It follows that

$$\left| \sum_i E\left[\left\{ \sum_\ell (Z_{\ell 1}^2 - EZ_{\ell 1}^2) \right\} \left\{ \sum_\ell (Z_{\ell,1+i}^2 - EZ_{\ell,1+i}^2) \right\} \right] \right| = O\left\{ p^2 \sum_i \alpha(i) \right\}.$$

Therefore,

$$\mathrm{var}\left(n^{-2} \sum_i \sum_\ell Z_{\ell i}^2 \right) = O(n^{-3} p^2) = o(n^{-2} p^2). \tag{A.25}$$

Next, we consider an upper bound for

$$E\left(\sum_{i_1 \neq i_2} \sum_\ell Z_{\ell i_1} Z_{\ell i_2} \right)^2 = \sum_{\ell_1} \sum_{\ell_2} \sum_{i_{11} \neq i_{12}} \sum_{i_{21} \neq i_{22}} E(Z_{\ell_1 i_{11}} Z_{\ell_1 i_{12}} Z_{\ell_2 i_{21}} Z_{\ell_2 i_{22}}).$$

This can be obtained by considering several cases of the indices.

Case 1. Suppose the indices satisfy $i_{11} = i_{21} = i_1$ and $i_{12} = i_{22} = i_2$. It follows from $\mathrm{supp}\phi \subseteq (-v, v)$ and the mixing property that

$$\sum_{\ell_1} \sum_{\ell_2} \sum_{i_1 \neq i_2} E(Z_{\ell_1 i_1} Z_{\ell_1 i_2} Z_{\ell_2 i_1} Z_{\ell_2 i_2}) = O(n^2 p). \tag{A.26}$$

Case 2. Suppose the indices satisfy $i_{11} = i_{21}$ and $i_{11} < i_{12} < i_{22}$. According to Corollary A.1 of Hall and Heyde (1980, p. 277) and Condition 3.6,

$$\sum_{i_{11} < i_{12} < i_{22}} \sum_{\ell_1} \sum_{\ell_2} E(Z_{\ell_1 i_{11}} Z_{\ell_1 i_{12}} Z_{\ell_2 i_{11}} Z_{\ell_2 i_{22}}) \tag{A.27}$$
$$\le O(1) \sum_{\ell_1} \sum_{\ell_2} \sum_{i_{11} < i_{12} < i_{22}} \max\{p^{-1}, 4p^2 \alpha(i_{22})\} = O(n^2 p^2).$$

Case 3. Suppose the indices satisfy $i_{11} < i_{12} < i_{21} < i_{22}$. Again, by Corollary A.1 of Hall and Heyde (1980, p. 277) and Condition 3.6,

$$\sum\sum_{i_1<i_2} E(Z_{\ell_1 i_1} Z_{\ell_1 i_2}) \le O(1) \sum_{i_1}\sum_{i_2} \max\{p^{-1}, p\alpha(i_2)\} = O(n),$$
$$E(Z_{\ell_1 i_{11}} Z_{\ell_1 i_{12}} Z_{\ell_2 i_{21}} Z_{\ell_2 i_{22}}) \le E(Z_{\ell_1 i_{11}} Z_{\ell_1 i_{12}}) E(Z_{\ell_2 i_{21}} Z_{\ell_2 i_{22}}) + O(1) p^2 \alpha(i_{21})$$

and, since $E(Z_{\ell_1 i_{11}}) = 0$,

$$E(Z_{\ell_1 i_{11}} Z_{\ell_1 i_{12}} Z_{\ell_2 i_{21}} Z_{\ell_2 i_{22}}) \le O(1) p^2 \alpha(i_{12}).$$

Using an argument similar to that in Lemma 9 of Truong and Stone (1992), we have

$$\sum\sum\sum\sum_{i_{11}<i_{12}<i_{21}<i_{22}} \sum_{\ell_1}\sum_{\ell_2} E(Z_{\ell_1 i_{11}} Z_{\ell_1 i_{12}} Z_{\ell_2 i_{21}} Z_{\ell_2 i_{22}}) = o(n^2 p^2). \tag{A.28}$$

The desired result follows from (A.24)–(A.28).

References

Auestad, B. and Tjøstheim, D. (1990). Identification of nonlinear time series: First order characterization and order determination, *Biometrika*, **77**, 669–687.

Daubechies, I. (1992). *Ten Lectures on Wavelets*, SIAM, Philadelphia.

Donoho, D. L. (1995). Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition, *Appl. Comput. Harmon. Anal.*, **2**, 101–126.

Donoho, D. L. and Johnstone, I. M. (1994*a*). Minimax risk over ℓ_p-balls for ℓ_q-error, *Probab. Theory Related Fields*, **99**, 277–303.

Donoho, D. L. and Johnstone, I. M. (1994*b*). Ideal spatial adaptation by wavelet shrinkage, *Biometrika*, **81**, 425–455.

Donoho, D. L. and Johnstone, I. M. (1998). Minimax estimation via wavelet shrinkage, *Ann. Statist.*, **26**, 879–921.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage, Asymptopia? (with Discussion), *J. Roy. Statist. Soc. Ser. B*, **57**, 301–369.

Gorodetskii, V. V. (1977). On the strong mixing property for linear sequences, *Theory Probab. Appl.*, **22**, 411–413.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theorems*, Academic Press, New York.

Hall, P. and Patil, P. (1995*a*). On wavelet methods for estimating smooth functions, *Bernoulli*, **1**, 42–58.

Hall, P. and Patil, P. (1995*b*). Formulae for mean integrated squared error of nonlinear wavelet-based density estimators, *Ann. Statist.*, **23**, 905–928.

Hall, P. and Patil, P. (1996*a*). On the choice of smoothing parameter, threshold and truncation in nonparametric regression by nonlinear wavelet methods, *J. Roy. Statist. Soc. Ser. B*, **58**, 361–377.

Hall, P. and Patil, P. (1996*b*). Effect of threshold rules on performance of wavelet-based curve estimators, *Statistica Sinica*, **6**, 331–345.

Härdle, W., Klinke, S. and Turlach, B. A. (1995). *XploRe-an Interactive Statistical Computing Environment*, Springer, New York.

Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, Approximation, and Statistical Applications*, Springer, New York.

Johnstone, I. M. (1999). Wavelet threshold estimators for correlated data and inverse problems: Adaptivity results, *Statistica Sinica*, **9**, 51–83.

Johnstone, I. M. and Silverman, B. W. (1997). Wavelet threshold estimators for data with correlated noise, *J. Roy. Statist. Soc. Ser. B*, **59**, 319–351.

Kerkyacharian, G. and Picard, D. (1992). Density estimation in Besov spaces, *Statist. Probab. Lett.*, **13**, 15–24.

Kerkyacharian, G. and Picard, D. (1993*a*). Density estimation by kernel and wavelet methods: Optimality of Besov spaces, *Statist. Probab. Lett.*, **18**, 327–336.

Kerkyacharian, G. and Picard, D. (1993*b*). Linear wavelet methods and other periodic kernel methods, Manuscript.

Kerkyacharian, G. and Picard, D. (1993*c*). *Introduction aux Ondelettes et Estimation de la Densité, 1: Introduction aux Ondelettes et a l'Analyse Multiresolution*, Lecture Notes.

Li, Y. and Xie, Z. (1999). The wavelet identification of thresholds and time delay of threshold autoregressive models, *Statistica Sinica*, **9**, 153–166.

Neumann, M. H. and von Sachs, R. (1997). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra, *Ann. Statist.*, **25**, 38–76.

Priestley, M. B. (1988). *Non-linear and Non-stationary Time Series.*, Academic Press, New York.

Strang, G. (1989). Wavelets and dilation equations: A brief introduction, *SIAM Rev.*, **31**, 614–627.

Tjøstheim, D. (1994). Non-linear time series: A selected review, *Scand. J. Statist.*, **21**, 97–130.

Tong, H. (1983). *Threshold Models in Non-linear Time Series Analysis*, Lecture Notes in Statist., **21**, Springer, New York.

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*, Oxford University Press, New York.

Tran, L. T. (1993). Nonparametric function estimation for time series by local average estimators, *Ann. Statist.*, **21**, 1040–1057.

Truong, Y. K. (1994). Nonparametric time series regression, *Ann. Inst. Statist. Math.*, **46**, 279–293.

Truong, Y. and Patil, P. (1996). On a local property of wavelet estimators involving time series, *J. Nonparametr. Statist.*, **6**, 143–156.

Truong, Y. K. and Stone, C. J. (1992). Nonparametric function estimation involving time series, *Ann. Statist.*, **20**, 77–97.

Truong, Y. K. and Stone, C. J. (1994). Semi-parametric time series regression, *J. Time Ser. Anal.*, **15**, 405–428.

Wang, Y. (1996). Function estimation via wavelet shrinkage for long-memory data, *Ann. Statist.*, **24**, 466–484.

Withers, C. S. (1981). Conditions for linear processes to be strong mixing, *Z. Wahrsch. Verw. Gebiete*, **57**, 477–480.