Acta Mathematica Sinica, English Series
June, 2005, Vol. 21, No. 3, pp. 585–592
Published online: Apr. 14, 2005
DOI: 10.1007/s10114-005-0528-7
Http://www.ActaMath.com

Acta Mathematica Sinica, English Series

Robust Depth-Weighted Wavelet for Nonparametric Regression Models

Lu LIN
School of Mathematics and Systems Science, Shandong University, Ji'nan 250100*, P. R. China*
E-mail: linlu@sdu.edu.cn

Abstract In the nonparametric regression models, the original regression estimators including kernel estimator, Fourier series estimator and wavelet estimator are always constructed by the weighted sum of data, and the weights depend only on the distance between the design points and estimation points. As a result these estimators are not robust to the perturbations in data. In order to avoid this problem, a new nonparametric regression model, called the depth-weighted regression model, is introduced and then the depth-weighted wavelet estimation is defined. The new estimation is robust to the perturbations in data, which attains very high breakdown value close to 1/2. On the other hand, some asymptotic behaviours such as asymptotic normality are obtained. Some simulations illustrate that the proposed wavelet estimator is more robust than the original wavelet estimator and, as a price to pay for the robustness, the new method is slightly less efficient than the original method.

Keywords Nonparametric regression, Wavelet, Statistical depth, Robustness
MR(2000) Subject Classification 62G08, 62G20, 62G35

1 Introduction

We first consider the following random design nonparametric regression model:

$$Y_i = r(X_i) + \varepsilon_i,\ i = 1, \dots, n, \tag{1}$$

where data $(X_1, Y_1), \dots, (X_n, Y_n)$ are independent and distributed as (X, Y) with $r(x) = E(Y|X = x)$, $x \in [a, b]$ and $\varepsilon_i = Y_i - r(X_i)$.

So far many methods concerning the regression function $r(\cdot)$ have been proposed such as kernel estimator, Fourier series estimator and wavelet estimator (Hart [1]). In this paper, for convenience, we only consider wavelet estimation in what follows and the results obtained can be extended to other cases.

As shown by Antoniadis, Gregoire and Mckeague [2], Härdle, et al. [3], in the class of wavelet estimations, a standard estimator of regression function $r(\cdot)$ is defined as

$$\tilde{r}(t) = \frac{1}{n}\sum_{i=1}^{n} Y_i E_m(t, X_i)/\tilde{f}(t), \tag{2}$$

where $\tilde{f}(\cdot)$ is a wavelet estimator of the density of X given by

$$\tilde{f}(t) = \frac{1}{n}\sum_{i=1}^{n} E_m(t, X_i)$$

and $E_m(\cdot, \cdot)$ is wavelet kernel. An ordinary choice of $E_m(\cdot, \cdot)$ is

$$E_m(t, s) = 2^m \sum_{k\in\mathbf{Z}} \varphi(2^m t - k)\varphi(2^m s - k),$$

Received September 10, 2003, Accepted July 6, 2004
This paper is supported by NNSF Projects (10371059 and 10171051) of China

where $\mathbf{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ and $\varphi(\cdot)$ is a scaling function such as Haar wavelet scaling function defined by $\varphi(x) = 1$ for $x \in [0, 1]$, otherwise $\varphi(x) = 0$.

From (2) we can see that the estimator $\tilde{r}(t)$ of the regression function is the weighted sum of data and the weights depend only on the distance between the design points X_i and estimation points t. Then this estimation is sensitive to the data $(X_1, Y_1), \ldots, (X_n, Y_n)$. So it is not robust and then some new designs are desired.

Recent study on statistical depth showed that the estimators of some moments based on depth-weighted methods have some desirable robustness, which attain very high breakdown values close to 1/2 (Zuo, Cui and He [4]). In this paper we use depth-weighted method to define depth-weighted regression model and then construct depth-weighted wavelet estimation. The remainder of this paper is organized as follows. In Section 2 a depth-weighted regression model is defined through common regression model (1) and depth-weighted mean, and then a depth-weighted wavelet estimation is introduced by the defined depth-weighted regression model and wavelet kernel estimation (2). In Section 3, it is proven that this new estimator has very high breakdown value close to 1/2. Some asymptotic behaviours such as asymptotic normality are obtained in Section 4. The asymptotic normality gives a principle about how to choose weight function so as to derive an efficient estimation from the class of the robust depth-weighted wavelet estimations. Some simulations are given in Section 5 to illustrate that the proposed wavelet estimator is more robust than the original wavelet estimator and, as a price to pay for the robustness, the new method is slightly less efficient than the original method. Selected proofs of the theorems are presented in Section 6.

2 Depth-Weighted Wavelet Estimation

Statistical depths, including halfspace depth (Tukey [5]), simplicial depth (Liu [6]), regression depth (Rousseeuw and Hubert [7]) and projection depth (Zuo and Serfling [8, 9]), are powerful tools for the robust and nonparametric statistics. For convenience, in what follows, we only use the projection depth to define depth-weighted regression model and then construct robust depth-weighted regression estimation.

The projection depth of an observation vector Z based on distribution F is defined by

$$D(Z, F) = \frac{1}{1 + O(Z, F)}, \tag{3}$$

where $O(Z, F) = \sup_{||u||=1} |u'Z - \mathrm{med}(F_u)|/\mathrm{mad}(F_u)$, F_u is the distribution function of $u'Z$, $\mathrm{med}(F_u)$ is the median of F_u, $\mathrm{mad}(F_u)$ is the median of the distribution of $|u'Z - \mathrm{med}(F_u)|$. If $u'Z - \mathrm{med}(F_u) = \mathrm{mad}(F_u) = 0$, then we define $|u'Z - \mathrm{med}(F_u)|/\mathrm{mad}(F_u) = 0$. According to the notion of depth-weighted mean of a random variable proposed by Zuo, Cui and He [4], for given depth function $D(Z, F)$ and distribution function F of Z, the depth-weighted mean of a function $g(Z)$ is defined by

$$DE(g(Z)) = \frac{\int g(z)\omega(D(z, F))dF(z)}{\int \omega(D(z, F))dF(z)}, \tag{4}$$

where $\omega(\cdot)$ is a given weight function. As shown by Zuo, Cui and He [4], a common weight function is chosen as

$$\omega(s) = \begin{cases} \dfrac{\exp(-K(1 - \frac{s}{\mathrm{med}(D(Z,F))})^2) - \exp(-K)}{1 - \exp(-K)}, & \text{if } \ s < \mathrm{med}(D(Z, F)), \\ 1, & \text{if } \ s \geq \mathrm{med}(D(Z, F)), \end{cases} \tag{5}$$

where K is a suitable positive constant.

We now suppose that, for given weight function $\omega(\cdot)$, a regression function $r(\cdot)$ satisfies

$$r(x) = DE(Y|X = x), \ \ x \in [a, b]. \tag{6}$$

From (4), we can rewrite the regression function as

$$r(x) = \frac{\int y\,\omega(D((x,y),F(y|x)))dF(y|x)}{\int \omega(D((x,y),F(y|x)))dF(y|x)},\quad x \in [a,b], \tag{7}$$

where $F(y|x)$ is the conditional distribution function of Y given $X = x$. Then model (6) or (7) is called the depth-weighted regression model. Combining the above depth-weighted regression model with wavelet estimator (2), we define the depth-weighted wavelet estimator to regression function $r(\cdot)$ as

$$\hat{r}(t) = \frac{\frac{1}{n}\sum_{i=1}^{n} Y_i\,\omega(D((X_i,Y_i),F_n))E_m(t,X_i)}{\frac{1}{n}\sum_{i=1}^{n} \omega(D((X_i,Y_i),F_n))E_m(t,X_i)} \tag{8}$$

for some $t \in [a,b]$ satisfying $g(t) > 0$, where F_n is the empirical distribution function based on $(X_1,Y_1),\ldots,(X_n,Y_n)$ and $g(t)$ is defined by

$$g(t) = \int \omega(D((t,y),F(y|x=t)))dF(y|x=t). \tag{9}$$

In practice we use the empirical version

$$\frac{1}{n}\sum_{i=1}^{n} \omega(D((t,Y_i),F_n)) > 0$$

to replace the condition $g(t) > 0$. Fortunately, if the weight function and depth function are well defined, it is ensured that the condition $g(t) > 0$ holds (Zuo, Cui and He [4]).

The depth-weighted wavelet estimator $\hat{r}(t)$ is actually a weighted sum of data $Y_1,\ldots,Y_n$. Here the weights depend on both the distance $|t - X_i|$ and the depth $D((X_i,Y_i),F_n)$, i.e., larger weights are given to Y_i's whose X_i's are close to the estimation point t and (X_i,Y_i)'s have larger statistical depths. By the latter property, we can expect that the new estimation has desirable robustness.

3 Robustness

We now consider the robustness of the depth-weighted wavelet estimator $\hat{r}(t)$. One usually uses breakdown value to measure the robustness of an estimator. As shown by Donoho and Huber [10], the replacement breakdown value of an estimator T at sample $(Z)^n = (Z_1,\ldots,Z_n)$ is defined as

$$RB(T,(Z)^n) = \min\left\{\frac{k}{n} : k = 0,\ldots,n, \text{and satisfies } \sup_{(Z)_k^n} ||T((Z)_k^n) - T((Z)^n)|| = \infty\right\}, \tag{10}$$

where $(Z)_k^n$ denotes the contaminated sample by replacing k points of $(Z)^n$ with arbitrary values.

A d-dimensional sample $(Z)^n$ is said to be in a general position if there are no more than d sample points of $(Z)^n$ lying in any $(d-1)$-dimensional subspace.

Theorem 1 *Under the projection depth function* (3)*, if the weight function satisfies that* $\omega(r) \le Mr$ *for some positive constant* M *and sample* $(X,Y)^n$ *is in the general position with* $n > 5$*, then* $RB(\hat{r}(t),(X,Y)^n) = \frac{[n/2]}{n}$.

Remark (a) The condition that $\omega(r) \le Mr$ is very common. For example weight function (5) satisfies this condition. (b) This theorem shows that depth-weighted wavelet estimator $\hat{r}(t)$ attains a very high breakdown value close to 1/2. Then this estimator has the desirable robustness.

4 Asymptotic Properties

To obtain the asymptotic results, we need the regression function $r(\cdot)$, the density function $f(\cdot)$ of X and the scaling function $\varphi(\cdot)$ to satisfy the following regularity conditions:

(i) $r(\cdot), f(\cdot), r(\cdot)f(\cdot) \in H^\nu$ for $\nu > 1/2$, where H^ν is a Sobolev space; $r(\cdot)$ and $f(\cdot)$ are Lipschitz of order $\gamma > 0$; $f(\cdot)$ does not vanish on $[a,b]$.

(ii) $\varphi(\cdot)$ has compact support; $\varphi(\cdot)$ is Lipschitz; $|\hat{\varphi}(\xi) - 1| = O(|\xi|)$ as $\xi \to 0$, where $\hat{\varphi}$ denotes the Fourier transform of φ.

The regularity conditions (i) and (ii) are common in wavelet and statistical depth theories. On the other hand, for the sake of convenience, in this section we choose the weight function such as (5). In fact, the results below can been extended to the case where the regularity weight functions are as given by Zuo, Cui and He [4].

Theorem 2 *Under the conditions* (i) *and* (ii)*, if the weight function is chosen as* (5)*, $E(Y^2|X = x)$ is bounded for x belonging to a neighborhood of $t \in [a,b]$, $m \to \infty$ and $n2^{-m} \to \infty$, then $\hat{r}(t) \xrightarrow{P} r(t)$.*

To obtain the asymptotic normality result, we need to consider an approximation to $\hat{r}(\cdot)$ based on its values at dyadic points of order m. That is, define

$$\hat{r}_d(t) = \hat{r}(t^{(m)}), \tag{11}$$

where $t^{(m)} = [2^m t]/2^m$. Thus $\hat{r}_d(t)$ is the piecewise-constant approximation to $\hat{r}(t)$ at resolution 2^{-m}.

Let $\nu^* = \min(3/2, \nu, \gamma + 1/2) - \varepsilon$ and $\varepsilon = 0$ for $\nu \neq 3/2$, $\varepsilon > 0$ for $\nu = 3/2$.

Theorem 3 *Under the conditions* (i) *and* (ii)*, if the weight function is chosen as* (5)*, $E(|Y|^{2+\delta}|X = x)$ is bounded for x belonging to a neighborhood of $t \in [a,b]$ and some $\delta > 0$, $m \to \infty$, $n2^{-m} \to \infty$ and $n2^{2m\nu^*} \to 0$, then*

$$\sqrt{n2^{-m}}(\hat{r}_d(t) - r(t)) \xrightarrow{\mathscr{D}} N(0, \sigma_\omega^2(t) w_0^2 / f(t)),$$

where $\sigma_\omega^2(t) = \frac{\int \omega^2(D(t,y), F(y|x=t))(y-r(t))^2 dF(y|x=t)}{(\int \omega(D(t,y), F(y|x=t)) dF(y|x=t))^2}$, $w_0^2 = \sum_{k \in \mathbf{Z}} \varphi^2(k)$ and $f(x)$ is the density function of X.

On the other hand, as shown by Antoniadis, Gregoire and Mckeague [2], the approximation to the original wavelet estimator (2), defined by $\tilde{r}_d(t) = \tilde{r}(t^{(m)})$, satisfies

$$\sqrt{n2^{-m}}(\tilde{r}_d(t) - r(t)) \xrightarrow{\mathscr{D}} N(0, \sigma^2(t) w_0^2 / f(t)),$$

where $\sigma^2(t) = \int (y - r(t))^2 dF(y|x = t)$. According to the monotonicity relative to the deepest point (Zuo Serfling [8]), in general, $D((x,y), F)$ has the form of $D((x,y), F) = D(|y - r(x)|, F)$ and is a decreasing function of $|y - r(x)|$. Then $\omega(D(x,y), F)$ is also a decreasing function of $|y - r(x)|$. In this case, from the Gurland inequality it follows that $\sigma_\omega^2(t) \leq \sigma^2(t)\zeta_\omega^2(t)$, where

$$\zeta_\omega^2(t) = \frac{\int \omega^2(D((t,y), F(y|x=t))) dF(y|x=t)}{(\int \omega(D((t,y), F(y|x=t))) dF(y|x=t))^2} \geq 1.$$

Then the asymptotic variance of $\hat{r}_d(t)$ is bounded by $\sigma^2(t)\zeta_\omega^2(t) w_0^2 / f(t)$. To get a smaller asymptotic variance, we should choose a suitable weight function $\omega(s)$ so as to make $\zeta_\omega^2(t)$ as small as possible. If the weight function is chosen as (5), then we should choose a suitable K to make $\zeta_\omega^2(t)$ as small as possible. But this selection depends on numerical simulation.

5 Simulations

We now carry out some simulations to illustrate the efficiency and the robustness of the proposed method. Consider the following nonparametric regression model:

$$Y_i = 100(X_i - 0.5)^3 - 15(X_i - 0.5) + \varepsilon_i,$$

where X_i's and ε_i's are independent with $X_i \sim U[0,1]$ and $\varepsilon_i \sim U[-0.25, 0.25]$. A similar model with mixed dependent data was studied by Lin [11]. Similarly to Antoniadis, Gregoire and Mckeague [2], we choose the wavelet kernel based on ${}_8\varphi$ and $m = 4$. The weight function is chosen as (5) with $K = 3$. Table 1 reports the simulation results of means of squares of residuals:

$$\hat{R} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{r}(X_i))^2 \text{ and } \tilde{R} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \tilde{r}(X_i))^2.$$

According to the numerical results, we can see that $\tilde{r}(t)$ is slightly more efficient than $\hat{r}(t)$. We also consider the contaminated data $X_i \sim (1-\varepsilon)U[0,1] + \varepsilon U(0.5, 1.5)$. Tables 2–3 show that $\hat{r}(t)$ is more robust than $\tilde{r}(t)$. These results are expected in our theory and similar to the simulation results for linear models given by Wang [12].

Table 1. Simulated values for uncontaminated data

n	200	300	400	500	600
$\hat{R}$	0.3021	0.2802	0.2404	0.2401	0.2367
$\tilde{R}$	0.3002	0.2710	0.2307	0.2360	0.2209

Table 2. Simulated values for contaminated data with $\varepsilon = 0.15$

n	200	300	400	500	600
$\hat{R}$	0.6680	0.6504	0.6690	0.6491	0.6209
$\tilde{R}$	0.6807	0.6698	0.6805	0.6640	0.6420

Table 3. Simulated values for contaminated data with $\varepsilon = 0.20$

n	200	300	400	500	600
$\hat{R}$	0.7070	0.6901	0.6823	0.6756	0.6802
$\tilde{R}$	0.7524	0.7417	0.7360	0.7502	0.7385

6 Selected Proofs

Proof of Theorem 1 In order to emphasize the influence from the sample, we now denote the depth-weighted wavelet estimator $\hat{r}(t)$ by $\hat{r}(t|(X,Y)^n)$.

(i) $k = [n/2]$ points are sufficient for breakdown of $\hat{r}(t|(X,Y)^n)$. Since $(Y,X)^n$ is in general position, we can select a straight line l determined by two sample points from $(X,Y)^n$. Let (X_j, Y_j) be a point not on l. Consider replacing k other points in $(X,Y)^n$ (not from l and (X_j,Y_j)) by (X^*, Y^*), a points on l. Choose two unit vectors u_1 and u_2 perpendicular to l and to the straight line connecting (X^*,Y^*) and (X_j,Y_j) respectively. Since $k+1 > n/2$, it follows that $O(Z_i,(Z)^n) = \infty$ for all points except at (X^*,Y^*), where $(Z)^n = (Z_1,\ldots,Z_n)$ denotes the contaminated sample. This means $\hat{r}(t|(Z)^n) = \hat{r}(t|(X^*,Y^*)) = Y^*$. Therefore a breakdown occurs as $Y^* \to \infty$.

(ii) $k = [n/2]-1$ points are not sufficient for breakdown of $\hat{r}(t|(X,Y)^n)$. Since $k < [(n+1)/2]$ and $n-k > n/2$, it is clear that there exists a constant C such that $\mathrm{med}(u'(Z)^n) \le C$ and $\mathrm{mad}(u'(Z)^n) \le C$ uniformly for any unit vector u and contaminated sample $(Z)^n$. It is sufficient to show that the numerator of $\hat{r}(t|(Z)^n)$ is bounded above and the denominator is bounded away from zero. To see the former, note that for any nonzero $Z_i \in (Z)^n$ and $u_0 = Z_i/||Z_i||$, we have

$$||Z_i|| \le \mathrm{mad}(u_0'(Z)^n)O(Z_i,(Z)^n) + |\mathrm{med}(u_0'(Z)^n)| \le C/D(Z_i,(Z)^n).$$

By assumption that $\omega(r) \le Mr$ and the property that $E_m(\cdot,\cdot)$ is bounded above (Härdle, Kerkyacharian, Picard and Tsybakov [3]), we have $|Y_i\,\omega(D((X_i,Y_i),F_n))| \le MC$ uniformly for

(X_i, Y_i) and then $Y_i\,\omega(D((X_i, Y_i), F_n))E_m(t, X_i)$ is bounded above, implying that the numerator of is bounded above.

Furthermore, as shown by Zuo, Cui and He [4], the denominator of $\hat{r}(t|(Z)^n)$ is bounded away from zero.

Proof of Theorem 2 Denote $\hat{q}(t) = n^{-1}\sum_{i=1}^{n}\omega(D((X_i, Y_i), F_n))E_m(t, X_i)$. Example 3.1 of Zuo, Cui and He [4] shows that

$$\omega(D((x, y), F_n)) = \omega(D((x, y), F)) + o_P(1) \tag{12}$$

for any (x, y). It is well known that, if a function $h(\cdot)$ belongs to the Sobolev Space and satisfies Lipschits Condition of order $\gamma > 0$, then

$$\int E_m(t, s)h(s)ds \to h(t) \quad \text{as } m \to \infty. \tag{13}$$

Then, by the above results, we have

$$\begin{aligned}
E(\hat{q}(t)) &= \int \omega(D((x, y), F_n))E_m(t, x)dF(x, y) \\
&= \int \omega(D((x, y), F))E_m(t, x)dF(x, y) + o(1) \\
&= \int g(x)E_m(t, x)f(x)dx + o(1) \\
&= g(t)f(t) + o(1),
\end{aligned} \tag{14}$$

where $g(t)$ is defined by (9) and $f(x)$ is the density function of X. Furthermore, from (12) and the Lemma 6.1 of Antoniadis, Gregoire and Mckeague [2], it follows that

$$\begin{aligned}
\mathrm{Var}(\hat{q}(t)) &\le n^{-1}\int \omega^2(D((x, y), F_n))E_m^2(t, x)dF(x, y) \\
&= n^{-1}\int p(x)f(x)E_m^2(t, x)dx + o(1) = O\left(\frac{2^m}{n}\right) + o(1) \to 0,
\end{aligned} \tag{15}$$

where $p(x) = \int \omega^2(D((x, y), F(y|x)))dF(y|x)$. Then (14) and (15) show

$$\hat{q}(t) = \frac{1}{n}\sum_{i=1}^{n}\omega(D((X_i, Y_i), F_n))E_m(t, X_i) \xrightarrow{P} g(t)f(t). \tag{16}$$

By the definition (6) or (7) and the same argument that is applied to prove (16), we obtain

$$\frac{1}{n}\sum_{i=1}^{n} Y_i\,\omega(D((X_i, Y_i), F_n))E_m(t, X_i) \xrightarrow{P} r(t)\,g(t)f(t).$$

The result above and (16) imply that $\hat{r}(t) \xrightarrow{P} r(t)$.

Proof of Theorem 3 Using the continuity of f and the same argument as in the proof of Theorem 2, we can prove that $\hat{q}_d(t) \xrightarrow{P} g(t)f(t)$, where $\hat{q}_d$ is the piecewise-constant approximation to $\hat{q}$ and $\hat{q}$ is defined as in the proof of Theorem 2. By the above result and

$$\hat{r}_d(t) - r(t) = [\hat{h}_d(t) - r(t)\,\hat{q}_d(t)]/\hat{q}_d(t),$$

where $\hat{h}_d(t) = \hat{h}(t^{(m)})$ and $\hat{h}(t) = n^{-1}\sum_{i=1}^{n} Y_i\,\omega(D((X_i, Y_i), F_n))E_m(t, X_i)$, we can turn to considering $\sqrt{n2^{-m}}(\hat{h}_d(t) - r(t)\,\hat{q}_d(t))$. Denote $W_i = E_m(t^{(m)}, X_i)\omega(D((X_i, Y_i), F_n))[Y_i - r(t)]$. Then

$$\sqrt{\frac{n}{2^m}}(\hat{h}_d(t) - r(t)\,\hat{q}_d(t)) = \sqrt{\frac{n}{2^m}}\,\frac{1}{n}\sum_{i=1}^{n}(W_i - E(W_i)) + \sqrt{\frac{n}{2^m}}\,E(W_1).$$

By the same argument as in (14), we can see $E(W_1) \to 0$ and consequently

$$\sqrt{\frac{n}{2^m}}(\hat{h}_d(t) - r(t)\,\hat{q}_d(t)) = \sqrt{\frac{n}{2^m}}\,\frac{1}{n}\sum_{i=1}^{n}(W_i - E(W_i)) + o(1). \tag{17}$$

To complete the proof we apply the Lindeberg–Feller theorem to the first term on the right-hand side of (17). First, by the same argument as in (14), we have

$$\frac{1}{2^m}\mathrm{Var}(E(W_1|X_1)) = \frac{1}{2^m}\left(\int E_m^2(t^{(m)}, x)g^2(x)[r(x) - r(t)]^2 f(x)dx - (E(W_1))^2\right) + o(1),$$

which tends to zero by Lemma 6.1 of Antoniadis, Gregoire and Mckeague [2]. Next,

$$\begin{aligned}\frac{1}{2^m}E(\mathrm{Var}(W_1|X_1)) = \frac{1}{2^m}&\left(\int E_m^2(t^{(m)}, x)g^2(x)\sigma_\omega^2(x)f(x)dx\right.\\ &\left. - \int E_m^2(t^{(m)}, x)g^2(x)(r(x) - r(t))^2 f(x)dx\right) + o(1)\\ \to\ & g^2(t)\sigma_\omega^2(t)f(t)w_0^2,\end{aligned}$$

again by Lemma 6.1 of Antoniadis, Gregoire and Mckeague [2]. Thus, from the conditional variance formula

$$\mathrm{Var}(W_1) = E(\mathrm{Var}(W_1|X_1)) + \mathrm{Var}(E(W_1|X_1)),$$

we can see that the variance of the first term on the right-hand of (17) tends to $g^2(t)\sigma_\omega^2(t)f(t)w_0^2$. It remains to check the Lindeberg condition, which amounts to showing that

$$E(U^2 I(|U| > \delta\sqrt{n})) \to 0, \quad \forall\, \delta > 0,$$

where $U = (W_1 - E(W_1))/\sqrt{\mathrm{Var}(W_1)}$. By the Cauchy–Schwarz and Chebyshev inequalities,

$$E(U^2 I(|U| > \delta\sqrt{n})) \le (E(U^4))^{1/2}(n\delta^2)^{-1/2}.$$

From Lemma 6.1 and the proof of Theorem 3.1 of Antoniadis, Gregoire and Mckeague [2], we get that $\sup_{t\in[a,b]}|E_m(t, x)| = O(2^m)$, $E(W_1) = o(1)$ and $\mathrm{Var}(W_1) = O(2^m)$. Then

$$\begin{aligned}E(U^4) &= O(2^{-2m})E(W_1)^4\\ &= O(2^{-2m})\int E_m^4(t^{(m)}, x)\omega^4(D(x, y), F)(y - r(t))^4 dF(x, y) + o(1)\\ &= O(1)\int E_m^2(t^{(m)}, x)\omega^4(D(x, y), F)(y - r(t))^4 dF(x, y) + o(1)\\ &= O(2^m).\end{aligned}$$

Therefore $E(U^2 I(|U| > \delta\sqrt{n})) = O(\sqrt{2^m/n}) = o(1)$, as required. The proof is completed.

Acknowledgements The author would like to thank the two referees for their constructive suggestions.

References

[1] Hart, J. D.: Nonparametric smoothing and lack-of-fit tests, Springer-Verlag, New York, Inc., 1997
[2] Antoniadis, A., Gregoire, G., Mckeague, I. W.: Wavelet methods for curve esitmation. *JASA.*, **89**, 1340–1353 (1994)
[3] Härdle, W., Kerkyacharian, G., Picard, D., Tsbakov, A. B.: Wavelet, approximation, and statistical applications, Springer-Verlag, New York, Inc., 1998
[4] Zuo, Y. J., Cui, H. J., He, X. M.: On the Stahel–Donoho estimator and depth-weighed means of multivariate data. *Ann. Statist.*, **32**, 167–188 (2004)
[5] Tukey, J. W.: Mathematics and picturing data. *Proc. Intern. Congr. Math. Vancouver*, **2**, 523–531 (1975)
[6] Liu, R. Y.: On a notion of data depth based on random simplices. *Ann of Statist.*, **18**, 405–414 (1990)
[7] Rousseeuw, P. J., Hubert M.: Regression depth. *JASA.*, **94**, 389–433 (1999)
[8] Zuo, Y. J., Serfling, R.: General notions of statistical depth function. *Ann. Statist.*, **28**, 461–482 (2000a)

[9] Zuo, Y. J., Serfling, R.: Structural properties convergence results for contours of sample statistical depth functions. *Ann. Statist.*, **28**, 483–499 (2000b)

[10] Donoho, D. L., Huber, P. J.: The notion of breakdown point, in A Festschrift for Erich L. Lehmann (P. J. Bickel, K. A. Doksum and J. L. Hodges, Jr., eds.) 157–184, Wadsworth, Belmont, CA, 1983

[11] Lin, L.: Some statistical models with dependent data and analysis, Doctoral dissertation, Nankai University, 2001

[12] Wang, Z.: Statistical depth function and its applications, Master dissertation, Nankai University, 2003