Nonlinear wavelet estimator of the regression function under left truncated dependent data

Jacobo de Uña-Álvarez[a], Han-Ying Liang[a,b, 1], Alberto Rodríguez-Casal[c]

[a]Department of Statistics and OR, Facultad de Ciencias Económicas y Empresariales, Universidad de Vigo, Campus Lagoas-Marcosende, 36310 Vigo, Spain

[b]Department of Mathematics, Tongji University, Shanghai 200092, P. R. China

[c]Department of Statistics an OR, Facultad de Matemáticas, Universidad de Santiago de Compostela, Campus Sur, 15782 Santiago, Spain

Abstract

In this paper, we define a new nonlinear wavelet-based estimator of the regression function under random left-truncation. We provide an asymptotic expression for the mean integrated squared error (MISE) of the estimator. It is assumed that the observations form a stationary α-mixing sequence. The nonlinear wavelet-based estimator of the covariate's density is considered as well. Unlike for kernel estimators, the MISE expression of the wavelet-based estimators is not affected by the presence of discontinuities in the curves. The finite sample behaviour of the proposed estimators is explored through simulations

Key words and phrases: Mean integrated squared error; nonlinear wavelet-based estimator; nonparametric regression; truncated data; α-mixing sequence

AMS 2000 subject classifications: 62G07; 62G20.

1 Introduction

The importance of wavelets in curve estimation is well known since the initial works by Kerkyacharian and Picard (1992, 1993), Donoho and Johnstone (1994, 1995), and Donoho et al. (1995, 1996). In these papers, adaptation of wavelets (in the minimax sense) to the degree of smoothness of the underlying function is analyzed, for a wide range of functional spaces and a number of loss functions. This is a remarkable property of the wavelet method when compared to other common estimation techniques (such as the kernel method) which may fail in unsmooth situations. Hall and Patil (1995) gave for the first time an asymptotic expression of the mean integrated squared error (MISE) of a nonlinear wavelet density estimator, comparing its performance to that corresponding to the kernel density estimator. These authors showed that the asymptotic MISE formula is the same in both the smooth and unsmooth density cases,

[1]Corresponding author. *E-mail address:* hyliang83@yahoo.com (H. Y. Liang)

a fact that is not true for the kernel method. Similar results are available for the problem of estimating a regression function, see for example Hall and Patil (1996).

In some fields as Reliability and Survival Analysis, right-censored or left-truncated data are often encountered. Some authors have investigated wavelet density estimation and wavelet regression with censored data. For example, Antoniadis et al. (1999) considered linear wavelet density estimation under random censoring, providing the MISE convergence rate under smoothness assumptions on the density function; Li (2003) proposed a non-linear wavelet estimator of the density function with censored data and derived a result similar to the main result, Theorem 2.1, of Hall and Patil (1995). Also, Rodríguez-Casal and de Uña-Álvarez (2004) investigated the asymptotic expression of the MISE for the non-linear wavelet estimator of the density function under the Koziol-Green model of random censorship. Finally, Li et al. (2008) considered nonlinear wavelet regression in the censored case. However, there is not much research on wavelet estimators with left-truncated data. This is an important gap we fill with the present work.

Another existing gap in wavelet estimation from incomplete data is that all the mentioned references are devoted to independent data. However, the dependent data scenario is an important one in a number of applications with survival data. For example, when sampling clusters of individuals (family members, or repeated measurements on the same individual, for example), lifetimes within clusters are typically correlated (see Kang and Koheler, 1997, or Cai et al., 2000). In these applications, short-range dependence conditions as α-mixing have been found to be realistic (Cai and Kim, 2003), and some theory has been adapted accordingly. With complete α-mixing data, Liang et al. (2005) discussed the global L_2 error of the nonlinear wavelet estimator of the density function in the Besov space; while Truoug and Patil (2001) gave the MISE result in nonlinear wavelet regression. The case of (complete) long memory data was considered by Li and Xiao (2006, 2007), who provided the asymptotic MISE of the nonlinear wavelet-based regression estimator. Interestingly, the long memory (or long-range dependence) situation differs from short-range dependence and from the independent case in that it leads to slower convergence rates of the estimator. Once again, there is little or no literature devoted to wavelet estimation with censored and/or truncated dependent data. In this paper, we focus on nonlinear wavelets for the estimation of the covariable density and the regression function with left-truncated, dependent data.

Let Y be a response variable with continuous distribution function (df) F and let X be a continuous univariate covariable taking its values in $[0,1]$ with df V and density v. In nonparametric statistics, a smooth regression function is commonly used to describe the relationship between Y and X. The regression function at a point $x \in [0,1]$ is the conditional expectation of Y given $X = x$, and it is given by

$$\mathbb{E}[Y|X = x] := m(x) \quad x \in [0,1], \tag{1.1}$$

which can be written as $m(x) = h(x)/v(x)$, where $h(x) = \int_{\mathbb{R}} yf(x,y)dy$ with $f(\cdot,\cdot)$ being the joint density function of (X,Y). In practice, the response variable Y – a variable of interest, referred to hereafter as the lifetime, may be subject to right censoring and/or left truncation. In this paper we are interested in the left truncation model. Left-truncated data occur in astronomy, economics, epidemiology and biometry; see, e.g., Woodroofe (1985), Feigelson and Babu (1992), Wang et al. (1986), Tsai et al. (1987) and He and Yang (1994).

Under the assumption that the lifetime observations are mutually independent, regression with left-truncated data has been considered in a number of papers. Gross and Lai (1996) introduced linear regression for left-truncated and right-censored data, while Park and Hwang (2003) investigated regression depth in the same scenario. In a completely nonparametric setup, Iglesias-Pérez and González-Manteiga (1999) and Iglesias-Pérez (2003) considered respectively estimation of a conditional distribution and its quantiles. Also, Park (2004) gave the optimal convergence rate for B-splines regression under truncation and censorship. Recently, Ould-Saïd and Lemdani (2006) introduced a kernel estimator of the regression function under left-truncation, and investigated its asymptotic properties under independent and identically distributed (i.i.d.) framework. In this paper we define a new nonlinear wavelet-based estimator of the regression function under the left-truncation model, and establish an asymptotic expression of the MISE for the estimator of the regression function when the data exhibit some kind of dependence. Also, the MISE result of the nonlinear wavelet-based estimator of the covariable's density is considered.

Let $\{(X_k, Y_k, T_k),\ k \geq 1\}$ be a sequence of random vectors from (X,Y,T), where T is the truncation variable. For the components of (X,Y,T), in addition to the assumptions and notation for X and Y we made above, we assume throughout that T is independent of (X,Y), and T has continuous df G. Let $F(\cdot,\cdot)$ be the joint df of the random variable (X,Y). Without loss of generality, we assume that both Y and T are nonnegative random variables, as usual in survival analysis. In the random left-truncation model, the lifetime Y_i is interfered by the truncation random variable T_i in such a way that both Y_i and T_i are observable only when $Y_i \geq T_i$, whereas neither is observed if $Y_i < T_i$ for $i = 1, \cdots, N$, where N is the potential sample size. Due to the occurrence of truncation, the N is unknown, and n – the size of the actually observed sample, is random with $n \leq N$. Let $\theta = \mathbb{P}(Y \geq T)$ be the probability that the random variable Y is observable. Since $\theta = 0$ implies that no data can be observed, we suppose throughout the paper that $\theta > 0$.

Since N is unknown and n is known (although random), our results will not be stated with respect to the probability measure $\mathbb{P}$ (related to the N-sample) but will involve the conditional probability P with respect to the actually observed n-sample. Also $\mathbb{E}$ and E will denote the expectation operators under $\mathbb{P}$ and P, respectively.

In the sequel, the observed sample $\{(X_k, Y_k, T_k),\ 1 \leq k \leq n\}$ is assumed to be a stationary α-mixing sequence. Recall that a sequence $\{\zeta_k, k \geq 1\}$ is said to be α-mixing if the α-mixing coefficient

$$\alpha(n) := \sup_{k \geq 1} \sup\{|P(AB) - P(A)P(B)| : A \in \mathcal{F}_{n+k}^{\infty}, B \in \mathcal{F}_1^k\}$$

converges to zero as $n \to \infty$, where $\mathcal{F}_l^m$ denotes the σ-algebra generated by $\zeta_l, \zeta_{l+1}, \ldots, \zeta_m$ with $l \leq m$. Among various mixing conditions used in the literature, α-mixing is reasonably weak and has many practical applications; see, e.g., Doukhan (1994), page 99, for more details. In particular, the stationary autoregressive-moving average (ARMA) processes, which are widely applied in time series analysis, are α-mixing with exponential mixing coefficient, i.e., $\alpha(k) = O(\rho^k)$ for some $0 < \rho < 1$. As, mentioned α-mixing has been used in applications with clustered survival data see, for instance, Cai and Kim 2003.

The rest of this paper is organized as follows. In the next section, we give some notations for the left-truncation model. Basic elements of the wavelet theory, and the definition of the nonlinear wavelet-based estimators of $m(\cdot)$, $v(\cdot)$ and $h(\cdot)$ are given too. Main results are described in Section 3, their proofs are given in Section 4. Section 5 analyzes the finite sample properties through a simulation study. In Appendix, we collect some preliminary lemmas, which are used in Section 4.

2 Notations and Wavelet-based Estimators

Following Stute (1993) the conditional dfs of Y and T given no occurrence of the truncation are

$$F^*(y) = P(Y \leq y) = \mathbb{P}(Y \leq y | Y \geq T) = \theta^{-1} \int_0^y G(u)dF(u)$$

and $G^*(y) = \mathbb{P}(T \leq y | Y \geq T) = \theta^{-1} \int_0^\infty G(y \wedge u)dF(u)$, which can be estimated by

$$F_n^*(y) = n^{-1} \sum_{i=1}^n I(Y_i \leq y) \quad \text{and} \quad G_n^*(y) = n^{-1} \sum_{i=1}^n I(T_i \leq y),$$

respectively, where $I(\cdot)$ is the indicator function.

Since $C(y) = \mathbb{P}(T \le y \le Y | Y \ge T) = \theta^{-1} G(y)[1 - F(y)] = G^*(y) - F^*(y)$, the empirical estimator of $C(y)$ is defined by $C_n(y) = n^{-1} \sum_{i=1}^n I(T_i \le y \le Y_i) = G_n^*(y) - F_n^*(y-)$, where $F_n^*(y-)$ denotes the left-limit of F_n^* at y.

Following the idea of Lynden-Bell (1971), the nonparametric maximum likelihood estimators of F and G are given by

$$1 - F_n(y) = \prod_{i:Y_i \le y} \Big(1 - \frac{1}{nC_n(Y_i)}\Big) \quad \text{and} \quad G_n(y) = \prod_{i:T_i > y} \Big(1 - \frac{1}{nC_n(T_i)}\Big).$$

The estimator of θ is defined by $\theta_n = G_n(y)[1 - F_n(y-)] C_n^{-1}(y)$. He and Yang (1998) (devoted to the i.i.d. setting) proved that θ_n does not depend on y and its value can then be obtained for any y such that $C_n(y) \ne 0$.

Now we introduce some notation corresponding to wavelets. Let $\phi(x)$ and $\psi(x)$ be father and mother wavelets, having the properties: ϕ and ψ are bounded and compactly supported; $\int \phi^2 = \int \psi^2 = 1$, $\mu_k = \int y^k \psi(y) dy = 0$ for $0 \le k \le r-1$ and $\mu_r = r!\kappa$, where $\kappa = (r!)^{-1} \int y^r \psi(y) dy$. Therefore, the functions

$$\phi_j(x) = p^{1/2} \phi(px - j), \quad \psi_{ij}(x) = p_i^{1/2} \psi(p_i x - j), \quad x \in \mathbb{R}$$

for arbitary $p > 0$, $i, j \in \mathbb{Z}$, $i \ge 0$ and $p_i = p2^i$, are orthonormal:

$$\int \phi_{j_1} \phi_{j_2} = \delta_{j_1 j_2}, \quad \int \psi_{i_1 j_1} \psi_{i_2 j_2} = \delta_{i_1 i_2} \delta_{j_1 j_2}, \quad \int \phi_{j_1} \psi_{i j_2} = 0, \tag{2.1}$$

where δ_{ij} denotes the Kronecker delta [i.e. $\delta_{ij} = 1$, if $i = j$; 0, otherwise], and the system $\{\phi_j(x), \psi_{ij}(x), i, j \in \mathbb{Z}, i \ge 0\}$ is an orthonormal basis for the space $L_2(\mathbb{R})$. For more on wavelets see Daubechies (1992) or Härdle et al. (1998).

In this note, the regression function m, function h and density v are supported on the unit interval $[0,1]$. Hence, without loss of generality, we may and will assume that ϕ and ψ are compactly supported on $[0,1]$. For every function v in $L_2([0,1])$, we have the following wavelet expansion:

$$v(x) = \sum_{j=0}^{p-1} a_j \phi_j(x) + \sum_{i=0}^{\infty} \sum_{j=0}^{p_i - 1} a_{ij} \psi_{ij}(x), \tag{2.2}$$

where $a_j = \int v\phi_j$ and $a_{ij} = \int v\psi_{ij}$ are the wavelet coefficients of the function $v(\cdot)$ and the series in (2.2) converges in $L_2([0,1])$. Notice that, in order to simplify the notation, $p - 1$ and $p_i - 1$ denote $\lfloor p \rfloor - 1$ and $\lfloor p_i \rfloor - 1$, respectively, where $\lfloor z \rfloor$ is the integer part of z.

Note that in our application $a_j = \int v\phi_j = \int \phi_j dV$ and hence an estimator of a_j (resp. of a_{ij}) can be constructed on the basis of any given estimator for the covariate's cumulative distribution function $V(\cdot)$. As usually with truncated data, ordinary empiricals fail to be consistent due to the presence of biased data in the sampling; hence, some building of a specific estimator is needed to overcome this issue.

For any df H, let $a_H = \inf\{y : H(y) > 0\}$ and $b_H = \sup\{y : H(y) < 1\}$ be its two endpoints. Following the idea of Ould-Saïd and Lemdani (2006), we build an estimator of $V(\cdot)$. First, we consider the conditional joint distribution of (X, Y, T)

$$H^*(x, y, t) \quad = \quad \mathbb{P}(X \le x, Y \le y, T \le t | Y \ge T) = \frac{1}{\theta} \int_{u \le x} \int_{a_G \le w \le y} G(w \wedge t) F(du, dw).$$

Taking $t = +\infty$, we get the conditional joint df of (X, Y)

$$F^*(x, y) := \mathbb{P}(X \le x, Y \le y | Y \ge T) = \theta^{-1} \int_{u \le x} \int_{a_G \le w \le y} G(w) F(du, dw),$$

which by differentiating gives

$$F(dx, dy) = \theta G^{-1}(y) F^*(dx, dy) \quad \text{for} \quad y > a_G. \tag{2.3}$$

Integrating over y we get the df of X: $V(x) = \theta \int_{u \le x} \int_{y \ge a_G} \frac{1}{G(y)} F^*(du, dy)$. A natural estimator of V is then given by

$$V_n(x) = \frac{\theta_n}{n} \sum_{k=1}^{n} \frac{1}{G_n(Y_k)} I(X_k \le x). \tag{2.4}$$

Note that in Eq. (2.4) and the forthcoming formulae, the sum is taken only for k such that $G_n(Y_k) \ne 0$. In view of (2.4), the proposed non-linear wavelet estimator of $v(x)$ is

$$\hat{v}(x) = \sum_{j=0}^{p-1} \hat{a}_j \phi_j(x) + \sum_{i=0}^{q-1} \sum_{j=0}^{p_i - 1} \hat{a}_{ij} I(|\hat{a}_{ij}| > \delta) \psi_{ij}(x), \tag{2.5}$$

where $\delta > 0$ is a "threshold" and $q \ge 1$ is another smoothing parameter, and the wavelet coefficients $\hat{a}_j$ and $\hat{a}_{ij}$ are defined as follows:

$$\hat{a}_j = \int \phi_j dV_n = \frac{\theta_n}{n} \sum_{k=1}^{n} \frac{1}{G_n(Y_k)} \phi_j(X_k), \qquad \hat{a}_{ij} = \int \psi_{ij} dV_n = \frac{\theta_n}{n} \sum_{k=1}^{n} \frac{1}{G_n(Y_k)} \psi_{ij}(X_k). \tag{2.6}$$

Similarly, as for v, if the function h is square-integrable then its wavelet expansion is given by

$$h(x) = \sum_{j=0}^{p-1} b_j \phi_j(x) + \sum_{i=0}^{\infty} \sum_{j=0}^{p_i - 1} b_{ij} \psi_{ij}(x), \quad x \in [0, 1], \tag{2.7}$$

where $b_j = \int h\phi_i$ and $b_{ij} = \int h\psi_{ij}$. Note that $H_n(x) = \frac{\theta_n}{n}\sum_{k=1}^n \frac{Y_k}{G_n(Y_k)} I(X_k \le x)$ is an estimator of $H(x) = \int_{u\le x} h(u)du$ (see Ould-Saïd and Lemdani (2006)). So, the proposed non-linear wavelet estimator of $h(x)$ is

$$\hat{h}(x) = \sum_{j=0}^{p-1} \hat{b}_j \phi_j(x) + \sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} \hat{b}_{ij} I(|\hat{b}_{ij}| > \delta)\psi_{ij}(x), \tag{2.8}$$

where $\hat{b}_j = \frac{\theta_n}{n}\sum_{k=1}^n \frac{Y_k}{G_n(Y_k)}\phi_j(X_k)$, $\hat{b}_{ij} = \frac{\theta_n}{n}\sum_{k=1}^n \frac{Y_k}{G_n(Y_k)}\psi_{ij}(X_k)$. Further, from (2.5) and (2.8), a wavelet estimator of $m(x)$ is given by $\hat{m}(x) = \hat{h}(x)/\hat{v}(x)$.

3 Main Results

In the sequel, let $C, C_0, C_1, \cdots$ and c denote generic finite positive constants, whose values are unimportant and may change from line to line. $A_n = O(B_n)$ stands for $A_n \le CB_n$. Throughout this paper, we assume that

$$a_G < a_F \quad b_G \le b_F < \infty. \tag{3.1}$$

In order to formulate the main results, we need to impose the following assumptions.

(A1) For all integers $j \ge 1$, the joint conditional density $v_j^*(\cdot,\cdot)$ of X_1 and X_{j+1} exists and satisfies $v_j^*(x,x') \le C_0$ for all $x, x' \in [0,1]$.

(A2) The density $v(\cdot)$ satisfies $v(x) \ge C_1$ for $x \in [0,1]$.

(A3) The smoothing parameters p, q and δ are functions of n. Suppose that $p \to \infty$, $q \to \infty$ as $n \to \infty$ in such a manner that $p_q\delta^2 = O(n^{-\epsilon})$ for some $0 < \epsilon < 1$, $p^{2r+1}\delta^2 \to \infty$, $\delta \ge C_3(n^{-1}\log n)^{1/2}$.

Remark 3.1 *If $v(x)$ in (A2) is continuous on $[0,1]$, then $v(x) \le C_2$. Truoug and Patil (2001) used the assumptions $C_1 \le v(x) \le C_2$, (A3) as well as other conditions. In i.i.d. setting, the assumption (A3) except $p_q\delta^2 = O(n^{-\epsilon})$ for some $0 < \epsilon < 1$ had been used by some authors, such as Hall and Patil (1995), Li (2003) and Rodríguez-Casal and de Uña-Álvarez (2004).*

Theorem 3.1 *In addition to the conditions on ϕ and ψ stated in Section 2 and the assumptions (A1)-(A3) and (3.1), let $\alpha(k) = O(k^{-\lambda})$ for some*

$$\lambda \ge \max\{(2-\epsilon)/\epsilon, 3+4r, 1+(2r+1)/\epsilon, (\tau-1)(2\tau+1)(2-\epsilon)/(2\epsilon(\tau-2))\}, \tag{3.2}$$

where $\tau > 2$. Assume that the r-th derivatives $h^{(r)}$ and $v^{(r)}$ are continuous and bounded, and

$$\epsilon(\lambda + 1 + 2b) + 2b/(2r+1) \geq 2(b+1) \quad \textit{for} \quad b > 1. \tag{3.3}$$

Then,

(i) $E\Big|\int(\hat{v}-v)^2 - \Big\{\theta n^{-1}p\int\int \frac{f(x,y)}{G(y)}dxdy + \kappa^2(1-2^{-2r})^{-1}p^{-2r}\int {v^{(r)}}^2\Big\}\Big| = o(n^{-1}p + p^{-2r}).$

(ii) $E\Big|\int(\hat{h}-h)^2 - \Big\{\theta n^{-1}p\int\int \frac{y^2 f(x,y)}{G(y)}dxdy + \kappa^2(1-2^{-2r})^{-1}p^{-2r}\int {h^{(r)}}^2\Big\}\Big| = o(n^{-1}p + p^{-2r}).$

(iii) Let $r > 1$. Suppose that $(r+1)/(2r+1) \leq \epsilon < 2r/(2r+1)$ and $p^{2r+1} = O(n)$, then

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

Moreover, if p is chosen of size $n^{1/(2r+1)}$, then $\int(\hat{m}-m)^2 = O_p(n^{-2r/(2r+1)})$.

Remark 3.2 *(a) In the proof procedure of Theorem 3.1, in order to handle covariance part (see Step 1 below) we use assumption (A1), which is redundant for independent setting.*

(b) In Theorem 3.1, if we replace $\alpha(k) = O(k^{-\lambda})$ by the exponential decay $\alpha(k) = O(\rho^k)$ for some $0 < \rho < 1$, then (3.2) and (3.3) are automatically satisfied. While, Truoug and Patil (2001) used the assumption $\alpha(k) = O(\rho^k)$, hence, our α-mixing conditions are weaker than that in Truoug and Patil (2001).

(c) For the sake of generality, we assume that $\tau > 2$ in (3.2) and $b > 1$ in (3.3), here $\tau > 2$ and $b > 1$ are any fixed, for example, taking $\tau = 3$, $b = 2$. Actually, by appropriate choosing for τ, b, ϵ and r, inequalities (3.2) and (3.3) can be specialized.

In Theorem 3.1, we have assumed that the functions h and v are r-times continuously differentiable for simplicity and convenience of the exposition. However, if $h^{(r)}$ and $v^{(r)}$ are only piecewise continuous, Theorem 3.1 still holds, as stated in the following result.

Theorem 3.2 *In Theorem 3.1, let the derivatives $h^{(r)}$ and $v^{(r)}$ is only piecewise continuous, i.e., there exist points $x_0 = 0 < x_1 < x_2 < \cdots < x_N < 1 = x_{N+1}$ such that the first r derivatives of h and v exist and are bounded and continuous on (x_i, x_{i+1}) for $0 \leq i \leq N$, with left- and right-hand limits. In particular, h and v themself may be only piecewise continuous. Assume that $p_q^{2r+1}n^{-2r} \to \infty$. Then the conclusions (i)-(ii) in Theorem 3.1 still hold, and also (iii) in Theorem 3.1 remains true when $v^{(r)}$ is continuous and bounded.*

Remark 3.3 *(a) The error rates in our Theorems are same as that in Hall and Patil (1996) for i.i.d. complete data, and that in Truoug and Patil (2001) for α-mixing complete data.*

(b) Compared with the corresponding kernel estimator, the wavelet analogue of the bandwidth h_n of the kernel estimator is p^{-1}. As point out by Hall and Patil (1996), the variance component of the integrated squared error is of size $n^{-1}p$ (Compare $(nh_n)^{-1}$ in the case of a kernel estimator) and the squared bias component is of p^{-2r} (Compare h_n^{2r} for an rth-order kernel estimator), the optimal size of p is $cn^{1/(2r+1)}$.

(c) By choosing $p \sim n^{1/(2r+1)}$ it can be shown that the mean integrated squared errors satisfy

$$E\int(\hat{v}-v)^2 \sim \theta n^{-1}p\int\int\frac{f(x,y)}{G(y)}dxdy + p^{-2r}\kappa^2(1-2^{-2r})^{-1}\int {v^{(r)}}^2 \sim n^{-2r/(2r+1)},$$

$$E\int(\hat{h}-h)^2 \sim \theta n^{-1}p\int\int\frac{y^2 f(x,y)}{G(y)}dxdy + \kappa^2(1-2^{-2r})^{-1}p^{-2r}\int {h^{(r)}}^2 \sim n^{-2r/(2r+1)}.$$

4 Proofs of Main Results

We are now ready to prove our main results.

Proof of Theorem 3.1. We only prove (i) and (iii), the proof of (ii) is similar to that of (i) under (3.1). First, we prove (i). It follows from the orthogonality of the wavelet basis functions that

$$\begin{aligned}&\int(\hat{v}-v)^2\\ &= \sum_{j=0}^{p-1}(\hat{a}_j-a_j)^2 + \sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1}a_{ij}^2 I(|\hat{a}_{ij}|\le\delta) + \sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1}(\hat{a}_{ij}-a_{ij})^2 I(|\hat{a}_{ij}|>\delta) + \sum_{i=q}^{\infty}\sum_{j=0}^{p_i-1}a_{ij}^2\\ &:= S_1+S_2+S_3+S_4. \end{aligned} \tag{4.1}$$

It suffices to show that

$$E|S_1 - \theta n^{-1}p\int\int\frac{f(x,y)}{G(y)}dxdy| = o(n^{-1}p); \tag{4.2}$$

$$E\left|S_2 - p^{-2r}\kappa^2(1-2^{-2r})^{-1}\int {v^{(r)}}^2\right| = o(p^{-2r}); \tag{4.3}$$

$$E(S_3) = o(n^{-2r/(2r+1)}); \tag{4.4}$$

$$S_4 = O(p_q^{-2r}) = o(p^{-2r}). \tag{4.5}$$

Step 1. We verify (4.2). By using Lemma 6.5 it follows that

$$
\begin{aligned}
&E\Big|S_1 - \theta n^{-1} p \int\int \frac{f(x,y)}{G(y)} dxdy\Big| \\
&\le E\Big|\sum_{j=0}^{p-1} (\tilde{a}_j - a_j)^2 - \theta n^{-1} p \int\int \frac{f(x,y)}{G(y)} dxdy\Big| + \sum_{j=0}^{p-1} EA_j^2 + 2E\sum_{j=0}^{p-1} |\tilde{a}_j - a_j| \cdot |A_j| \\
&:= S_{11} + S_{12} + S_{13}.
\end{aligned}
\tag{4.6}
$$

Note that $E\tilde{a}_j = a_j$ and

$$
\begin{aligned}
nE(\tilde{a}_j - a_j)^2 &= \frac{\theta^2}{n} \mathrm{Var}\Big(\sum_{k=1}^{n} \frac{\phi_j(X_k)}{G(Y_k)}\Big) \\
&= \theta^2 \mathrm{Var}\Big(\frac{\phi_j(X_1)}{G(Y_1)}\Big) + 2\theta^2 \sum_{l=1}^{n-1} \Big(1 - \frac{l}{n}\Big) \mathrm{Cov}\Big(\frac{\phi_j(X_1)}{G(Y_1)}, \frac{\phi_j(X_{1+l})}{G(Y_{1+l})}\Big).
\end{aligned}
\tag{4.7}
$$

According to (2.3) we have

$$
\begin{aligned}
&\sum_{j=0}^{p-1} \theta^2 \mathrm{Var}\Big(\frac{\phi_j(X_1)}{G(Y_1)}\Big) = \sum_{j=0}^{p-1} \theta^2 \Big[\int\int \frac{\phi_j^2(x)}{G^2(y)} F^*(dx,dy) - \Big(\int\int \frac{\phi_j(x)}{G(y)} F^*(dx,dy)\Big)^2\Big] \\
&= \sum_{j=0}^{p-1} \Big[\theta \int\int \frac{\phi^2(u)}{G(y)} f\Big(\frac{u+j}{p}, y\Big) dudy - \Big(\int p^{-1/2} \phi(u) v\Big(\frac{u+j}{p}\Big) du\Big)^2\Big].
\end{aligned}
\tag{4.8}
$$

Now, by $\int \phi^2 = 1$ and the compactness of the support of ϕ, we get

$$
\sum_{j=0}^{p-1} \Big(\int p^{-1/2} \phi(u) v(\frac{u+j}{p}) du\Big)^2 \le C \sum_{j=0}^{p-1} \int \phi^2(u) p^{-1} v^2(\frac{u+j}{p}) du \to C \int v^2(x) dx.
$$

Hence, from $\sum_{j=0}^{p-1} p^{-1} f(\frac{u+j}{p}, y) \to \int f(x,y) dx$ and (4.8), it follows

$$
\sum_{j=0}^{p-1} \theta^2 \mathrm{Var}\Big(\frac{\phi_j(X_1)}{G(Y_1)}\Big) = p\theta \int\int \frac{f(x,y)}{G(y)} dxdy + o(p).
\tag{4.9}
$$

Note that, from (3.1), (2.3), (A1) and $v(x) \le C_2$ we have

$$
\begin{aligned}
&\Big|\mathrm{Cov}\Big(\frac{\phi_j(X_1)}{G(Y_1)}, \frac{\phi_j(X_{1+l})}{G(Y_{1+l})}\Big)\Big| \\
&\le G^{-2}(a_F) E|\phi_j(X_1)\phi_j(X_{1+l})| + E\Big|\frac{\phi_j(X_1)}{G(Y_1)}\Big| E\Big|\frac{\phi_j(X_{1+l})}{G(Y_{1+l})}\Big| \\
&= G^{-2}(a_F) \int\int |\phi_j(x)\phi_j(x')| v_l^*(x,x') dxdx' + \theta^2 \Big(\int |\phi_j(x)| v(x) dx\Big)^2 \\
&\le C_0 G^{-2}(a_F) p^{-1} \int\int |\phi(s)\phi(t)| dsdt + C_2^2 \theta^2 p^{-1} \Big(\int |\phi(u)| du\Big)^2 \\
&= O(p^{-1}) \quad \text{for} \quad j = 0, 1, \cdots, p-1.
\end{aligned}
$$

On the other hand, since $|\phi_j(X_k)/G(Y_k)| \le G^{-1}(a_F)p^{1/2}\|\phi\|_\infty = O(p^{1/2})$, according to Lemma 6.1 we have

$$|\mathrm{Cov}(\phi_j(X_1)/G(Y_1), \phi_j(X_{1+l})/G(Y_{1+l}))| = O(p\alpha(l)).$$

Note that $p^{2r+1}\delta^2 \to \infty$ and $p_q\delta^2 = O(n^{-\epsilon})$ imply $p \ge Cn^{\epsilon/(2r)}$. Choosing $M_n = p/\log\log n$, we have

$$2\theta^2\Big|\sum_{l=1}^{n-1}\Big(1-\frac{l}{n}\Big)\mathrm{Cov}\Big(\frac{\phi_j(X_1)}{G(Y_1)}, \frac{\phi_j(X_{1+l})}{G(Y_{1+l})}\Big)\Big| \le C\Big(\sum_{l\le M_n} + \sum_{l>M_n}\Big)\min(p^{-1}, p\alpha(l)) = o(1). \quad (4.10)$$

In the Appendix, it is proved that

$$\mathrm{Var}\Big(\sum_{j=0}^{p-1}(\tilde{a}_j - a_j)^2\Big) = o(n^{-2}p^2). \quad (4.11)$$

Then (4.7) and (4.9)-(4.11) yield that $S_{11} = o(n^{-1}p)$.

Following the line as for S_{11}, it is easy to see that

$$\begin{aligned} S_{12} &= O\Big(\frac{\ln\ln(n)}{n}\Big)\sum_{j=0}^{p-1}E\Big[\frac{\theta}{n}\sum_{k=1}^{n}\frac{|\phi_j(X_k)|}{G(Y_k)}\Big]^2 \\ &\le O\Big(\frac{\ln\ln(n)}{n}\Big)\sum_{j=0}^{p-1}\Big\{E\Big[\frac{\theta}{n}\sum_{k=1}^{n}\Big(\frac{|\phi_j(X_k)|}{G(Y_k)} - E\Big(\frac{|\phi_j(X_k)|}{G(Y_k)}\Big)\Big)\Big]^2 + \theta\Big(E\Big|\frac{\phi_j(X_1)}{G(Y_1)}\Big|\Big)^2\Big\} \\ &= O\Big(\frac{\ln\ln(n)}{n}\Big)\cdot O\Big(\frac{p}{n}\Big) + O\Big(\frac{\ln\ln(n)}{n}\Big) = o(n^{-1}p). \end{aligned}$$

As to S_{13}, we have $S_{13} \le 2\Big(\sum_{j=0}^{p-1}E(\tilde{a}_j - a_j)^2\Big)^{1/2}\Big(\sum_{j=0}^{p-1}EA_j^2\Big)^{1/2} = o(n^{-1}p)$.

Step 2. We prove (4.3). Let $\zeta > 0$, and define

$$S_{21} = \sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1}a_{ij}^2 I(|a_{ij}| \le (1+\zeta)\delta),\ \ S_{22} = \sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1}a_{ij}^2 I(|a_{ij}| \le (1-\zeta)\delta),$$

$\Delta = \sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1}a_{ij}^2 I(|\hat{a}_{ij} - a_{ij}| > \zeta\delta)$. Then $S_{22} - \Delta \le S_2 \le S_{21} + \Delta$.

By using a Taylor expansion, we have

$$\begin{aligned} a_{ij} &= p_i^{-1/2}\int\psi(u)v(\frac{u+j}{p_i})du = p_i^{-1/2}\int\psi(u)\Big[\sum_{l=0}^{r-1}\frac{1}{l!}(u/p_i)^l v^{(l)}(j/p_i) \\ &\quad + \frac{1}{(r-1)!}(u/p_i)^r\int_0^1(1-t)^{r-1}v^{(r)}((j+tu)/p_i)dt\Big]du \\ &= p_i^{-(r+1/2)}\frac{1}{(r-1)!}\int u^r\psi(u)\Big[\int_0^1(1-t)^{r-1}v^{(r)}((j+tu)/p_i)dt\Big]du \\ &= \kappa p_i^{-(r+1/2)}(g_{ij} + \eta_{ij}), \end{aligned} \quad (4.12)$$

where $g_{ij} = v^{(r)}(j/p_i)$ and $\sup_{0\le i\le q-1, 0\le j\le p_i-1} |\eta_{ij}| \to 0$.

Note that $\sup_j |a_{ij}| \le Cp_i^{-(r+1/2)} \le Cp^{-(r+1/2)}$ and $p^{r+1/2}\delta \to \infty$. Hence, for n large enough we have

$$\begin{aligned} S_{21} &= S_{22} = \sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} a_{ij}^2 = \sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} \kappa^2 p_i^{-(2r+1)}(g_{ij}+\eta_{ij})^2 \\ &= \kappa^2(1-2^{-2r})^{-1}p^{-2r}\int v^{(r)2} + o(p^{-2r}). \end{aligned}$$

Therefore, to prove (4.3), it suffices to show that $E\Delta = o(S_{21})$.

According to Lemma 6.5 we have

$$E\Delta \le \sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} a_{ij}^2 P(|\tilde{a}_{ij}-a_{ij}| > \gamma_1\zeta\delta) + \sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} a_{ij}^2 P(|A_{ij}| > \gamma_2\zeta\delta), \tag{4.13}$$

where γ_1 and γ_2 are positive constants such that $\gamma_1+\gamma_2 = 1$.

In order to evaluate $E\Delta$, we first use Lemma 6.2 to bound $P(|\tilde{a}_{ij}-a_{ij}| > \gamma_1\zeta\delta)$. Set $\xi_{ijk} = \frac{\theta}{G(Y_k)}\psi_{ij}(X_k)$. Then $E\xi_{ijk} = a_{ij}$ and $|\xi_{ijk} - E\xi_{ijk}| \le Cp_i^{1/2} := S$,

$$E(\xi_{ijk}-E\xi_{ijk})^2 \le E\xi_{ijk}^2 \le C, \ |\mathrm{Cov}(\xi_{ijs}, \xi_{ijt})| = O(p_i^{-1}) \ \text{ for } \ s \ne t.$$

Hence, by Lemma 6.3, taking $m=\infty$, for $N\in\mathbb{N}, 0<N\le n/2$ we have

$$D_N = \max_{1\le l\le 2N} \mathrm{Var}\Big(\sum_{k=1}^{l}\xi_{ijk}\Big) \le CN((p_i^{1/2})^{2/r}(p_i^{-1})^{1-1/r} + C) \le CN. \tag{4.14}$$

Note that $p_q\delta^2 = O(n^{-\epsilon})$, $\delta \ge C_3(n^{-1}\log n)^{1/2}$ and $\lambda \ge (2-\epsilon)/\epsilon$ imply $p_i^{\lambda+1}\delta^{2(\lambda-1)} < p_q^{\lambda+1}\delta^{2(\lambda-1)} \to 0$. So, according to Lemma 6.2, taking $N = [(\delta^2 p_i)^{-1/2}]$, it follows that

$$\begin{aligned} &P(|\tilde{a}_{ij}-a_{ij}| > \gamma_1\zeta\delta) = P(|\sum_{k=1}^{n}(\xi_{ijk}-E\xi_{ijk})| > n\gamma_1\zeta\delta) \\ &\le 4\exp\Big\{-\frac{n^2\gamma_1^2\zeta^2\delta^2/16}{nN^{-1}D_N + Cn\gamma_1\zeta\delta SN}\Big\} + \frac{32S}{n\gamma_1\zeta\delta}n\alpha(N) \\ &\le 4\exp\{-C_5\delta^2 n\} + C(p_i^{\lambda+1}\delta^{2(\lambda-1)})^{1/2} \to 0. \end{aligned} \tag{4.15}$$

By using arguments similar to those behind (4.14), it follows that $\mathrm{Var}(\theta\sum_{k=1}^n |\psi_{ij}(X_k)|G^{-1}(Y_k)) \le Cn$. Hence

$$\begin{aligned} EA_{ij}^2 &= O\Big(\frac{\ln\ln(n)}{n}\Big)E\Big(\frac{\theta}{n}\sum_{k=1}^{n}\frac{|\psi_{ij}(X_k)|}{G(Y_k)}\Big)^2 \\ &= O\Big(\frac{\ln\ln(n)}{n}\Big)\Big\{\mathrm{Var}\Big(\frac{\theta}{n}\sum_{k=1}^{n}\frac{|\psi_{ij}(X_k)|}{G(Y_k)}\Big) + (\theta E|\psi_{ij}(X_1)|G^{-1}(Y_1))^2\Big\} \\ &= O\Big(\frac{\ln\ln(n)}{n}\Big)\Big\{\frac{1}{n}+\frac{1}{p_i}\Big\} = o\Big(\frac{1}{n}\Big). \end{aligned} \tag{4.16}$$

From (4.13), (4.15) and (4.16), and noticing $n\delta^2 \to \infty$, it yields that

$$E\Delta \le o\Big(\sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} a_{ij}^2\Big) + \sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} a_{ij}^2 \frac{EA_{ij}^2}{\gamma_2^2\zeta^2\delta^2} = o(S_{21}).$$

Step 3. We prove (4.4). Let γ_3, γ_4 denote positive numbers satisfying $\gamma_3 + \gamma_4 = 1$. Then, from $I(|\hat{a}_{ij}| > \delta) \le I(|a_{ij}| > \gamma_3\delta) + I(|\hat{a}_{ij} - a_{ij}| > \gamma_4\delta)$, we have

$$E(S_3) \le S_{31} + S_{32}, \tag{4.17}$$

where $S_{31} = \sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} E\{(\hat{a}_{ij} - a_{ij})^2 I(|a_{ij}| > \gamma_3\delta)\}$, and $S_{32} = \sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} E\{(\hat{a}_{ij} - a_{ij})^2 I(|\hat{a}_{ij} - a_{ij}| > \gamma_4\delta)\}$. According to Lemma 6.5, it follows that

$$S_{31} \le 2\sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} E\{(\tilde{a}_{ij} - a_{ij})^2 I(|a_{ij}| > \gamma_3\delta)\} + 2\sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} E\{A_{ij}^2 I(|a_{ij}| > \gamma_3\delta)\}. \tag{4.18}$$

The proof of (4.14) shows that $E(\tilde{a}_{ij} - a_{ij})^2 \le C/n$, and (4.12) implies $\sup_j |a_{ij}| \le Cp_i^{-(r+1/2)}$. Hence, from $n^{1/2}\delta \to \infty$, we find

$$\begin{aligned}
&\sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} E\{(\tilde{a}_{ij} - a_{ij})^2 I(|a_{ij}| > \gamma_3\delta)\} = O(n^{-1})\sum_{i=0}^{q-1} p_i I(p_i \le (C/\gamma_3\delta)^{2/(2r+1)}) \\
&= O(n^{-1}\delta^{-2/(2r+1)}) = o(n^{-2r/(2r+1)}).
\end{aligned} \tag{4.19}$$

Note that $p_q\delta^2 = O(n^{-\epsilon})$ implies that $q = O(\ln(n))$, and $p_q \ln(n)/n \to 0$ by $\delta^2 \ge C^2\ln(n)/n$. Then, by using (4.16) we have

$$\begin{aligned}
&\sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} E\{A_{ij}^2 I(|a_{ij}| > \gamma_3\delta)\} = O\Big(\frac{\ln\ln(n)}{n}\Big)\sum_{i=0}^{q-1}\Big\{\frac{p_i}{n} + 1\Big\} \\
&= O\Big(\frac{\ln\ln(n)}{n}\Big)\Big\{\frac{p_q}{n} + q\Big\} = o(n^{-2r/(2r+1)}).
\end{aligned} \tag{4.20}$$

Equations (4.18)-(4.20) yield that $S_{31} = o(n^{-2r/(2r+1)})$.

Let β_1, β_2 denote positive numbers satisfying $\beta_1 + \beta_2 = 1$. On applying Lemma 6.5, we have

$$\begin{aligned}
S_{32} \;\le\;& 2\sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} E\{A_{ij}^2 I(|\hat{a}_{ij} - a_{ij}| > \gamma_4\delta)\} + 2\sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} E\{(\tilde{a}_{ij} - a_{ij})^2 I(|\hat{a}_{ij} - a_{ij}| > \gamma_4\delta)\} \\
\le\;& 2\sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} EA_{ij}^2 + 2\sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} E\{(\tilde{a}_{ij} - a_{ij})^2 I(|\tilde{a}_{ij} - a_{ij}| > \beta_1\gamma_4\delta)\} \\
&+2\sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} E\{(\tilde{a}_{ij} - a_{ij})^2 I(|A_{ij}| > \beta_2\gamma_4\delta)\}.
\end{aligned} \tag{4.21}$$

Note that

$$\begin{aligned}
& \sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} E\{(\tilde{a}_{ij}-a_{ij})^2 I(|A_{ij}|>\beta_2\gamma_4\delta)\} \\
= & \sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} E\{(\tilde{a}_{ij}-a_{ij})^2 I(|A_{ij}|>\beta_2\gamma_4\delta, |\tilde{a}_{ij}-a_{ij}|>\beta_1\gamma_4\delta)\} \\
& +\sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} E\{(\tilde{a}_{ij}-a_{ij})^2 I(|A_{ij}|>\beta_2\gamma_4\delta, |\tilde{a}_{ij}-a_{ij}|\le\beta_1\gamma_4\delta)\} \\
\le & \sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} E\{(\tilde{a}_{ij}-a_{ij})^2 I(|\tilde{a}_{ij}-a_{ij}|>\beta_1\gamma_4\delta)\} + C\sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} EA_{ij}^2,
\end{aligned}$$

which, together with (4.21) and the proof of (4.20), leads to

$$\begin{aligned}
S_{32} & \le 3\sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} E\{(\tilde{a}_{ij}-a_{ij})^2 I(|\tilde{a}_{ij}-a_{ij}|>\beta_1\gamma_4\delta)\} + C\sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} EA_{ij}^2 \\
& \le 3\sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} E\{(\tilde{a}_{ij}-a_{ij})^2 I(|\tilde{a}_{ij}-a_{ij}|>\beta_1\gamma_4\delta)\} + o(n^{-2r/(2r+1)}).
\end{aligned}$$

Therefore, it suffices to show that

$$\sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} E\{(\tilde{a}_{ij}-a_{ij})^2 I(|\tilde{a}_{ij}-a_{ij}|>\beta_1\gamma_4\delta)\} = o(n^{-2r/(2r+1)}). \tag{4.22}$$

Let a denote a positive number such that $a^{-1}+b^{-1}=1$. By using Lemma 6.7 and (4.15), according to Hölder's inequality, we have

$$\begin{aligned}
& \sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} E\{(\tilde{a}_{ij}-a_{ij})^2 I(|\tilde{a}_{ij}-a_{ij}|>\beta_1\gamma_4\delta)\} \\
& \le \sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} \Big[E|\tilde{a}_{ij}-a_{ij}|^{2a}\Big]^{1/a}\Big[P(|\tilde{a}_{ij}-a_{ij}|>\beta_1\gamma_4\delta)\Big]^{1/b} \\
& \le C\sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} \frac{1}{n}\Big\{\exp\{-C_6\delta^2 n\} + (p_i^{\lambda+1}\delta^{2(\lambda-1)})^{1/(2b)}\Big\} \\
& \le C\frac{p_q}{n}\exp\{-C_6\delta^2 n\} + Cn^{-1}p_q^{(\lambda+1)/(2b)+1}\delta^{(\lambda-1)/b} = o(n^{-2r/(2r+1)})
\end{aligned}$$

by choosing $\delta \ge C_7(n^{-1}\ln(n))^{1/2}$ with C_7 such that $C_6C_7 = 2r/(2r+1)$, and by noticing that $p_q\delta^2 = O(n^{-\epsilon})$, $\delta \ge C_3(n^{-1}\log n)^{1/2}$ and $\epsilon(\lambda+1+2b)+2b/(2r+1) \ge 2(b+1)$ imply $n^{-1}p_q^{(\lambda+1)/(2b)+1}\delta^{(\lambda-1)/b} = o(n^{-2r/(2r+1)})$.

Step 4. We verify (4.5). From (4.12), it follows that

$$S_4 = \sum_{i=q}^{\infty} \sum_{j=0}^{p_i-1} \kappa^2 p_i^{-(2r+1)} (g_{ij} + \eta_{ij})^2 \le 2\kappa^2 \sum_{i=q}^{\infty} p_i^{-(2r+1)} \sum_{j=0}^{p_i-1} g_{ij}^2 = O(p_q^{-2r}) = o(p^{-2r}).$$

Now, we prove (iii). It is easy to see that

$$\hat{m}(x) - m(x) = \frac{\hat{h}(x) - h(x)}{\hat{v}(x)} + \frac{h(x)}{v(x)} \cdot \frac{v(x) - \hat{v}(x)}{\hat{v}(x)}.$$

Then

$$\begin{aligned} \int (\hat{m}(x) - m(x))^2 dx \quad \le \quad & \frac{2}{\inf_{x\in[0,1]} v^2(x) - \sup_{x\in[0,1]} |\hat{v}^2(x) - v^2(x)|} \\ & \times \Big\{ \int (\hat{h}(x) - h(x))^2 dx + \sup_{x\in[0,1]} \Big(\frac{h(x)}{v(x)}\Big)^2 \int (\hat{v}(x) - v(x))^2 dx \Big\}. \end{aligned}$$

From (3.1) and (A2), it suffices to show that

$$\sup_{x\in[0,1]} |\hat{v}(x) - v(x)| = o_p(1), \tag{4.23}$$

$$\int (\hat{h}(x) - h(x))^2 dx = O_p(n^{-1}p + p^{-2r}), \tag{4.24}$$

$$\int (\hat{v}(x) - v(x))^2 dx = O_p(n^{-1}p + p^{-2r}). \tag{4.25}$$

Note that, (i) and (ii) imply, respectively, that

$$E \int (\hat{h}(x) - h(x))^2 dx \le C(n^{-1}p + p^{-2r}), \; E \int (\hat{v}(x) - v(x))^2 dx \le C(n^{-1}p + p^{-2r}).$$

Therefore, by using the fact that $|\eta| = O_p(E|\eta|)$ for any random variable η, it yields (4.24) and (4.25).

Next, we prove (4.23). Since ϕ and ψ are compactly supported,

$$\sup_{x\in[0,1]} |\hat{v}(x) - v(x)| \le I_1 + I_2 + I_3 + I_4, \tag{4.26}$$

where

$$I_1 = \sum_{j=0}^{p-1} |\hat{a}_j - a_j| \|\phi_j\|_\infty, \quad I_2 = \sum_{i=0}^{q-1} \sum_{j=0}^{p_i-1} |a_{ij}| I(|\hat{a}_{ij}| \le \delta) \|\psi_{ij}\|_\infty,$$

$$I_3 = \sum_{i=0}^{q-1} \sum_{j=0}^{p_i-1} |\hat{a}_{ij} - a_{ij}| I(|\hat{a}_{ij}| > \delta) \|\psi_{ij}\|_\infty, \quad I_4 = \sum_{i=q}^{\infty} \sum_{j=0}^{p_i-1} |a_{ij}| \|\psi_{ij}\|_\infty.$$

Note that $\|\phi_j\|_\infty = O(p^{1/2})$ and $\|\psi_{ij}\|_\infty = O(p_i^{1/2})$. By Hölder's inequality and (4.2), from $p^{2r+1} = O(n)$ and $r > 1$ it follows that

$$I_1 = O(p)\Big(\sum_{j=0}^{p-1}(\hat{a}_j - a_j)^2\Big)^{1/2} = p\{O_p(n^{-1}p)\}^{1/2} = o_p(1). \tag{4.27}$$

Similarly, from $p_q\delta^2 = O(n^{-\epsilon})$ with $(r+1)/(2r+1) \leq \epsilon < 2r/(2r+1)$ and $p^{2r+1}\delta^2 \to \infty$ we have

$$\begin{aligned} I_2 &\leq C\sum_{i=0}^{q-1} p_i^{1/2}\sum_{j=0}^{p_i-1}|a_{ij}|I(|\hat{a}_{ij}| \leq \delta) \leq C\Big\{\sum_{i=0}^{q-1} p_i \cdot \sum_{i=0}^{q-1}\Big(\sum_{j=0}^{p_i-1}|a_{ij}|I(|\hat{a}_{ij}| \leq \delta)\Big)^2\Big\}^{1/2} \\ &\leq C\Big\{p_q \cdot \sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1} a_{ij}^2 I(|\hat{a}_{ij}| \leq \delta)\sum_{j=0}^{p_i-1} 1\Big\}^{1/2} = p_q\{O_p(p^{-2r})\}^{1/2} = o_p(1). \qquad (4.28) \\ I_3 &\leq Cp_q\Big\{\sum_{i=0}^{q-1}\sum_{j=0}^{p_i-1}(\hat{a}_{ij} - a_{ij})^2 I(|\hat{a}_{ij}| > \delta)\Big\}^{1/2} = p_q\{o_p(n^{-2r/(2r+1)})\}^{1/2} = o_p(1). \qquad (4.29) \end{aligned}$$

From (4.12) we have

$$I_4 \leq 2\kappa\sum_{i=q}^{\infty}\sum_{j=0}^{p_i-1} p_i^{-r}|g_{ij}| = 2\kappa\sum_{i=q}^{\infty} p_i^{-(r-1)}\sum_{j=0}^{p_i-1} p_i^{-1}|g_{ij}| = O(p_q^{-(r-1)}) = o(1). \tag{4.30}$$

Then, (4.23) follows from (4.26)-(4.30). ■

Proof of Theorem 3.2. Similarly to the proof as in Theorem 3.1, we prove only (i), the proof of (ii) is analogous, and (iii) can be proved by using (i) and (ii). By the orthogonality properties of ϕ and ψ, $\int(\hat{v} - v)^2 = \Gamma_q(\mathbb{Z}, \mathbb{Z}, \cdots)$, where $\mathbb{Z}$ denotes the set of all integers (for instance, $\mathcal{L}_i = \{0, 1, \cdots, p_i - 1\}$) and

$$\begin{aligned} \Gamma_q(\mathcal{L}, \mathcal{L}_0, \mathcal{L}_1, \cdots) &= \sum_{j\in\mathcal{L}}(\hat{a}_j - a_j)^2 + \sum_{i=0}^{q-1}\sum_{j\in\mathcal{L}_i} a_{ij}^2 I(|\hat{a}_{ij}| \leq \delta) \\ &\quad + \sum_{i=0}^{q-1}\sum_{j\in\mathcal{L}_i}(\hat{a}_{ij} - a_{ij})^2 I(|\hat{a}_{ij}| > \delta) + \sum_{i=q}^{\infty}\sum_{j\in\mathcal{L}_i} a_{ij}^2 \\ &:= \Gamma_1(\mathcal{L}) + \Gamma_2(\mathcal{L}_0, \mathcal{L}_1, \cdots) + \Gamma_3(\mathcal{L}_0, \mathcal{L}_1, \cdots) + \Gamma_4(\mathcal{L}_0, \mathcal{L}_1, \cdots). \end{aligned}$$

When v is only piecewise continuous, let $\mathcal{X}$ denote the finite set of points where $v^{(s)}$ has discontinuities for some $0 \leq s \leq r$. Suppose supp$\phi \subseteq (-v, v)$, supp$\psi \subseteq (-v, v)$ and let

$\mathbb{k} = \{k, k \in (py - v, py + v)$ for some $y \in \mathcal{X}\}$, $\mathbb{k}_i = \{k, k \in (p_i y - v, p_i y + v)$ for some $y \in \mathcal{X}\}$.

Also let $\Bbbk^c$, $\Bbbk_i^c$ denote their complements. Then, unless $j \in \Bbbk_i$, a_{ij} and $\hat{a}_{ij}$ are constructed entirely from an integral over or an average of data values from an interval where $v^{(r)}$ exists and is bounded and continuous. Also, unless $j \in \Bbbk$, a_j and $\hat{a}_j$ are constructed solely from such regions. Then we may write

$$\begin{aligned}\Gamma_q(\mathcal{L}, \mathcal{L}_0, \mathcal{L}_1, \cdots) &= \Gamma_1(\Bbbk) + \Gamma_2(\Bbbk_0, \Bbbk_1, \cdots) + \Gamma_3(\Bbbk_0, \Bbbk_1, \cdots) + \Gamma_4(\Bbbk_0, \Bbbk_1, \cdots) \\ &\quad + \Gamma_1(\Bbbk^c) + \Gamma_2(\Bbbk_0^c, \Bbbk_1^c, \cdots) + \Gamma_3(\Bbbk_0^c, \Bbbk_1^c, \cdots) + \Gamma_4(\Bbbk_0^c, \Bbbk_1^c, \cdots).\end{aligned}$$

The proof of (4.7) shows $E(\tilde{a}_j - a_j)^2 = O(n^{-1})$, the evaluation for S_{12} shows $EA_j^2 = O(\frac{\ln\ln(n)}{n})(n^{-1} + p^{-1})$. Since ϕ and ψ have compactly supported, both $\Bbbk$ and $\Bbbk_i$ have no more than $(2v+1)(\#\mathcal{X})$ elements for each i. Then by Lemma 6.5 it follows that

$$E\Gamma_1(\Bbbk) \le 2\sum_{j\in\Bbbk} E(\tilde{a}_j - a_j)^2 + 2\sum_{j\in\Bbbk} EA_j^2 = O\Big(\frac{1}{n}\Big) + O\Big(\frac{\ln\ln(n)}{n}\Big)\Big[\frac{1}{n} + \frac{1}{p}\Big] = o\big(n^{-2r/(2r+1)}\big).$$

Note that

$$\begin{aligned}E\Gamma_2(\Bbbk_0, \Bbbk_1, \cdots) &\le \sum_{i=0}^{q-1}\sum_{j\in\Bbbk_i} a_{ij}^2 I(|a_{ij}| \le (1+\zeta)\delta) + \sum_{i=0}^{q-1}\sum_{j\in\Bbbk_i} a_{ij}^2 P(|\hat{a}_{ij} - a_{ij}| > \zeta\delta) \\ &= O\big(q\delta^2\big) + \sum_{i=0}^{q-1}\sum_{j\in\Bbbk_i} p_i^{-1}\big\{P(|\tilde{a}_{ij} - a_{ij}| > c\delta) + P(|A_{ij}| > c\delta)\big\}. \quad (4.31)\end{aligned}$$

From (2.3) and $\delta \ge C_3(n^{-1}\log n)^{1/2}$ we have $E|A_{ij}|\delta^{-1} \le C\sqrt{\ln\ln(n)/(n\delta^2)}\cdot\theta E(|\psi_{ij}(X_1)|/G(Y_1)) = C((\ln\ln(n)/(n\delta^2 p_i))^{1/2}) \le C((\ln\ln(n)/(p\ln(n)))^{1/2}) \to 0$, hence similarly to the proof as for (4.15) one can verify that

$$P(|A_{ij}| > c\delta) \le P(|A_{ij} - EA_{ij}| > c\delta) \le 4\exp\big\{-C_5\delta^2 n\big\} + Cp_i^{(\lambda+1)/2}\delta^{\lambda-1}.$$

Therefore, in view of $p_q^{2r+1}n^{-2r} \to \infty$ and (A3), from (4.31) and $n^{2r/(2r+1)} \cdot p_q^{(\lambda-1)/2}\delta^{\lambda-1} \le Cn^{-[\epsilon(\lambda-1)/2 - 2r/(2r+1)]} \to 0$ since $\epsilon(\lambda-1)/2 - 2r/(2r+1) > o$ by $\lambda \ge 1 + (2r+1)/\epsilon$ we have

$$\begin{aligned}E\Gamma_2(\Bbbk_0, \Bbbk_1, \cdots) &\le O\Big(q(p_q\delta^2)\cdot p_q^{-1}\Big) + C\sum_{i=0}^{q-1} p_i^{-1}\Big(\exp\big\{-C_5\delta^2 n\big\} + p_i^{(\lambda+1)/2}\delta^{\lambda-1}\Big) \\ &\le o\big(n^{-2r/(2r+1)}\big) + Cp^{-1}\exp\big\{-C_5\delta^2 n\big\} + Cp_q^{(\lambda-1)/2}\delta^{\lambda-1} = o\big(n^{-2r/(2r+1)}\big).\end{aligned}$$

By Lemmas 6.5 and 6.7, from (4.16) it follows that

$$E\Gamma_3(\Bbbk_0, \Bbbk_1, \cdots) \le 2\sum_{i=0}^{q-1}\sum_{j\in\Bbbk_i}\big\{E(\tilde{a}_{ij} - a_{ij})^2 + EA_{ij}^2\big\} = O(q/n) = o\big(n^{-2r/(2r+1)}\big).$$

Thus $\Gamma_1(\Bbbk)+\Gamma_2(\Bbbk_0, \Bbbk_1, \cdots)+\Gamma_3(\Bbbk_0, \Bbbk_1, \cdots)$ is negligible compared to the main terms of MISE. In view of $a_{ij} = O(p_i^{-1/2})$ and $p_q^{2r+1} n^{-2r} \to \infty$ we have $\Gamma_4(\Bbbk_0, \Bbbk_1, \cdots) = O(p_q^{-1}) = o(n^{-2r/(2r+1)})$. By tracing the whole proof of Theorem 3.1 carefully, $\Gamma_q(\Bbbk^c, \Bbbk_0^c, \Bbbk_1^c, \cdots)$ has precisely the asymptotic properties claimed for $\int (\hat{v} - v)^2$ in Theorem 3.1. ■

5 Simulation Study

In order to analyze the finite sample properties of the proposed estimates we have conducted a small simulation study. We have focused in the estimator for the regression function, which is probably the most interesting case in applications. Following Cai and Kim (2003), it was assumed that the observed covariables were clustered in m groups of K correlated observations ($m = 25$, 50 and $K = 3$, 5). The joint survival function for L correlated values of the covariable, $(X_1, \ldots, X_L)$, is given by

$$P(X_1 > x_1, \ldots, X_L > x_L) = \left\{ \sum_{i=1}^{L} S_i(x_i)^{-1/\lambda} - (L-1) \right\}^{-\lambda}$$

where S_i is the marginal survival function for X_i and $\lambda > 0$ is a parameter which controls the degree of dependence: large values of λ correspond with a weak dependence within each cluster whereas small ones indicate a strong dependence structure. Here we have chosen $\lambda = 0.8$ and $\lambda = 3$. As in Cai and Kim (2003), the marginal distribution of X_i was the exponential with mean one. Since we have assumed that m clusters of K dependent observation were observed, the sample size of the observed sample was $n = K \times m$. Since the clusters are statistically independent, the α-mixing property is automatically satisfied if K remains bounded as m goes to infinity

Given an observed value of a covariable, $X_i = x_i$ $(i = 1, \ldots, n)$, the variable Y_i was generated according to the model

$$Y_i = m(x_i) + \varepsilon_i,$$

where $m(x) = x$ is the true regression function and $\{\varepsilon_i\}_{i=1}^n$ is a sequence of iid random variables with zero-mean normal distribution. The values for the standard deviation of the random errors ϵ_i were $\sigma = 0.2$ and $\sigma = 1$. The truncation time, T, was independently generated according to a normal distribution with mean μ and standard deviation one. Note that, since the iid truncation times are independent of everything else, the α-mixing property of the observed sample follows from that of the original sample.The parameter μ was chosen in order to ensure that $1-\theta = P(Y < T)$ were, approximately, 10%, 30% and 60%. Table 1 shows the values of μ for the different values of σ and truncation percentages.

	10%	30%	60%
$\sigma = 1$	-1.1	0.1	1.3
$\sigma = 0.2$	-0.7	0.2	1.2

Table 1: *Values of μ to get truncation rates of* 10%, 30% *and* 60%

We have simulated $B = 1000$ random samples from the above mentioned models (48 in total). For each sample we have computed the proposed estimator for the regression function, $\hat{m}$, for several values of the parameters. The parameter q was fixed at zero (so we used a linear wavelet, for which the δ parameter plays no role). Then, in order to investigate the influence of the p parameter on the performance of the estimator, we have chosen 80 equispaced values of p between 0.05 and 4. Therefore, given a random sample, we have computed 80 estimates of m, one for each value of p. The error criteria was the integrated square error (ISE),

$$\int (\hat{m}(x) - m(x))^2 dx,$$

which was only computed between $x = 0$ and $x = 2$. This ensures that density of X, v, remains large enough (see Condition A2). Table 2 reports, for each model, the median ISE along the simulations for the wavelet based on the parameter p which minimizes this error criterion. Of course, this optimal p is not available in practice, and a very interesting topic which is left for future research is that of the development of some data-driven selection rule for the estimate's smoothing degree. The influence of p on the error is clearly seen from Figures 1 to 3, in which the median ISE is averaged along a number of simulated models. In Figure 1, performance in models with m=25, K=3 (n=75) is compared to that in models with m=50, $K = 3$ (n=150). Comparison for the case K=5 (not shown) reports a similar result. From this Figure 1 we see that the error decreases with an increasing sample size n. Figure 2 compares performance in models under 10% of truncation to 60% of truncation; this figure suggests that the regression wavelet estimator behaves worse in the second scenario. Finally, in Figure 3, the ISEs with small and large variance of the error term ε are compared. As expected, the latter scenario gives larger ISEs. Note the bath-tube shape of the curves in Figure 1 to 3, indicating the risk of undersmoothing (small p) and oversmoothing (large p). As usual, the optimal value of p

is a compromise between bias and variance. As a final remark, from Table 2 we see that the estimate is relatively robust against dependence in the data. For instance, the lowest mean error is achieved for the same value of p (0.6) both $\lambda = 0.8$ and $\lambda = 3$ and the minimum is also almost the same (0.0962 and 0.0973, respectively). This robustness is in accordance to the provided theory.

m	K	λ	σ	Tr.	Err.
25	3	3.0	1.0	10	0.095
25	3	3.0	1.0	30	0.144
25	3	3.0	1.0	60	0.351
25	3	3.0	0.2	10	0.019
25	3	3.0	0.2	30	0.020
25	3	3.0	0.2	60	0.043
25	3	0.8	1.0	10	0.096
25	3	0.8	1.0	30	0.142
25	3	0.8	1.0	60	0.378
25	3	0.8	0.2	10	0.019
25	3	0.8	0.2	30	0.018
25	3	0.8	0.2	60	0.035
25	5	3.0	1.0	10	0.268
25	5	3.0	1.0	30	0.109
25	5	3.0	1.0	60	0.323
25	5	3.0	0.2	10	0.011
25	5	3.0	0.2	30	0.012
25	5	3.0	0.2	60	0.027
25	5	0.8	1.0	10	0.077
25	5	0.8	1.0	30	0.108
25	5	0.8	1.0	60	0.308
25	5	0.8	0.2	10	0.010
25	5	0.8	0.2	30	0.010
25	5	0.8	0.2	60	0.025

m	K	λ	σ	Tr.	Err.
50	3	3.0	1.0	10	0.063
50	3	3.0	1.0	30	0.092
50	3	3.0	1.0	60	0.018
50	3	3.0	0.2	10	0.009
50	3	3.0	0.2	30	0.010
50	3	3.0	0.2	60	0.061
50	3	0.8	1.0	10	0.009
50	3	0.8	1.0	30	0.089
50	3	0.8	1.0	60	0.252
50	3	0.8	0.2	10	0.009
50	3	0.8	0.2	30	0.009
50	3	0.8	0.2	60	0.017
50	5	3.0	1.0	10	0.046
50	5	3.0	1.0	30	0.068
50	5	3.0	1.0	60	0.204
50	5	3.0	0.2	10	0.006
50	5	3.0	0.2	30	0.006
50	5	3.0	0.2	60	0.013
50	5	0.8	1.0	10	0.046
50	5	0.8	1.0	30	0.067
50	5	0.8	1.0	60	0.009
50	5	0.8	0.2	10	0.005
50	5	0.8	0.2	30	0.005
50	5	0.8	0.2	60	0.012

Table 2: *Error for the best smoothing parameter*

Acknowledgments. The authors are grateful to the Editor, an associate Editor, and two

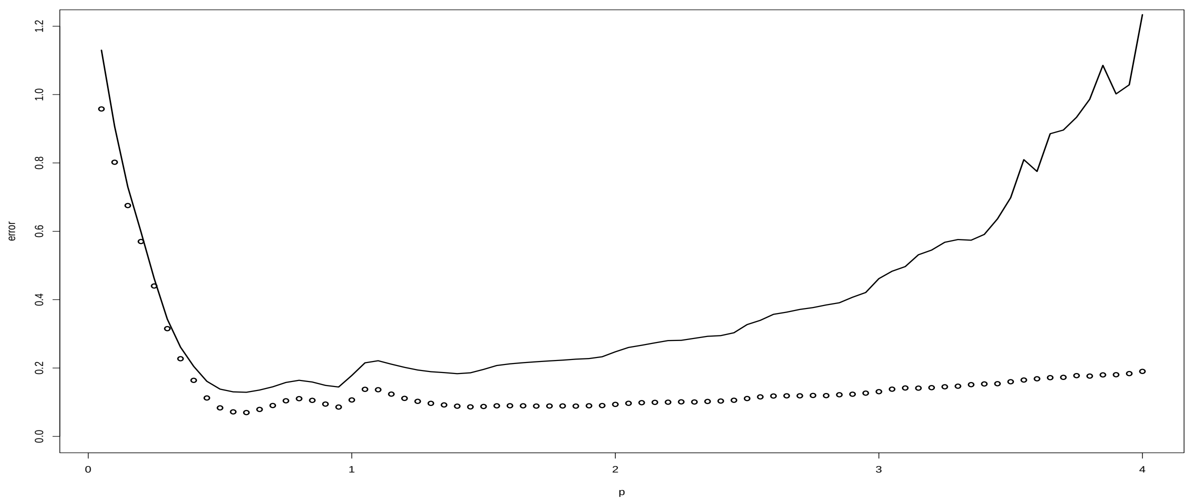

Figure 1: *Mean error for sample sizes* $n = 75$ *(continuous line) and* $n = 150$ *(dotted line)*

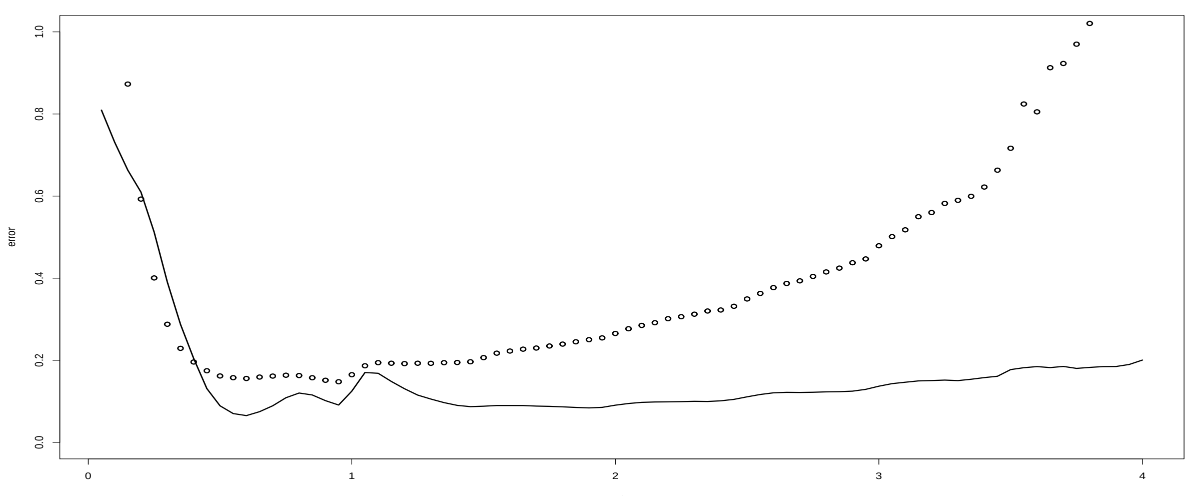

Figure 2: *Mean error for sample truncation percentages* 10% *(continuous line) and* 60% *(dotted line)*

anonymous referees for providing a detailed list of comments which greatly improved the presentation of the paper. First and second authors were supported by the Grants MTM2008-03129 of the Spanish Ministry of Science and Innovation, and PGIDIT07PXIB300191PR of the Xunta de Galicia, Spain, and also by the National Natural Science Foundation of China (10571136,10871146).Third author acknowledges financial support from the Grants MTM2008-

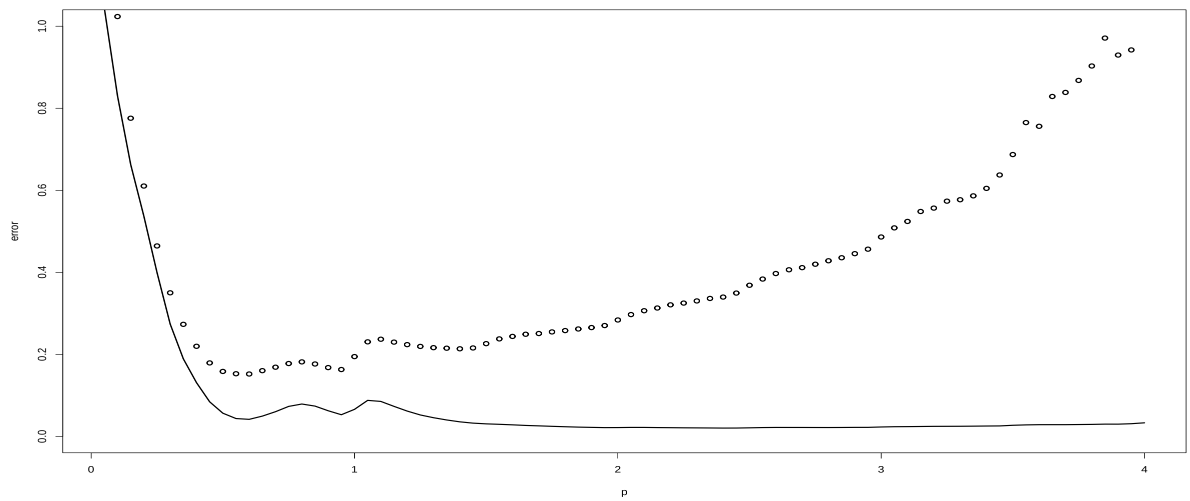

Figure 3: *Mean error for standard errors $\sigma = 0.2$ (continuous line) and $\sigma = 1$ (dotted line)*

03010 (Spanish Ministry of Science), PGIDIT06PXIB207009PR (Xunta de Galicia, Spain) and the IAP research network grant no. P6/03 of the Belgian government.

References

[1] Antoniadis, A., Grégoire, G. and Nason, G. (1999). Density and hazard rate estimation for right-censored data by using wavelet methods. *J. R. Statist. Soc.* **B 61**, 63-84.

[2] Bradley, R. C. (1983). Approximation theorems for strongly mixing random variables. *Michigan. Math. J.* **30**, 69-81.

[3] Cai, J. and Kim, J. (2003). Nonparametric quantile estimation with correlated failure time data. *Lifetime Data Analysis* **9**, 357-371.

[4] Cai, T., Wei, L.J. and Wilcox, M. (2000). Semiparametric regression analysis for clustered failure time data. *Biometrika* **87**, 867-878.

[5] Cohen, A., Daubechies, I. and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.* **1**, 54-82.

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

[7] Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81**, 425-455.

[8] Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *JASA* **90**, 1200-1224.

[9] Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage-asymptopia ? (with discussion). *J. R. Statist. Soc.* **B 57**, 301-369.

[10] Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24**, 508-539.

[11] Doukhan. P. (1994). *Mixing: Properties and Examples.* Lecture Notes in Statistics, vol. 85. Springer, Berlin.

[12] Feigelson, E. D. and Babu, G. J. (1992). *Statistical challenges in modern astronomy.* Berlin Heidelberg, Springer, New York.

[13] Gross, S. T., Lai, T.L. (1996). Nonparametric estimation and regression analysis with left-truncated and right-censored data. *J. Amer. Statist. Association* **91**, 1166-1180.

[14] Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and its Application.* Academic Press, New York.

[15] Hall, P. and Patil, P. (1995). Formulae for mean integrated squared error of non-linear wavelet-based density estimators. *Ann. Statist.* **23**, 905-928.

[16] Hall, P. and Patil, P. (1996). On the choice of smoothing parameter, threshold and truncation in nonparametric regression by non linear wavelet methods. *J. Roy. Statist. Soc. Ser. B* **58(2)**, 361-377.

[17] Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, Approximation and Statistical Applications.* Lecture Notes in Statistics, **129**, Springer, New York.

[18] He, S. and Yang, G. (1994). Estimating a lifetime distribution under different sampling plan. In S.S. Gupta J. O. Berger (Eds.) *Statistical decision theory and related topics* (Vol. 5, 73-85) Berlin Heidelberg, Springer, New York.

[19] He, S., Yang, G. (1998). Estimation of the truncation probability in the random truncation model. *Ann. Statist.* **26**, 1011-1027.

[20] Iglesias-Pérez, M. C., González-Manteiga, W. (1999). Strong representation of a generalized product-limit estimator for truncated and censored data with some applications. *J. Nonparametr. Statist.* **10**, 213-244.

[21] Iglesias-Pérez, M. C. (2003). Strong representation of a conditional quantile function estimator with truncated and censored data. *Statist. Probab. Lett.* **65**, 79-91.

[22] Kang, S. S. and Koehler, K. J. (1997). Modifcation of the greenwood formula for correlated failure times. *Biometrics* **53**, 885-899.

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

[24] Kerkyacharian, G. and Picard, D. (1993). Density estimation by kernel and wavelets methods: optimality of Besov spaces. *Statist. Probab. Lett.* **18(4)**, 327-336.

[25] Li, L. (2003). Non-linear wavelet-based density estimators under random censorship. *J. Statist. Plann. Inference* **117** 35-58.

[26] Li, L., MacGibbon, B. and Valenta, C. (2008). On the optimality of wavelet-based nonparametric regression with censored data. *J. Appl. Probab. Statist.* **3**, 243-261.

[27] Li, L. and Xiao, Y. (2006). Wavelet-based estimators of mean regression function with long memory data. *Appl. Math. Mech. (English Ed.)* **27(7)**, 901-910.

[28] Li, L. and Xiao, Y. (2007). Mean integrated squared error of non-linear wavelet-based estimators with long memory data. *Ann. Inst. Statist. Math.* **59**, 299-324.

[29] Liang, H. Y., Mammitzsch, V. and Steinebach, J. (2005). Nonlinear wavelet density and hazard rate estimation for censored data under dependent observations. *Statist. Decisions.* **23**, 161-180.

[30] Liang, H. Y., Li, D. L. and Qi, Y. C. (2009). Strong convergence in nonparametric regression with truncated dependent data. *J. Multivariate Anal.* **100**, 162-174.

[31] Liebscher, E. (1996). Strong convergence of sums of α-mixing random variables with applications to density estimation. *Stochastic Processes Appl.*, **65(1)**, 69-80.

[32] Liebscher, E. (2001). Estimation of the density and the regression function under mixing conditions. *Statist. Decisions*, **19(1)**, 9-26.

[33] Ould-Saïd, E. and Lemdani, M. (2006) Asymptotic properties of a nonparametric regression function estimator with randomly truncated data. *Ann. Inst. Statist. Math.* **58**, 357-378.

[34] Park, J. (2004). Optimal global rate of convergence in nonparametric regression with left-truncated and right-censored data. *J. Multivariate Anal.* **89**, 70-86.

[35] Park, J., Hwang, J. (2003). Regression depth with censored and truncated data. *Comm. Statist. Theory Methods.* **32**, 997-1008.

[36] Petrov, V. V. (1995). *Limit Theorems of Probability Theory.* Oxford Univ. Press Inc., New York.

[37] Rodríguez-Casal, A. and de Uña-Álvarez, J. (2004). Nonlinear wavelet density estimation under the Koziol-Green model. The International Conference on Recent Trends and Directions in Nonparametric Statistics. *J. Nonparametr. Statist.* **16(1-2)**, 91-109.

[38] Stute, W. (1993). Almost sure representations of the product-limit estimator for truncated data. *Ann. Statist.* **21**, 146-156.

[39] Truoug, Y. and Patil, P. (2001). Asymptotics for wavelet based estimates of piecewise smooth regression for stationary time series. *Ann. Inst. Statist. Math.* **53**, 159-178.

[40] Tsai, W. Y., Jewell, N. P. and Wang, M. C. (1987). A note on the product-limit estimator under right censoring and left truncation. *Biometrika* **74**, 883-886.

[41] Wang, M. C., Jewell, N. P. and Tsai, W. Y. (1986). Asymptotic properties of the product-limit estimate under random truncation. *Ann. Statist.* **14**, 1597-1605.

[42] Woodroofe, M. (1985). Estimating a distribution function with truncated data. *Ann. Statist.* **13**, 163-177.

6 Appendix

In this section, we give some preliminary Lemmas, which have been used in Section 4. Let $\{Z_i, i \geq 1\}$ be a sequence of α-mixing real random variables with the mixing coefficients $\{\alpha(k)\}$.

Lemma 6.1 (Hall and Heyde (1980), Corllary A.1) *Suppose that X and Y are random variables such that $|X| < C_1$, $|Y| < C_2$. Then*

$$|EXY - EXEY| \leq 4C_1C_2\Big\{\sup_{A\in\sigma(X), B\in\sigma(Y)} |P(A\cap B) - P(A)P(B)|\Big\}.$$

Lemma 6.2 (Liebscher (2001), Proposition 5.1) *Assume that* $EZ_i = 0$ *and* $|Z_i| \le S < \infty$ *a.s.* $(i = 1, 2, \cdots, n)$. *Set* $D_N = \max_{1\le j\le 2N} \mathrm{Var}(\sum_{i=1}^{j} Z_i)$. *Then, for* $n, N \in \mathbb{N}, 0 < N \le n/2$, $\epsilon > 0$,

$$P\Big(\Big|\sum_{i=1}^{n} Z_i\Big| > \epsilon\Big) \le 4\exp\Big\{-\frac{\epsilon^2}{16}\Big(nN^{-1}D_N + \frac{1}{3}\epsilon SN\Big)^{-1}\Big\} + 32\frac{S}{\epsilon}n\alpha(N).$$

Lemma 6.3 (Liebscher (1996), Lemma 2.3) *Assume* $\alpha(k) \le C_1 k^{-r}$, *for some* $r > 1, C_1 > 0$. *Let* $\sup_{1\le i,j\le n, i\ne j} |\mathrm{Cov}(Z_i, Z_j)| := R^*(n) < \infty$ *be satisfied. Moreover, let* $R_m(n) < \infty$ *for some* m, $2r/(r-1) < m \le \infty$, *where* $R_m(n) = \sup_{1\le i\le n}(E|Z_i|^m)^{1/m}$, *for* $1 \le m < \infty$, *and* $R_\infty(n) = \sup_{1\le i\le n} \operatorname{ess\,sup}_{w\in\Omega}|Z_i|$. *Then*

$$Var\Big(\sum_{i=1}^{n} Z_i\Big) \le n\Big\{C_2(r,m)(R_m(n))^{2m/(r(m-2))}(R^*(n))^{1-m/(r(m-2))} + R_2^2(n)\Big\}$$

holds with $C_2(r,m) := \frac{20r-40r/m}{r-1-2r/m}C_1^{1/r}$.

Lemma 6.4 (Liang et al. (2009)) *Suppose that* $\alpha(k) = O(k^{-r})$ *for some* $r > 3$. *Then*

$$\sup_y |G_n(y) - G(y)| = O((\ln\ln(n)/n)^{1/2}) \quad a.s., \quad |\theta_n - \theta| = O((\ln\ln(n)/n)^{1/2}) \quad a.s.$$

Lemma 6.5 *Let* $\hat{a}_j$, $\hat{a}_{ij}, \hat{b}_j$, $\hat{b}_{ij}$ *be as defined in Section 2. Set*

$$\tilde{a}_j = \frac{\theta}{n}\sum_{k=1}^{n}\frac{1}{G(Y_k)}\phi_j(X_k), \quad \tilde{a}_{ij} = \frac{\theta}{n}\sum_{k=1}^{n}\frac{1}{G(Y_k)}\psi_{ij}(X_k),$$
$$\hat{b}_j = \frac{\theta}{n}\sum_{k=1}^{n}\frac{Y_k}{G(Y_k)}\phi_j(X_k), \quad \hat{b}_{ij} = \frac{\theta}{n}\sum_{k=1}^{n}\frac{Y_k}{G(Y_k)}\psi_{ij}(X_k).$$

Then, under the assumption $\alpha(n) = O(n^{-r})$ *for some* $r > 3$, *we have*

$$\hat{a}_j = \tilde{a}_j + A_j, \quad \hat{a}_{ij} = \tilde{a}_{ij} + A_{ij}, \quad \hat{b}_j = \tilde{b}_j + B_j, \quad \hat{b}_{ij} = \tilde{b}_{ij} + B_{ij},$$

where

$$A_j = O\Big(\Big(\frac{\ln\ln(n)}{n}\Big)^{1/2}\Big)\cdot\frac{\theta}{n}\sum_{k=1}^{n}\frac{|\phi_j(X_k)|}{G(Y_k)} \quad a.s.,$$
$$A_{ij} = O\Big(\Big(\frac{\ln\ln(n)}{n}\Big)^{1/2}\Big)\cdot\frac{\theta}{n}\sum_{k=1}^{n}\frac{|\psi_{ij}(X_k)|}{G(Y_k)} \quad a.s.,$$
$$B_j = O\Big(\Big(\frac{\ln\ln(n)}{n}\Big)^{1/2}\Big)\cdot\frac{\theta}{n}\sum_{k=1}^{n}\frac{|Y_k\phi_j(X_k)|}{G(Y_k)} \quad a.s.,$$
$$B_{ij} = O\Big(\Big(\frac{\ln\ln(n)}{n}\Big)^{1/2}\Big)\cdot\frac{\theta}{n}\sum_{k=1}^{n}\frac{|Y_k\psi_{ij}(X_k)|}{G(Y_k)} \quad a.s..$$

Proof. We observe that

$$\begin{aligned}\hat{a}_j &= \frac{\theta}{n}\sum_{k=1}^{n}\frac{1}{G(Y_k)}\phi_j(X_k) + \Big[\frac{\theta_n-\theta}{n}\sum_{k=1}^{n}\frac{1}{G_n(Y_k)}\phi_j(X_k)\\ &\quad +\frac{\theta}{n}\sum_{k=1}^{n}\Big(\frac{1}{G_n(Y_k)}-\frac{1}{G(Y_k)}\Big)\phi_j(X_k)\Big]\\ &:= \tilde{a}_j + A_j.\end{aligned}$$

According to (3.1) and Lemma 6.4 we have

$$\begin{aligned}|A_j| &\leq \frac{|\theta_n-\theta|}{nG_n(a_F)}\sum_{k=1}^{n}|\phi_j(X_k)| + \frac{\theta\sup_y|G_n(y)-G(y)|}{nG_n(a_F)}\sum_{k=1}^{n}\frac{|\phi_j(X_k)|}{G(Y_k)}\\ &\leq \frac{|\theta_n-\theta|}{n[G(a_F)-\sup_y|G_n(y)-G(y)|]}\sum_{k=1}^{n}|\phi_j(X_k)|\\ &\quad +\frac{\theta\sup_y|G_n(y)-G(y)|}{n[G(a_F)-\sup_y|G_n(y)-G(y)|]}\sum_{k=1}^{n}\frac{|\phi_j(X_k)|}{G(Y_k)}\\ &= O\Big(\Big(\frac{\ln\ln(n)}{n}\Big)^{1/2}\Big)\cdot\frac{\theta}{n}\sum_{k=1}^{n}\frac{|\phi_j(X_k)|}{G(Y_k)} \quad a.s..\end{aligned}$$

The other quantities can be analyzed in the same manner. ■

Lemma 6.6 (Bradley (1983), Theorem 3) *Let η and ξ be real-valued random variables. Suppose U is a uniform-$[0,1]$ random variable, independent of (η,ξ). Then there exists a real-valued random variable ξ^*, measurable w.r.t. (η,ξ,U), such that*

(1) ξ^* is independent of η,

(2) the probability distributions of ξ^* and ξ are identical, and

(3) $P(|\xi^*-\xi|\geq\epsilon)\leq 18(\|\xi\|_r/\epsilon)^{r/(2r+1)}\{\sup_{A\in\sigma(\xi),B\in\sigma(\eta)}|P((A\cap B)-P(A)P(B)|\}^{2r/(2r+1)}$, where $0<\epsilon\leq\|\xi\|_r$, when $\|\xi\|_r>0$, and $\epsilon>0$, when $\|\xi\|_r=0$ and $\|\xi\|_\infty=\operatorname{ess\,sup}|\xi|$.

Remark 6.1 *Bradley (1983), Theorem 3, considers only the case $\|\xi\|_r>0$. Actually, if $\|\xi\|_r=0$, then $\xi=0$ a.s.; hence, on choosing $\xi^*=\xi=0$ a.s., then, for any $\epsilon>0$, (1), (2) and (3) in Lemma 6.6 are still true.*

Lemma 6.7 *Let $\tau>2$. Under the assumptions of Theorem 3.1, if $\lambda\geq(\tau-1)(2\tau+1)(2-\epsilon)/(2\epsilon(\tau-2))$, then $E|\tilde{a}_{ij}-a_{ij}|^\tau=O(n^{-\tau/2})$, $E|\tilde{b}_{ij}-b_{ij}|^\tau=O(n^{-\tau/2})$.*

Proof. Following the lines of Lemma 4.5 in Liang et al. (2005), one can verify Lemma 6.7. For the sake of completeness, here we give the proof of the second equation, the proof of the first equation is analogous. Choosing $r(n) = [(n/p_q)^{(\tau-2)/(2(\tau-1))}]$, and positive integers $k(n)$ and $\gamma(n)$ such that $n = r(n)k(n) + \gamma(n)$, with $0 \le \gamma(n) < r(n)$. Set $W_k = \frac{1}{n}(\frac{\theta Y_k \psi_{ij}(X_k)}{G(Y_k)} - b_{ij})$. Then

$$\tilde{b}_{ij} - b_{ij} = \sum_{l=1}^{k(n)} \sum_{j=(l-1)r(n)+1}^{lr(n)} W_j + \sum_{j=r(n)k(n)+1}^{n} W_j.$$

The contribution of the remainder term $\sum_{i=r(n)k(n)+1}^{n} W_i$ is negligible (and is subsequently ignored) since it consists of at most $r(n)$ terms. So, without loss of generality, we assume $\gamma(n) = 0$, and further $k(n) = 2s(n)$. Then

$$\begin{aligned} \tilde{b}_{ij} - b_{ij} &= \sum_{l=1}^{2s(n)} \sum_{j=(l-1)r(n)+1}^{lr(n)} W_j := \sum_{l=1}^{2s(n)} \xi_n(l) \\ &= \sum_{l=1}^{s(n)} \xi_n(2l) + \sum_{l=1}^{s(n)} \xi_n(2l-1) := S(n) + T(n), \end{aligned} \tag{6.1}$$

where $\xi_n(l) = \sum_{j=(l-1)r(n)+1}^{lr(n)} W_j$. Hence $E|\tilde{\beta}_{ij} - b_{ij}|^\tau \le C\{E|S(n)|^\tau + E|T(n)|^\tau\}$. Next, we evaluate only $E|T(n)|^\tau$, since the evaluation of $E|S(n)|^\tau$ is similar. In view of Lemma 6.6, there exist i.i.d. random variables $\xi_n^*(2l-1)$, $l = 1, 2, \ldots, s(n)$ such that $\xi_n^*(2l-1)$ has the same distribution as $\xi_n(2l-1)$ for each l, and satisfies

$$P(|\xi_n^*(2l-1) - \xi_n(2l-1)| \ge \epsilon_l) \le 18(\frac{\|\xi_n(2l-1)\|_\infty}{\epsilon_l})^{1/2} \alpha(r(n)), \tag{6.2}$$

where $0 < \epsilon_l \le \|\xi_n(2l-1)\|_\infty$, if $\|\xi_n(2l-1)\|_\infty > 0$, and $\epsilon_l > 0$, if $\|\xi_n(2l-1)\|_\infty = 0$. Then,

$$\begin{aligned} E|T(n)|^\tau &\le C\{E|\sum_{l=1}^{s(n)} \xi_n^*(2l-1)|^\tau + E|\sum_{l=1}^{s(n)} (\xi_n^*(2l-1) - \xi_n(2l-1))|^\tau\} \\ &:= C\{T_1(n) + T_2(n)\}. \end{aligned}$$

Let us take $M_n > 0$ such that $s(n)M_n \asymp n^{-1/2}$, where $a_n \asymp b_n$ means $0 < \liminf a_n/b_n \le \limsup a_n/b_n < \infty$, and assume $\|\xi_n(2l-1)\|_\infty \ge M_n$, for $l = 1, 2, \ldots, s(n)$. Otherwise, by rearranging the terms appropriately, we may assume, without loss of generality, that $\|\xi_n(2l-1)\|_\infty \ge M_n$, for $l = 1, 2, \ldots, s_1(n)$, and $\|\xi_n(2l-1)\|_\infty < M_n$, for $l = s_1(n)+1, \ldots, s(n)$, where $s_1(n)$ is a positive integer with $s_1(n) \le s(n)$, in this case we have

$$T_2(n) \le C\Big\{(M_n s(n))^\tau + E\Big(\sum_{l=1}^{s_1(n)} |\xi_n^*(2l-1) - \xi_n(2l-1)|\Big)^\tau\Big\}.$$

Therefore,

$$T_2(n) \le C\Big\{(M_n s(n))^\tau + E\Big(\sum_{=1}^{s(n)} |\xi_n^*(2l-1) - \xi_n(2l-1)| I(|\xi_n^*(2l-1) - \xi_n(2l-1))| \ge M_n)\Big)^\tau\Big\},$$

where $\|\xi_n(2l-1)\|_\infty \ge M_n$. Observe that

$$|\xi_n^*(2l-1) - \xi_n(2l-1)| \le 2r(n)\Big(\frac{\theta b_F p_i^{1/2}\|\psi\|_\infty}{G(a_F)} + |b_{ij}|\Big)\frac{1}{n} \le \frac{C}{n} r(n) p_q^{1/2}.$$

Note that $p_q\delta^2 = O(n^{-\epsilon})$, $\delta \ge C_3(n^{-1}\log n)^{1/2}$ and $\lambda \ge (\tau-1)(2\tau+1)(2-\epsilon)/(2\epsilon(\tau-2))$ imply $n^{-(\frac{\lambda(\tau-2)}{2(\tau-1)}-\frac{1}{4})} p_q^{\frac{\lambda(\tau-2)}{2(\tau-1)}+\frac{\tau}{2}+\frac{1}{4}} = o(n^{-\tau/2})$. Then, according to (6.2) and $M_n s(n) = O(n^{-1/2})$, it follows that

$$\begin{aligned} T_2(n) &\le C\Big\{\Big(\frac{1}{n} r(n) p_q^{1/2}\Big)^\tau (s(n))^{\tau-1} \sum_{l=1}^{s(n)} P(|\xi_n^*(2l-1) - \xi_n(2l-1))| \ge M_n)\Big\} + O(n^{-\tau/2}) \\ &\le C\Big(\frac{1}{n} r(n) p_q^{1/2}\Big)^\tau (s(n))^\tau \Big(\frac{r(n)p_q^{1/2}}{nM_n}\Big)^{1/2} (r(n))^{-\lambda} + O(n^{-\tau/2}) \\ &\le C n^{-(\frac{\lambda(\tau-2)}{2(\tau-1)}-\frac{1}{4})} p_q^{\frac{\lambda(\tau-2)}{2(\tau-1)}+\frac{\tau}{2}+\frac{1}{4}} + O(n^{-\tau/2}) = O(n^{-\tau/2}). \end{aligned}$$

Next, we estimate $T_1(n)$. Applying the Rosenthal inequality for sums of independent random variables (cf. Petrov (1995), Theorem 2.9, page 59), we get

$$\begin{aligned} T_1(n) &\le C\Big\{\sum_{l=1}^{s(n)} E|\xi_n^*(2l-1)|^\tau + \Big(\sum_{l=1}^{s(n)} E(\xi_n^*(2l-1))^2\Big)^{\tau/2}\} \\ &\le C\{s(n)E|\xi_n(1)|^\tau + [s(n)E(\xi_n(1))^2]^{\tau/2}\Big\}. \end{aligned} \tag{6.3}$$

From (3.1) we have

$$\begin{aligned} E|\xi_n(1)|^\tau &= E\Big|\sum_{k=1}^{r(n)} W_k\Big|^\tau \le (r(n))^\tau E|W_1|^\tau \le (r(n))^\tau \Big(\frac{\theta b_F}{nG(a_F)}\Big)^\tau E|\psi_{ij}(X_1)|^\tau \\ &\le C(r(n))^\tau n^{-\tau} \cdot p_i^{\tau/2-1} \int |\psi(u)|^\tau v(\frac{u+j}{p_i}) du \\ &\le C(r(n))^\tau n^{-\tau} p_q^{\tau/2-1}. \end{aligned}$$

Then

$$s(n)E|\xi_n(1)|^\tau = O(n^{-\tau/2}). \tag{6.4}$$

As to $E(\xi_n(1))^2$, by using Lemma 6.3, it follows that

$$E(\xi_n(1))^2 = E\Big|\sum_{k=1}^{r(n)} W_k\Big|^2 \le r(n)\{C(R_\infty(r(n))^{2/\lambda}(R^*(r(n)))^{1-1/\lambda} + R_2^2(r(n))\},$$

where

$$R_\infty(r(n)) := \sup_{1\le k\le r(n)} esssup_{w\in\Omega}|W_k| \le C\Big(\frac{\theta b_F p_i^{1/2}\|\psi\|_\infty}{G(a_F)} + |b_{ij}|\Big)\frac{1}{n} = O(p_q^{1/2}/n);$$
$$R_2^2(r(n)) := E|W_1|^2 \le \frac{C}{n^2}\int |\psi(u)|^2 v(\frac{u+j}{p_i})du = O(n^{-2});$$
$$R^*(r(n)) := \sup_{1\le s,t\le r(n), s\ne t} |\mathrm{Cov}(W_s, W_t)| \le \frac{C}{n^2 p_q}.$$

Therefore, $E(\xi_n(1))^2 \le Cr(n)n^{-2}$, and $s(n)E(\xi_n(1))^2 \le Cs(n)r(n)n^{-2} = O(n^{-1})$, which, together with (6.3) and (6.4), yields $T_1(n) = O(n^{-\tau/2})$. ■

Proof of (4.11). Assume that supp$\phi \subseteq (-L, L)$. Set $V_{jk} = \frac{\theta\phi_j(X_k)}{G(Y_k)} - a_j$. Then $EV_{jk} = 0$, $\|V_{jk}\|_\infty = O(p^{1/2})$, $E|V_{jk}| = O(p^{-\frac{1}{2}})$, and

$$n^2\sum_{j=0}^{p-1}(\tilde{a}_j - a_j)^2 = \sum_{k=1}^{n}\sum_{j=0}^{p-1} V_{jk}^2 + \sum_{1\le k_1,k_2\le n, k_1\ne k_2}\sum_{j=0}^{p-1} V_{jk_1}V_{jk_2}.$$

Hence

$$\mathrm{Var}\Big\{n^2\sum_{j=0}^{p-1}(\tilde{a}_j - a_j)^2\Big\} \le C\Big\{\mathrm{Var}\Big(\sum_{k=1}^{n}\sum_{j=0}^{p-1}V_{jk}^2\Big) + E\Big(\sum_{k_1\ne k_2}\sum_{j=0}^{p-1}V_{jk_1}V_{jk_2}\Big)^2\Big\}. \tag{6.5}$$

It is easy to see that

$$n^{-1}\mathrm{Var}\Big(\sum_{k=1}^{n}\sum_{j=0}^{p-1}V_{jk}^2\Big) = n^{-1}E\Big\{\sum_{k=1}^{n}\sum_{j=0}^{p-1}(V_{jk}^2 - EV_{jk}^2)\Big\}^2$$
$$= E\Big\{\sum_{j=0}^{p-1}V_{j1}^2 - E\Big(\sum_{j=0}^{p-1}V_{j1}^2\Big)\Big\}^2 + 2\sum_{l=1}^{n-1}(1-\frac{l}{n})\mathrm{Cov}\Big(\sum_{j=0}^{p-1}V_{j1}^2, \sum_{j=0}^{p-1}V_{j,1+l}^2\Big).$$

Since ϕ has compact support and it is a bounded function,

$$\frac{1}{2}\sum_{j=0}^{p-1}V_{jk}^2 \le \sum_{j=0}^{p-1}\frac{\theta^2\phi_j^2(X_1)}{G^2(Y_1)} + \sum_{j=0}^{p-1}a_j^2$$
$$\le \frac{\theta^2 p}{G^2(a_F)}\sum_{j=0}^{p-1}\phi^2(pX_1 - j) + \sum_{j=0}^{p-1}\frac{1}{p}\Big(\int\phi(u)v(\frac{u+j}{p})du\Big)^2 = O(p).$$

Hence, by Lemma 6.1 we have

$$\sum_{l=1}^{n-1}\Big|\mathrm{Cov}\Big(\sum_{j=0}^{p-1}V_{j1}^2, \sum_{j=0}^{p-1}V_{j,1+l}^2\Big)\Big| = O\Big(p^2\sum_{l=1}^{n-1}\alpha(l)\Big) = O(p^2).$$

This yields $\mathrm{Var}(\sum_{k=1}^{n}\sum_{j=0}^{p-1}V_{jk}^2)=O(np^2)=o(n^2p^2)$. Therefore, from (6.5), it suffices to show that

$$E\Big(\sum_{k_1\neq k_2}\sum_{j=0}^{p-1}V_{jk_1}V_{jk_2}\Big)^2=E\Big(\sum_{k_{11}\neq k_{12}}\sum_{k_{21}\neq k_{22}}\sum_{j_1=0}^{p-1}\sum_{j_2=0}^{p-1}V_{j_1k_{11}}V_{j_1k_{12}}V_{j_2k_{21}}V_{j_2k_{22}}\Big)=o(n^2p^2). \quad (6.6)$$

In order to verify (6.6), we consider the sums above by several cases of the indices.

Case 1. Suppose the indices satisfy $k_{11}=k_{21}=k_1$ and $k_{12}=k_{22}=k_2$. First, when $|j_1-j_2|\leq 4L$,

$$\begin{aligned}
&E\Big(\sum_{k_1\neq k_2}\sum_{j_1=0}^{p-1}\sum_{j_2=0}^{p-1}V_{j_1k_1}V_{j_1k_2}V_{j_2k_1}V_{j_2k_2}\Big)\leq Cpn^2\sum_{j_1=0}^{p-1}\sum_{j_2:|j_1-j_2|\leq 4L}E|V_{j_11}V_{j_12}|\\
&\leq Cpn^2L\sum_{j_1=0}^{p-1}E\Big|\Big(\frac{\theta p^{1/2}\phi(pX_1-j_1)}{G(Y_1)}-a_{j_1}\Big)\Big(\frac{\theta p^{1/2}\phi(pX_2-j_1)}{G(Y_2)}-a_{j_1}\Big)\Big|\\
&\leq Cpn^2L\sum_{j_1=0}^{p-1}\Big\{\frac{\theta^2p}{G^2(a_F)}E|\phi(pX_1-j_1)\phi(pX_2-j_1)|+a_{j_1}^2\Big\}\\
&\leq Cpn^2L\sum_{j_1=0}^{p-1}\Big\{\frac{\theta^2}{G^2(a_F)}\int\frac{1}{p}|\phi(u)|v(\frac{u+j_1}{p}\Big)du+p^{-1}\Big\}\\
&=O(n^2p).
\end{aligned} \quad (6.7)$$

When $|j_1-j_2|>4L$, since ϕ is supported on $(-L,L)$, at least one among $\phi_{j_1}(X_{k_1})$ and $\phi_{j_2}(X_{k_1})$ is zero, the same being true for $\phi_{j_1}(Z_{k_2})$ and $\phi_{j_2}(Z_{k_2})$. Here, we will assume $\phi_{j_1}(X_{k_1})=0,\phi_{j_1}(Z_{k_2})=0$, the other cases are investigated similarly. In this case, similarly to the arguments as in (6.7) we have

$$E\Big(\sum_{k_1\neq k_2}\sum_{j_1=0}^{p-1}\sum_{j_2=0}^{p-1}V_{j_1k_1}V_{j_1k_2}V_{j_2k_1}V_{j_2k_2}\Big)\leq\sum_{k_1\neq k_2}\sum_{j_1=0}^{p-1}a_{j_1}^2\sum_{j_2=0}^{p-1}E\Big|V_{j_2k_1}V_{j_2k_2}\Big|=O(n^2).$$

Therefore, (6.6) holds.

Case 2. Suppose the indices satisfy $k_{11}=k_{21}$ and $k_{11}<k_{12}<k_{22}$. Then

$$\begin{aligned}
&E\Big(\sum_{1\leq k_{11}<k_{12}<k_{22}\leq n}\sum_{j_1=0}^{p-1}\sum_{j_2=0}^{p-1}V_{j_1k_{11}}V_{j_1k_{12}}V_{j_2k_{11}}V_{j_2k_{22}}\Big)\\
&=E\Big(\sum_{l_1=1}^{n-2}\sum_{l_2=l_1+1}^{n-1}\sum_{l_3=l_2+1}^{n}\sum_{j_1=0}^{p-1}\sum_{j_2=0}^{p-1}V_{j_1l_1}V_{j_1l_2}V_{j_2l_1}V_{j_2l_3}\Big)\\
&=\sum_{l_1=1}^{n-2}\sum_{l_2=1}^{n-l_1-1}[n-(l_1+l_2)]\sum_{j_1=0}^{p-1}\sum_{j_2=0}^{p-1}E\Big(V_{j_11}V_{j_1(1+l_1)}V_{j_21}V_{j_2(1+l_1+l_2)}\Big)\\
&\leq n\sum_{l_1=1}^{n-2}\sum_{l_2=1}^{n-l_1-1}\sum_{j_1=0}^{p-1}\sum_{j_2=0}^{p-1}\Big|E\Big(V_{j_11}V_{j_1(1+l_1)}V_{j_21}V_{j_2(1+l_1+l_2)}\Big)\Big|.
\end{aligned} \quad (6.8)$$

Take $Q_n = \delta^{-2/(2r+1)}$. Introduce $D_1 = \{(l_1, l_2) : l_1 \le Q_n,\ l_2 \le Q_n,\ 1 < l_1 + l_2 < n\}$,

$$D_2 = \{(l_1, l_2) : l_1 \le Q_n, l_2 > Q_n,\ 1 < l_1 + l_2 < n\},\ D_3 = \{(l_1, l_2) : l_1 > Q_n,\ 1 < l_1 + l_2 < n\}.$$

Then, according to Lemma 6.1, from $EV_{j_2(1+l_1+l_2)} = 0$ and $|V_{jk}| \le Cp^{1/2}$, it follows that

$$\begin{aligned}
& \sum_{l_1=1}^{n-2}\sum_{l_2=1}^{n-l_1-1} \Big|E\Big(V_{j_1 1}V_{j_1(1+l_1)}V_{j_2 1}V_{j_2(1+l_1+l_2)}\Big)\Big| \\
\le\ & \Big(\sum_{(l_1,l_2)\in D_1} + \sum_{(l_1,l_2)\in D_2} + \sum_{(l_1,l_2)\in D_3}\Big)\Big|E\Big(V_{j_1 1}V_{j_1(1+l_1)}V_{j_2 1}V_{j_2(1+l_1+l_2)}\Big)\Big| \\
\le\ & \sum_{(l_1,l_2)\in D_1} \Big|E\Big(V_{j_1 1}V_{j_1(1+l_1)}V_{j_2 1}V_{j_2(1+l_1+l_2)}\Big)\Big| \\
& + Cp^2 \sum_{(l_1,l_2)\in D_2} \alpha(l_2) + \sum_{(l_1,l_2)\in D_2} \Big|E\Big(V_{j_1 1}V_{j_1(1+l_1)}V_{j_2 1}\Big)EV_{j_2(1+l_1+l_2)}\Big| \\
& + Cp^2 \sum_{(l_1,l_2)\in D_3} \alpha(l_1) + \sum_{(l_1,l_2)\in D_3} \Big|E\Big(V_{j_1 1}V_{j_2 1}\Big)E\Big(V_{j_1(1+l_1)}V_{j_2(1+l_1+l_2)}\Big)\Big| \\
\le\ & \sum_{(l_1,l_2)\in D_1} \Big|E\Big(V_{j_1 1}V_{j_1(1+l_1)}V_{j_2 1}V_{j_2(1+l_1+l_2)}\Big)\Big| + Cnp^2 Q_n^{-(\lambda-1)} \\
& + C\sum_{(l_1,l_2)\in D_3} \Big|E\Big(V_{j_1(1+l_1)}V_{j_2(1+l_1+l_2)}\Big)\Big|. \qquad (6.9)
\end{aligned}$$

Note that $\sum_{(l_1,l_2)\in D_3} |E(V_{j_1(1+l_1)}V_{j_2(1+l_1+l_2)})| = o(n)$ from (4.10), $|E(V_{j_1 1}V_{j_1(1+l_1)}V_{j_2 1}V_{j_2(1+l_1+l_2)})| \le CpE|V_{j_1 1}V_{j_1(1+l_1)}| = O(1)$, and that $\lambda \ge 3+4r$, $p_q\delta^2 = O(n^{-\epsilon})$ and $\delta \ge C_3(n^{-1}\log n)^{1/2}$ implies $Q_n^2/n = O(n^{-(2r-1)/(2r+1)}(\ln(n))^{-2/(2r+1)}) \to 0$ and

$$p^2 Q_n^{-(\lambda-1)} = O(p^{-[\lambda-(3+4r)]/(2r+1)}2^{-q(\lambda-1)/(2r+1)}n^{-\epsilon(\lambda-1)/(2r+1)}) \to 0.$$

So, from (6.8) and (6.9) we find

$$\begin{aligned}
& E\Big(\sum_{1\le k_{11}<k_{12}<k_{22}\le n}\sum_{j_1=0}^{p-1}\sum_{j_2=0}^{p-1} V_{j_1k_{11}}V_{j_1k_{12}}V_{j_2k_{11}}V_{j_2k_{22}}\Big) \\
& \le Cnp^2Q_n^2 + Cn^2p^4Q_n^{-(\lambda-1)} + o(n^2p^2) = o(n^2p^2).
\end{aligned}$$

Case 3. Suppose the indices satisfy $k_{11} < k_{12} < k_{21} < k_{22}$. By the stationarity of X_k we have

$$
\begin{aligned}
&E\Big(\sum_{1\le k_{11}<k_{12}<k_{21}<k_{22}\le n}\sum_{j_1=0}^{p-1}\sum_{j_2=0}^{p-1} V_{j_1k_{11}}V_{j_1k_{12}}V_{j_2k_{21}}V_{j_2k_{22}}\Big)\\
&=\sum_{l_1=1}^{n-3}\sum_{l_2=1}^{n-l_1-2}\sum_{l_3=1}^{n-l_1-l_2-1}\sum_{l_4=1}^{n-(l_1+l_2+l_3)}\sum_{j_1=0}^{p-1}\sum_{j_2=0}^{p-1} E\Big(V_{j_11}V_{j_1(1+l_2)}V_{j_2(1+l_2+l_3)}V_{j_2(1+l_2+l_3+l_4)}\Big)\\
&=\sum_{l_1=1}^{n-3}\sum_{l_2=1}^{n-l_1-2}\sum_{l_3=1}^{n-l_1-l_2-1}[n-(l_1+l_2+l_3)]\sum_{j_1=0}^{p-1}\sum_{j_2=0}^{p-1} E\Big(V_{j_11}V_{j_1(1+l_1)}V_{j_2(1+l_1+l_2)}V_{j_2(1+l_1+l_2+l_3)}\Big).
\end{aligned}
$$

Let $E_1 = \{(l_1, l_2, l_3) : l_1 \le Q_n,\ l_2 \le Q_n,\ l_3 \le Q_n,\ 1 < l_1 + l_2 + l_3 < n\}$,

$$
\begin{aligned}
E_2 &= \{(l_1, l_2, l_3) : l_1 \le Q_n,\ l_2 \le Q_n,\ l_3 > Q_n,\ 1 < l_1 + l_2 + l_3 < n\},\\
E_3 &= \{(l_1, l_2, l_3) : l_1 \le Q_n,\ l_2 > Q_n,\ 1 < l_1 + l_2 + l_3 < n\},\\
E_4 &= \{(l_1, l_2, l_3) : l_1 > Q_n,\ 1 < l_1 + l_2 + l_3 < n\}.
\end{aligned}
$$

Note that $|E(V_{j_11}V_{j_1(1+l_1)})| = O(p^{-1})$ and (4.10) imply that

$$
\sum_{(l_1,l_2,l_3)\in E_3}\Big|E\Big(V_{j_11}V_{j_1(1+l_1)}\Big)E\Big(V_{j_2(1+l_1+l_2)}V_{j_2(1+l_1+l_2+l_3)}\Big)\Big| = o(nQ_np^{-1}).
$$

$\delta \ge C_3(n^{-1}\log n)^{1/2}$ implies $n^{-1}Q_n^3 \le Cn^{-2(r-1)}(\ln(n))^{-3/(2r+1)} \to 0$; $\lambda \ge \max\{1 + (2r + 1)/\epsilon, 3 + 4r\}$ and $p_q\delta^2 = O(n^{-\epsilon})$ imply

$$
np^2/Q_n^{\lambda-1} \le Cn^{-[\epsilon\lambda-(2r+1+\epsilon)]/(2r+1)}2^{-q(\lambda-1)/(2r+1)}p^{-[\lambda-(4r+3)]/(2r+1)} \to 0.
$$

Hence, from $p^{2r+1}\delta^2 \to \infty$, similarly to the arguments as in *Case 2*, it follows that

$$
\begin{aligned}
&E\Big(\sum_{1\le k_{11}<k_{12}<k_{21}<k_{22}\le n}\sum_{j_1=0}^{p-1}\sum_{j_2=0}^{p-1} V_{j_1k_{11}}V_{j_1k_{12}}V_{j_2k_{21}}V_{j_2k_{22}}\Big)\\
&\le Cnp^2Q_n^3 + Cn^3p^4Q_n^{-(\lambda-1)} + o(n^2pQ_n) = o(n^2p^2).
\end{aligned}
$$

■