Ann. Inst. Statist. Math.
Vol. 50, No. 4, 729–754 (1998)

RECURSIVE ESTIMATION OF REGRESSION FUNCTIONS BY LOCAL POLYNOMIAL FITTING

J. A. VILAR-FERNÁNDEZ AND J. M. VILAR-FERNÁNDEZ

Departamento de Matemáticas, Facultad de Informática, Universidad de A Coruña, 15071 A Coruña, Spain

(Received March 6, 1996; revised October 27, 1997)

Abstract. The recursive estimation of the regression function $m(x) = E(Y/X = x)$ and its derivatives is studied under dependence conditions. The examined method of nonparametric estimation is a recursive version of the estimator based on locally weighted polynomial fitting, that in recent articles has proved to be an attractive technique and has advantages over other popular estimation techniques. For strongly mixing processes, expressions for the bias and variance of these estimators are given and asymptotic normality is established. Finally, a simulation study illustrates the proposed estimation method.

Key words and phrases: Recursive nonparametric estimation, regression models, local polynomial fitting, strongly mixing processes.

1. Introduction

Nonparametric regression is a smoothing method for estimating the regression function from noisy data and it has become a powerful and useful diagnostic tool for data analysis. See the monographs of Eubank (1988), Härdle (1990) and Müller (1988) for a good introduction and many interesting examples of specific applications of this method with real data.

Let $\{X_t, Y_t\}_{t=1}^n$ be a observed sample of the stationary stochastic processes $\{X, Y\} = \{X_t, Y_t; -\infty < t < \infty\}$ with unknown joint density $f(x, y)$ and denote the marginal density of X by $f_X(x)$ and the conditional variance of Y given x by $\sigma_Y^2(x)$. Several smoothing methods have been proposed for estimating the regression function, $m(x) = E(Y/X = x)$. Among the most extensively analyzed, we find: kernel, spline and orthogonal series methods.

Recently, local polynomial fitting has gained acceptance as an attractive method for the nonparametric estimation of $m(x)$. It was introduced by Stone (1977) and studied by Cleveland (1979), Lejeune (1985), Müller (1988), Cleveland and Devlin (1988), Fan (1992), in a context of independent observations, and by Masry and Fan (1997) and Masry (1996*a*, 1996*b*) for dependent observations. The motivation and study of this smoothing method can be found in the recent monograph of Fan and Gijbels (1996). The advantages of local polynomial fitting include its simplicity, its easy interpretation and computation, its nice minimax properties (Fan (1993), Fan *et al.* (1997)), its adaptation to the boundary of design points (Fan and Gijbels (1992), Hastie and Loader (1993)), its application to various design situations and its adaptation to the estimation of derivatives of $m(x)$ (Ruppert and Wand (1994), Fan and Gijbels (1995)).

If the $(p+1)$-th derivative of the regression function at the point x exits, local polynomial fitting permits estimating the parameter vector $\vec{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^t$, where $\beta_j = m^{(j)}(x)/(j!)$, with $j = 0, 1, \ldots, p$, by minimizing the function

$$\Psi_{(n)}(\vec{\beta}) = \sum_{t=1}^{n} \left(Y_t - \sum_{k=0}^{p} \beta_k (X_t - x)^k \right)^2 \omega_t^{(n)}, \tag{1.1}$$

where the weights are of the form $\omega_t^{(n)} = n^{-1} K_n(X_t - x)$ with $K_n(u) = (\frac{1}{h_n})K(\frac{u}{h_n})$, $K(\cdot)$ a kernel function and h_n a sequence of smoothing parameters. In particular, when $p = 0$, the minimizer of expression (1.1) is the classical Nadaraya-Watson estimator of $m(x)$.

In this report we study the local polynomial fitting modifying the weights $\omega_t^{(n)}$ as follows

$$\omega_t^{(n)} = \frac{1}{nh_t} K\left(\frac{X_t - x}{h_t}\right) = \frac{1}{n} K_t(X_t - x). \tag{1.2}$$

The problem of minimizing expression (1.1) with the sequence of weights proposed in (1.2) would then provide estimators $\hat{\vec{\beta}}_{(n)} = (\hat{\beta}_0^{(n)}, \hat{\beta}_1^{(n)}, \ldots, \hat{\beta}_p^{(n)})^t$ which are recursive, as will be proved in Section 2 using the recursivity algorithm of Plackett-Kalman for regression models. It is important to point out that, by making $p = 0$ again, we obtain as a solution estimator $\hat{\beta}_0^{(n)}$, which coincides with the recursive version of the Nadaraya-Watson kernel estimator, whose explicit expression is

$$\hat{\beta}_0^{(n)}(x) = \frac{\sum_{t=1}^{n} K_t(X_t - x) Y_t}{\sum_{t=1}^{n} K_t(X_t - x)}. \tag{1.3}$$

As it is well known, recursive estimators are specially useful when the observations are gathered sequentially. In this context, the recursive techniques allow easy updating of the estimates as additional observations are obtained, unlike non recursive methods where estimates must be completely recalculated when each additional item of data is received. Thus, the recursive estimates provide two important advantages with respect to non recursive ones. First, the fact that not all data have to be stored in order to compute the estimates with each additional observation leads to considerable saving of memory. Second, since it is not necessary to evaluate all the sample data again when the recursive estimator is used, each updating of the estimate is carried out independently of the initial sample size, unlike non recursive methods where the computing time increases with the sample size. Moreover, as it will be indicated in Section 4, the recursive algorithm we propose in next section allows to update the estimation, for one fixed point, with a speed $2/n$ times that of his analogous nonrecursive method, when direct implementations are employed and n denotes the sample size. From the above considerations it is clear that the recursive property is particularly interesting when larger sample sizes are considered.

These computational advantages have a cost in terms of inefficiency. In fact, the recursive estimates exhibit larger mean square error than the non recursive ones, even though both show the same convergence rates. In any case, their use in sequential methods or in the study of temporal processes seems advisable. Works studying recursive nonparametric estimators of the regression function are those by Devroye and Wagner (1980), Krzyzak and Pawlak (1984), Greblicky and Pawlak (1987), Roussas and Tran (1992) and Rousas (1992), among others.

From the previous paragraph, the use of recursive estimates is specially appropriate in time series analysis. However, in this setting, the assumption of sampling independence is not realistic so that the present study has been performed for dependent observations. In fact, we have adopted a standard approach to model the dependence which assumes that the processes $\{X_t, Y_t\}$ satisfy some mixing condition. In particular, the strongly mixing (α-mixing) condition has been considered.

This paper is organized as follows: in Section 2, the proposed estimation method is shown to be recursive and a iterative algorithm is derived. In Section 3, preliminary results are established, expressions for the bias and variance of the studied estimators are obtained and asymptotic normality is shown. In Section 4 a simulation study is performed and, finally, an outline of the proofs are presented in the Appendix.

2. Recursive algorithm

In the first place, it is more convenient to write the previous weighted least squares problem (given in (1.1)) in matrix notation.

Let us denote $\vec{Y}_{(n)}^t = (Y_1, \ldots, Y_n)$, let $X_{(n)}$ be the matrix design for a sample of size n, whose (i,j)-th element is $(X_i - x)^{j-1}$ with $i = 1, \ldots, n$ and $j = 1, \ldots, p+1$, and let $W_{(n)}$ be the diagonal matrix $W_{(n)} = \mathrm{diag}(\omega_1^{(n)}, \omega_2^{(n)}, \ldots, \omega_n^{(n)})$, where $\omega_t^{(n)}$ is given by (1.2).

Then, the function $\Psi_{(n)}(\vec{\beta})$ can be written in matrix form as

$$\Psi_{(n)}(\vec{\beta}) = (\vec{Y}_{(n)} - X_{(n)}\vec{\beta})^t W_{(n)} (\vec{Y}_{(n)} - X_{(n)}\vec{\beta}). \tag{2.1}$$

Minimizing this function provides the estimator

$$\hat{\vec{\beta}}_{(n)} = (X_{(n)}^t W_{(n)} X_{(n)})^{-1} (X_{(n)}^t W_{(n)} \vec{Y}_{(n)}) = S_{(n)}^{-1} \vec{T}_{(n)}, \tag{2.2}$$

where $S_{(n)}$ is the matrix $(p+1)\times(p+1)$ whose (r,t)-th element is $S_{r,t}^{(n)} = S_{r+t-2}^{(n)}$ with

$$S_j^{(n)} = \frac{1}{n}\sum_{t=1}^{n} (X_t - x)^j K_t(X_t - x), \qquad \text{for} \quad 0 \le j \le 2p, \tag{2.3}$$

and $\vec{T}_{(n)} = (T_0^{(n)}, T_1^{(n)}, \ldots, T_p^{(n)})^t$ with

$$T_j^{(n)} = \frac{1}{n}\sum_{t=1}^{n} (X_t - x)^j K_t(X_t - x) Y_t, \qquad \text{for} \quad 0 \le j \le p. \tag{2.4}$$

For estimator $\hat{\vec{\beta}}_{(n)}$ to be well defined, it is necessary for matrix $S_{(n)}$ to be non singular, and a sufficient condition for this is that at least in $(p+1)$ different points we find positive weights. For compactly supported kernel this is true with probability tending to one because we will assume that $nh_n \to \infty$. The problem can be also solved if a non compact kernel is employed (for instance, Gauss weights). Nevertheless, the lack of compactness presents the drawback of not being optimal and computationally slowly. This issue is studied in Seifert and Gasser (1996).

We now prove that the estimator defined in (2.2) is recursive. Let (X_{n+1}, Y_{n+1}) be an additional observation to the original sample of size n. According to (2.1) and (2.2), the new estimator of $\vec{\beta}$ based on the sample of size $n+1$ has the form

$$\begin{aligned} \hat{\vec{\beta}}_{(n+1)} &= (X^t_{(n+1)} W_{(n+1)} X_{(n+1)})^{-1} (X^t_{(n+1)} W_{(n+1)} \vec{Y}_{(n+1)}) \\ &= S^{-1}_{(n+1)} \vec{T}_{(n+1)} \end{aligned} \tag{2.5}$$

and it is straightforward to deduce that

$$S_{(n+1)} = \frac{n}{n+1} S_{(n)} + \omega^{(n+1)}_{n+1} \vec{x}_{(n+1)} \vec{x}^{\,t}_{(n+1)} \tag{2.6}$$

$$\vec{T}_{(n+1)} = \frac{n}{n+1} \vec{T}_{(n)} + \omega^{(n+1)}_{n+1} Y_{n+1} \vec{x}_{(n+1)} \tag{2.7}$$

where $\vec{x}_{(n+1)} = (1, (X_{n+1} - x), \ldots, (X_{n+1} - x)^p)^t$.

Substituting expressions (2.6) and (2.7) in (2.2), we obtain

$$(S_{(n+1)} - \omega^{(n+1)}_{n+1} \vec{x}_{(n+1)} \vec{x}^{\,t}_{(n+1)}) \hat{\vec{\beta}}_{(n)} = \vec{T}_{(n+1)} - \omega^{(n+1)}_{n+1} Y_{n+1} \vec{x}_{(n+1)}. \tag{2.8}$$

From (2.5) and (2.8) it can be deduced that

$$S_{(n+1)} (\hat{\vec{\beta}}_{(n+1)} - \hat{\vec{\beta}}_{(n)}) = \omega^{(n+1)}_{n+1} (Y_{n+1} - \vec{x}^{\,t}_{(n+1)} \hat{\vec{\beta}}_{(n)}) \vec{x}_{(n+1)},$$

therefore

$$\hat{\vec{\beta}}_{(n+1)} = \hat{\vec{\beta}}_{(n)} + \omega^{(n+1)}_{n+1} (Y_{n+1} - \vec{x}^{\,t}_{(n+1)} \hat{\vec{\beta}}_{(n)}) S^{-1}_{(n+1)} \vec{x}_{(n+1)}. \tag{2.9}$$

Thus, expression (2.9) indicates how to obtain the estimator $\hat{\vec{\beta}}_{(n+1)}$ from the previous estimator $\hat{\vec{\beta}}_{(n)}$ and the additional observation (X_{n+1}, Y_{n+1}).

However, the usefulness of (2.9) presents the drawback that with each additional observation it is necessary to recalculate the inverse of matrix $S_{(n+1)}$. In order to solve this problem we will make use of the following property of matrix algebra: if A is a non singular matrix and $\vec{v}$ is a vector, then

$$(A + \vec{v}\vec{v}^{\,t})^{-1} = A^{-1} - \frac{A^{-1}\vec{v}\vec{v}^{\,t}A^{-1}}{1 + \vec{v}^{\,t}A^{-1}\vec{v}}. \tag{2.10}$$

Applying property (2.10) to expression (2.6) with $A = (n/(n+1))S_{(n)}$ and $\vec{v} = \sqrt{\omega^{(n+1)}_{n+1}}\, \vec{x}_{(n+1)}$, we conclude that

$$S^{-1}_{(n+1)} = \left(1 + \frac{1}{n}\right) \left(S^{-1}_{(n)} - \frac{K_{n+1}(X_{n+1} - x) S^{-1}_{(n)} \vec{x}_{(n+1)} \vec{x}^{\,t}_{(n+1)} S^{-1}_{(n)}}{n + K_{n+1}(X_{n+1} - x) \vec{x}_{(n+1)} S^{-1}_{(n)} \vec{x}^{\,t}_{(n+1)}} \right). \tag{2.11}$$

From expressions (2.9) and (2.11) a recursive algorithm can be easily deduced for obtaining estimator $\hat{\vec{\beta}}_{(n+1)}$ of the regression function and its derivatives by local polynomial fitting. The practical application of this algorithm requires the definition of the initial values for the iteration. For this, we suggest to obtain an initial estimation from the model based on the first r observations ($r > p+1$) and then to modify recursively the model as new observations are gathered.

The second problem is the selection of the smoothing parameter associated with each new observation. In order to solve this problem it can be assumed that h_n has a functional form, for instance $h_n = Cn^{-\nu}$. In this case we could take $\nu = 1/5$ (which is the value that minimizes the MSE as will be shown in Section 3) and C may be computed using a plug-in type technique. Other alternative suggested by a referee is to consider the idea of global variable bandwidths (see Fan and Gijbels (1992)). For this, we should take $h_i = g_i/\alpha(X_i)$, with the usual choice $\alpha(x) = f_X(x)^\nu$, $\nu \in R$. This way provides an important motivation of (1.3) and it is very useful when curves are spatially inhomogeneous.

3. Asymptotic analysis

In this section the joint asymptotic normality of the estimator $\hat{\vec{\beta}}_{(n)}$ defined in (2.2) is established. In our analysis, the following assumptions will be used:

(A.1) The kernel function $K(u)$ is bounded with compact support.

(A.2) The sequence of bandwidths $\{h_n\}$ satisfies that $h_n > 0$, $\forall n$, $h_n \downarrow 0$ and $nh_n \uparrow \infty$ as $n \uparrow \infty$.

(A.3) If we denote by $\theta_j = \lim_{n\to\infty} 1/n \sum_{t=1}^{n} (h_t/h_n)^j$, then $\theta_j < \infty$, for $-1 \le j \le 4p$.

(A.4) The stationary processes $\{X_t, Y_t\}$ are strongly mixing (α-mixing) and such that $\sum_{t=1}^{\infty} t^{\epsilon}[\alpha(t)]^{1-2/\delta} < \infty$, for some $\delta > 2$ and $\epsilon > 1 - 2/\delta$.

(A.5) The joint probability density of X_t and X_{t+s}, $f_{X_tX_{t+s}}(x_t, x_{t+s})$, satisfies $|f_{X_tX_{t+s}}(x_t, x_{t+s}) - f_X(x_t)f_X(x_{t+s})| \le \text{cst.} < \infty$, for all x_t, x_{t+s} and $s \ge 1$.

(A.6) For all $s \ge 1$, it is verified that $f_{X_tX_{t+s}/Y_tY_{t+s}}(x_tx_{t+s}/y_ty_{t+s}) \le \text{cst.} < \infty$ and $f_{X_t/Y_t}(x_t/y_t) \le \text{cst.} < \infty$.

(A.7) $E(|Y_t|^\delta) < \infty$, for some $\delta > 2$.

Assumptions (A.1) and (A.2) are not very restrictive and they are classical when nonparametric regression is being considered. Although condition (A.3) appears to be more complex, it is also usual in the recursive estimation setting. If the usual selection of bandwidth $h_n = Cn^{-\nu}$ is considered, then θ_j may be calculated explicitly and is given by $\theta_j = 1/(1-\nu j)$. Thus, it would be enough to choose $\nu < 1/j$.

The condition (A.4) is a summability requirement on the α-mixing coefficients usually needed in strongly mixing dependence setting. Note that, among the numerous asymptotic independence conditions often imposed in this context, the α-mixing condition (introduced by Rosenblatt (1956)) is one of the least restrictive and it is satisfied by many processes. A wide and complete study about this condition can be seen in Doukhan (1995).

Finally, it is interesting to point out that (A.6) and (A.7) are technical assumptions which must be imposed in order to obtain the proofs.

A similar approach to that employed in Masry and Fan (1997) to obtain the asymptotic normality of the local polynomial kernel estimator of $\vec{\beta}$ will be followed. First, the asymptotic properties for the entries of matrices $S_{(n)}$ and $\vec{T}_{(n)}$ are studied in Theorems 1 and 2. Next, in Theorem 3, we establish additional conditions for concluding the asymptotic normality of vector $\vec{T}_{(n)}$. The proofs of these three previous results may be found in the Appendix. Finally, from results stated in the previous theorems, the asymptotic normality of the recursive estimator defined in (2.2) will be established.

THEOREM 1. *Under assumptions* (A.1), (A.2) *and* (A.3) *we have at every continuity point of* f_X

$$\lim_{n\to\infty} E(h_n^{-j} S_j^{(n)}) = \theta_j f_X(x)\mu_j, \quad \text{for} \quad j = 0, 1, \ldots, 2p, \tag{3.1}$$

where $\mu_j = \int u^j K(u)\, du$.

If, in addition, (A.4) *and* (A.5) *are assumed, then we have*

$$\lim_{n\to\infty} nh_n \mathrm{Var}(h_n^{-j} S_j^{(n)}) = \theta_{2j-1} f_X(x)\nu_{2j}, \quad \text{for} \quad j = 0, 1, \ldots, 2p, \tag{3.2}$$

The next conclusion follows directly from Theorem 1

$$h_n^{-j} S_j^{(n)} \xrightarrow{m.s.} \theta_j f_X(x)\mu_j, \tag{3.3}$$

which can be expressed in matrix form as

$$H_{(n)}^{-1} S_{(n)} H_{(n)}^{-1} \xrightarrow{m.s.} f_X(x) S, \tag{3.4}$$

where $H_{(n)} = \mathrm{diag}\left(1, h_n, h_n^2, \ldots, h_n^p\right)$ and S is the $(p+1)\times(p+1)$ matrix whose (r,t)-th element is $S_{r,t} = S_{r+t-2}$ with $S_j = \theta_j\mu_j$, $0 \le j \le 2p$. The convergence result in (3.4) is interpreted in the sense that each element of the matrix converges in mean square.

The next theorem establishes the asymptotic structure of the variance/covariance matrix of vector $\vec{T}_{(n)}$ centered with respect to vector $(m(X_1), \ldots, m(X_n))^t$. Let

$$\vec{T}^{\star}_{(n)} := \left(T^{\star}_{0,(n)}, \ldots, T^{\star}_{p,(n)}\right)^t,$$

where

$$T^{\star}_{j,(n)} = \frac{1}{n}\sum_{t=1}^{n} (X_t - x)^j\, K_t\,(X_t - x)\,(Y_t - m(X_t)), \quad \text{for} \quad 0 \le j \le p. \tag{3.5}$$

THEOREM 2. *Let us assume that conditions* (A.1) (A.7) *are satisfied. Then we have*

$$\lim_{n\to\infty} nh_n \mathrm{Cov}(h_n^{-j} T^{\star}_{j,(n)}, h_n^{-r} T^{\star}_{r,(n)}) - \theta_{j+r-1} f_X(x)\sigma_Y^2(x)\nu_{j+r}, \tag{3.6}$$

for $0 \leq j, r \leq p$ and for every x continuity point of σ_Y^2 and f_X.

Again as in (3.4), (3.6) may be expressed in matrix form as

$$nh_n E(H_{(n)}^{-1} \vec{T}_{(n)}^{\star} \vec{T}_{(n)}^{\star\, t} H_{(n)}^{-1}) \longrightarrow f_X(x) \sigma_Y^2(x) \tilde{S} \quad \text{as} \quad n \to \infty, \tag{3.7}$$

where $\tilde{S}$ is the $(p+1) \times (p+1)$ matrix whose (r,t)-th element is $\tilde{S}_{r,t} = \tilde{S}_{r+t-2}$ with $\tilde{S}_j = \theta_{j-1}\nu_j, 0 \leq j \leq 2p$.

Note that the assumptions (A.4)–(A.7) have been imposed to eliminate any effect due to the dependence of the data over first order terms in the expansion of the variance of the estimates, so that the influence of dependence is only noticeable over second order terms (these observations turn up along the proofs in Section 5). Thus, when the observations are independent, conditions (A.1)–(A.3) are sufficient to obtain the same results as in Theorem 1 and Theorem 2.

The following result establishes the asymptotic normality of $\vec{T}_{(n)}^{\star}$.

THEOREM 3. *Let us assume that conditions* (A.1)–(A.7) *are satisfied. In addition, the following assumptions hold for some $0 < \gamma < 1$:*

(A.8) *Let be $\pi_1 = 3(2p-1)$ for $p > 0$ and $\pi_1 = 3 - 2\gamma$ for $p = 0$. Then*

a) *h_n is such that $nh_n^{\pi_1} \to \infty$ as $n \to \infty$.*

b) *There exits a sequence of positive integers $\{s_n\}$, $s_n \to \infty$ as $n \to \infty$, with $s_n = o(\sqrt{nh_n^{\pi_1}})$ and such that*

$$\sqrt{nh_n^{\pi_2}} \sum_{t=s_n}^{\infty} [\alpha(t)]^{1-\gamma} \longrightarrow 0, \qquad as \quad n \to \infty,$$

with $\pi_2 = 1 - 2p$ for $p > 0$ and $\pi_2 = -1$ for $p = 0$.

(A.9) *The conditional distribution, $Y_t/X_t = x$, is continuous at the point x.*

Then, if x is a continuity point of σ_Y^2 and f_X, we have that

$$\sqrt{nh_n} H_{(n)}^{-1} \vec{T}_{(n)}^{\star} \xrightarrow{\mathcal{L}} N_{(p+1)}(\vec{0}, \sigma_Y^2(x) f_X(x) \tilde{S}). \tag{3.8}$$

Theorem 3 has been obtained by using Bernstein's method, which consists on approximating mixing sequences by independent ones. More specifically, the idea is to represent the sum of dependent variables as a sum of "almost" independent random variables alternating with other terms whose sum is asymptotically negligible. Assumption (A.8) is needed in order to develope this procedure. Indeed, s_n is the size of the groups of variables whose sums can be ignored. (A.9) is a technical assumption that is required in the proof of Theorem 3 to guarantee the continuity of the function

$$\eta(x) = \text{Var}\left(Y_t I\left\{|Y_t| > M\right\} / X_t = x\right)$$

for each $M > 0$ and for every continuity point, x, of σ_Y^2 and f_X.

From Theorems 1 and 3 (Expressions (3.4) and (3.8)) it can be straightly deduced that

$$\text{(3.9)} \qquad \sqrt{nh_n}H_{(n)}S_{(n)}^{-1}\vec{T}_{(n)}^{\star} \xrightarrow{\mathcal{L}} N_{(p+1)}\left(\vec{0}, \frac{\sigma_Y^2(x)}{f_X(x)}S^{-1}\tilde{S}S^{-1}\right).$$

With these results, we can now establish the joint asymptotic normality of $\hat{\vec{\beta}}_{(n)}$. Let $\vec{M}_{(n)} = (m(X_1), \ldots, m(X_n))^t$, then taking the conditioned expected value in (2.2) we obtain

$$\text{(3.10)} \qquad \vec{\beta}_{(n)}^{\star} = E[\hat{\vec{\beta}}_{(n)}/(X_1, \ldots, X_n)] = (X_{(n)}^t W_{(n)} X_{(n)})^{-1} X_{(n)}^t W_{(n)} \vec{M}_{(n)}.$$

On the other hand, we have assumed the continuity of the first $p+1$ derivatives of $m(\cdot)$ and as the polynomial fit is locally carried out in a neighbourhood of x, then $\vec{M}_{(n)}$ can be approximated by a Taylor series in a sufficiently small interval

$$\vec{M}_{(n)} = X_{(n)}\vec{\beta} + \frac{m^{(p+1)}(x)}{(p+1)!}\begin{pmatrix} (X_1 - x)^{p+1} \\ \vdots \\ (X_n - x)^{p+1} \end{pmatrix} + o_p\begin{pmatrix} h_1^{p+1} \\ \vdots \\ h_n^{p+1} \end{pmatrix},$$

which, after substituted in (3.10), leads to

$$\vec{\beta}_{(n)}^{\star} = \vec{\beta} + S_{(n)}^{-1}\left[\frac{m^{(p+1)}(x)}{(p+1)!}\begin{pmatrix} S_{p+1}^{(n)} \\ \vdots \\ S_{2p+1}^{(n)} \end{pmatrix} + o_p\left\{\frac{1}{n}\sum_{i=1}^{n} h_i^{p+1}\begin{pmatrix} h_n^{p+1} \\ \vdots \\ h_n^{2p+1} \end{pmatrix}\right\}\right],$$

and using the established convergence in (3.4) and Toeplitz's lemma we can deduce that

$$\text{(3.11)} \qquad \vec{\beta}_{(n)}^{\star} = \vec{\beta} + H_{(n)}^{-1}\left[\frac{m^{(p+1)}(x)}{(p+1)!}h_n^{p+1}S^{-1}\vec{\mu}_\theta + o_p\left(h_n^{p+1}(1, \ldots, 1)^t\right)\right],$$

where $\vec{\mu}_\theta = (\theta_{p+1}\mu_{p+1}, \ldots, \theta_{2p+1}\mu_{2p+1})^t$.

It must be noted that from Definitions (2.2) and (3.5) it follows

$$S_{(n)}^{-1}\vec{T}_{(n)}^{\star} = \hat{\vec{\beta}}_{(n)} - \vec{\beta}_{(n)}^{\star}$$

and using (3.11) we conclude that

$$\text{(3.12)} \quad \begin{aligned} H_{(n)}S_{(n)}^{-1}\vec{T}_{(n)}^{\star} &= H_{(n)}(\hat{\vec{\beta}}_{(n)} - \vec{\beta}) \\ &\quad - \left[\frac{m^{(p+1)}(x)}{(p+1)!}h_n^{p+1}S^{-1}\vec{\mu}_\theta + o_p\left(h_n^{p+1}(1, \ldots, 1)^t\right)\right]. \end{aligned}$$

From Expressions (3.9) and (3.12) we can deduce the following corollary that establishes the joint asymptotic normality of estimators $\hat{\beta}_j^{(n)} = \hat{m}^{(j)}(x)/(j!)$.

COROLLARY 1. *Under the hypotheses of Theorem 3, we have*

$$\sqrt{nh_n}\left[H_{(n)}\left(\hat{\vec{\beta}}_{(n)}-\vec{\beta}\right)-\frac{h_n^{p+1}m^{(p+1)}(x)}{(p+1)!}S^{-1}\vec{\mu}_0\right] \xrightarrow{\mathcal{L}} N_{(p+1)}\left(\vec{0},\frac{\sigma_Y^2(x)}{f_X(x)}S^{-1}\tilde{S}S^{-1}\right)$$

at the continuity points of σ_Y^2 *and* f_X *whenever* $f_X(x)>0$.

Corollary 1 is an analogous result, in a recursive setting, to those obtained by Masry and Fan (1997) for the local polynomial fitting under ρ-mixing and α-mixing conditions. Moreover, the first implication of Corollary 1 is that both the recursive estimate and non recursive one exhibit the same rate of convergence for their mean square errors (although with different values for the asymptotic expressions of both the bias and the variance).

The asymptotic normality of the individual components, $\hat{\beta}_k^{(n)}=m^{(k)}(x)/k!$, for $0\le k\le p$, follows directly from Corollary 1. The local linear fitting, corresponding $p=1$, is particularly interesting and simple. In this case, if the selected kernel function satisfies that $\mu_0=1$ and $\mu_1=0$, then Corollary 1 leads to

$$\sqrt{nh_n}\left[\left(\hat{\beta}_0^{(n)}-\beta_0\right)-\frac{m''(x)}{2}h_n^2\theta_2\mu_2\right]\xrightarrow{\mathcal{L}} N\left(0,\frac{\sigma_Y^2(x)}{f_X(x)}\theta_{-1}\nu_0\right), \tag{3.13}$$

$$\sqrt{nh_n^3}\left[\left(\hat{\beta}_1^{(n)}-\beta_1\right)-\frac{m''(x)}{2}h_n\frac{\theta_3\mu_3}{\theta_2\mu_2}\right]\xrightarrow{\mathcal{L}} N\left(0,\frac{\sigma_Y^2(x)\theta_1\nu_2}{f_X(x)\theta_2^2\mu_2^2}\right). \tag{3.14}$$

Statement (3.13) allows us to present a comparative study of the asymptotic behaviour of the following nonparametric regression estimates: the estimator obtained by lineal local fitting (LL), the recursive version of the Nadaraya-Watson estimator (RNW) given in (1.3), and, finally, the recursive estimator by linear local fitting (RLL) proposed in this paper. The bias and variance of the above three estimators are listed in Table 1.

Table 1. Pointwise bias and variance of the local linear fitting (LL), the recursive Nadaraya-Watson estimator (RNW) and the recursive local linear fitting (RLL).

Estimator	Bias	Variance
LL	$\frac{1}{2}m''(x)h_n^2\mu_2$	$\frac{1}{nh_n}\frac{\sigma_Y^2(x)}{f_X(x)}\nu_0$
RNW	$\left(\frac{1}{2}m''(x)+m'(x)\frac{f_X'(x)}{f_X(x)}\right)h_n^2\mu_2\theta_2$	$\frac{1}{nh_n}\frac{\sigma_Y^2(x)}{f_X(x)}\nu_0\theta_{-1}$
RLL	$\frac{1}{2}m''(x)h_n^2\mu_2\theta_2$	$\frac{1}{nh_n}\frac{\sigma_Y^2(x)}{f_X(x)}\nu_0\theta_{-1}$

In the first place, from Table 1 we see that the rates of convergence of both bias and variance are h_n^2 and $1/nh_n$ respectively, the same for the three estimators.

As it is well-known, the local linear fitting has been shown to reduce the bias of the Nadaraya-Watson estimator, while the variance remains the same. Table 1 shows how these properties are conserved when the recursive versions are used.

Next, for the correct interpretation Table 1, the usual selection of bandwidth $h_n = Cn^{-\nu}$ will be considered. In such case, $\theta_2 = 1/1 - 2\nu > 1$ and $\theta_{-1} = 1/1 + \nu < 1$. Therefore, if the recursive (RLL) and nonrecursive (LL) estimators by local linear fitting are compared, then it can be observed that the RLL has a larger bias and smaller variance than the LL. This result is expected because the recursive estimator uses a bigger bandwidth (on average, or any other sense) than the nonrecursive estimator.

By contrast, since the sum of variance and square bias is a complicated expression involving the unknown functions f_X, m' and m'' and depending of the kernel and bandwidth selected, it is not possible to establish a general comparison among the mean square error of the three analyzed estimators. Nevertheless, the extensive simulation study performed in next section allows to deduce, in a heuristic sense, that the RLL estimator presents a little higher mean square error than the LL estimator when both of them use the optimal bandwidth. This drawback is not very serious and, in any case, the mentioned simulation study also shows that the MSE of the RLL is lower than the ones of Nadaraya-Watson estimator and his recursive version given by (1.3).

We can conclude that the estimator studied in this work and defined in (2.2) presents similar properties to those of the estimator of the regression function and its derivatives obtained through local polynomial fitting. Moreover, it has the additional advantage that it is recursive, as proved in Section 2.

4. Simulation study

This section contains a numerical study of the proposed nonparametric estimation method comparing it to other classical methods. Three regression models of the form $Y_t = m_j(X_t) + \varepsilon_t$ with $t = 1, \ldots, n$ and $j = 1, 2, 3$, being $m_j(x)$ the regression function and ε_t the error of the model, are considered.

Model 1. The values of X_t come from a uniform distribution in $[0, 1]$, the regression function is $m_1(x) = \sin(5\pi x)$ and the error, ε_t, follows an AR(1) structure, $\varepsilon_t = \rho\varepsilon_{t-1} + e_t$, being $\{e_t\}$ a sequence of independent random variables with a common normal distribution $N(0, 0.3)$.

Model 2. We choose $m_2(x) = 16x^2(1 - x)^2$ and ε_t are independent and with common distribution $N(0, 0.5)$. Here, the values of X_t come from a variable with density $f(x) = 6x(1 - x)I_{[0,1]}(x)$.

In the previous two models the regression functions are studied in the interval $[0, 1]$, making the difference between the boundary region, $[0, 0.3] \cup [0.7, 1]$, and the central region, $[0.3, 0.7]$.

Model 3. The regression function is $m_3(x) = \sin(2x) + 2\exp(-16x^2)$, ε_t are independent and with common distribution $N(0, 0.8)$. The variable X_t comes from a $N(0, 1)$ and the study is performed in the interval $[-2, 2]$, being the boundary region $[-2, -1] \cup [1, 2]$.

In all models, the regression function is estimated in $N = 200$ equally spaced points in the interval under study and four nonparametric estimators are used: the Nadaraya-Watson estimator (NW), the recursive NW given in (1.3) (RNW), the local linear fitting (LL) and the recursive local linear fitting (RLL). The used kernel function is the quartic kernel ($K(u) = (15/16)(1 - u^2)^2$ if $|u| \leq 1$) and, for the recursive estimators, we have used bandwidths of the form $h_t = Ct^{-1/5}$, being C a constant to be empirically determined.

In the first step of the simulation, we took 500 random samples of size $n = 200$ and $n = 500$, for each one of the three regression models. Next, for every sample, we computed the bandwidth (h_{opt}) by minimizing the average squared error (ASE),

$$ASE(\hat{m}_h) = \sum_{j=1}^{N} (\hat{m}_h(x_j) - m(x_j))^2 \omega(x_j),$$

where $\omega(x)$ is a weight function that varies depending on whether the study is performed in the center region, boundary region or global region. With this bandwidth the ASE is computed for each sample as the sum of the squared bias and the variance. The obtained results are averaged for the 500 samples. In Tables 2, 3 and 4 we present the results for Model 1 with independent errors ($n = 200$ and $\rho = 0$), Model 1 with dependent errors ($n = 200$ and $\rho = 0.6$) and Model 3 ($n = 500$), respectively. For Model 2 similar results were achieved.

Table 2. Squared bias, variance and average square error of the Model 1 ($n = 200$), with independent data ($\rho = 0$), in boundary region, central region and global region.

	Estimator			
	NW	RNW	LL	RLL
Boundary R.				
$\bar{h}_{opt}$	0.045	0.104	0.059	0.135
Squared Bias	0.002 903	0.002 855	0.001 980	0.001 956
Variance	0.009 594	0.009 773	0.007 489	0.007 622
ASE	0.012 497	0.012 628	0.009 469	0.009 578
Central R.				
$\bar{h}_{opt}$	0.055	0.129	0.058	0.134
Squared Bias	0.001 334	0.001 465	0.001 567	0.001 642
Variance	0.007 463	0.007 656	0.005 675	0.005 836
ASE	0.008 797	0.009 121	0.007 242	0.007 478
Global R.				
$\bar{h}_{opt}$	0.047	0.110	0.058	0.135
Squared Bias	0.002 288	0.002 273	0.001 738	0.001 814
Variance	0.009 292	0.009 483	0.007 102	0.007 191
ASE	0.011 580	0.011 756	0.008 839	0.009 005

Table 3. Squared bias, variance and average square error of the Model 1 ($n = 200$), with dependent data ($\rho = 0.6$), in boundary region, central region and global region.

	Estimator			
	NW	RNW	LL	RLL
Boundary R.				
h_{opt}	0.044	0.092	0.060	0.132
Squared Bias	0.002 772	0.003 854	0.001 816	0.002 030
Variance	0.011 186	0.012 249	0.008 983	0.009 528
ASE	0.013 958	0.016 104	0.010 799	0.011 559
Central R.				
$\bar{h}_{opt}$	0.054	0.135	0.057	0.143
Squared Bias	0.001 366	0.001 362	0.001 645	0.001 648
Variance	0.008 704	0.008 736	0.006 773	0.006 821
ASE	0.010 071	0.010 099	0.008 419	0.008 471
Global R.				
$\bar{h}_{opt}$	0.047	0.103	0.060	0.137
Squared Bias	0.002 271	0.003 141	0.001 745	0.001 909
Variance	0.010 616	0.011 373	0.008 384	0.008 739
ASE	0.012 887	0.014 514	0.010 129	0.010 647

Table 4. Squared bias, variance and average square error of the Model 3 ($n = 500$) in boundary region, central region and global region.

	Estimator			
	NW	RNW	LL	RLL
Boundary R.				
$\bar{h}_{opt}$	0.168	0.189	0.103	0.292
Squared Bias	0.001 351	0.001 465	0.000 791	0.000 854
Variance	0.004 510	0.004 699	0.002 842	0.002 886
ASE	0.005 861	0.006 164	0.003 633	0.003 740
Central R.				
$\bar{h}_{opt}$	0.034	0.095	0.035	0.096
Squared Bias	0.000 750	0.000 802	0.000 815	0.000 871
Variance	0.003 181	0.003 298	0.002 783	0.002 900
ASE	0.003 931	0.004 100	0.003 598	0.003 771
Global R.				
$\bar{h}_{opt}$	0.044	0.123	0.047	0.131
Squared Bias	0.001 204	0.001 311	0.001 333	0.001 439
Variance	0.004 925	0.005 088	0.004 385	0.004 530
ASE	0.006 129	0.006 399	0.005 718	0.005 969

In these three tables we can observe the good behaviour of the recursive estimator by local polynomial fitting (RLL). It presents a little lower (but close to) efficiency than its corresponding nonrecursive estimator (LL). Moreover, it conserves its properties in the sense that it improves the obtained results when using the Nadaraya-Watson kernel estimator, both recursive and nonrecursive (RNW and NW), specially in the boundary region.

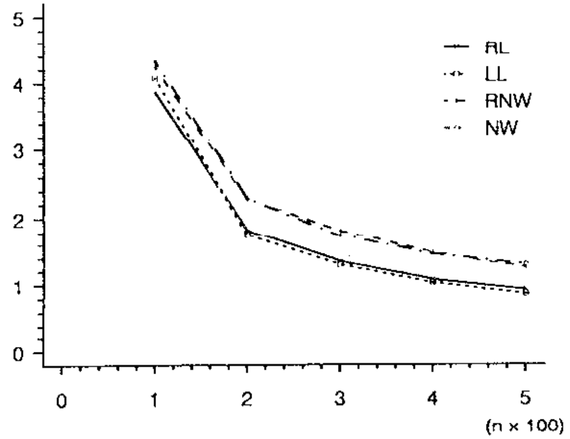

Fig. 1. Squared bias of Model 1, with independent data ($\rho = 0$), in the global region of the four analyzed estimators.

In Section 3, we have shown that the behaviour of the proposed recursive estimator under α-mixing dependence conditions are asymptotically equal to the case of independent observations. Nevertheless, for finite samples, it becomes evident that every kind of dependence must affect the efficiency of the estimators. Due to this fact, Table 2 (independent observations) shows better results than Table 3 ($AR(1)$ structure for the error) for all estimators. The squared errors of the estimators would be increased if a higher autocorrelation degree was employed.

In the second step of the simulation, we have studied the influence of the sample size. In order to do this, we have carried out the study for $n = 100$, 200, 300, 400 and 500. In Figs. 1 and 2 we present the results for Model 1, with $\rho = 0$, in the global region. In Fig. 1 we have represented the graph of the squared bias as a function of n for the four studied estimators. In Fig. 2 we have represented the quotient between the mean squared error of the NW, RNW and LL estimators and the estimator we propose (RLL). Thus, Fig. 2 presents a measure of the relative efficiency of this estimator with respect to the previous ones. Similar graphs to those of Figs. 1 and 2 were obtained when the studies were performed in both the boundary and the central region or when the observations were dependent.

From Fig. 1 it can be observed that the estimators by polynomial fitting (recursive and non recursive) present smaller bias than the Nadaraya-Watson kernel estimators (both the recursive and the nonrecursive one). In fact, our simulation study allowed us to observe that this property becomes more noticeable in the

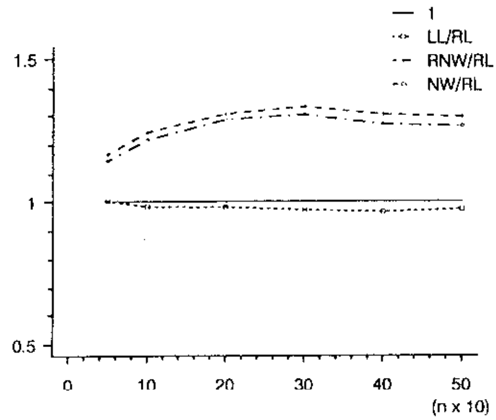

Fig. 2. Relative efficiency, as a function of n, of estimators LL, NW and RNW with respect to the RLL estimator of Model 1 with ($\rho = 0$) in the global region.

boundary region. Figure 2 shows clearly that the estimators obtained by polynomial fitting (recursive and non recursive) have smaller squared error. That is, they are more efficient than the classical Nadaraya-Watson estimators. Note that this advantage is present for the different employed sample sizes. On the other hand, it is interesting to observe that the differences between the results provided by the non recursive and recursive estimators are very small, both in the case of the estimation by local polynomial fitting and with the Nadaraya-Watson kernel estimator. We find a slight improvement in the non recursive estimator, which is the price that has to be paid for the advantage of the recursivity of the estimator. The same conclusions can be deduced from the study of the other simulated models.

As we have pointed out in the introduction, other relevant property of the recursive estimator is his computational efficiency when the observations are received sequentially. In fact, from (2.9) it is clear that if a previous estimation $\vec{\tilde{\beta}}_n$ has been calculated from n data, then the updating of the recursive estimate with each additional observation is independent on the initial sample size n. Therefore a better behaviour in terms of computing expedience of recursive estimate over non recursive one may be expected.

In order to evaluate the computing time we simulated random samples of size n (with $n = 100$, 200, 500 and 1000) from Model 1 and we computed $\vec{\tilde{\beta}}_n$ by using both the LL estimator and the RLL estimator. To make fair comparisons, the estimations were made for one fixed point ($x_0 = 0.5$). First, we computed the LL estimation by direct calculation. Secondly, from an initial estimation based on $n-1$ data (which is again obtained by using the LL estimator), we computed the RLL estimation based on n data by employing the updating algorithm given in (2.9). Finally, we compared the computing times concluding that, for a given sample size n, the quotient between the computing time for LL estimator and the one employed to update the estimation by the proposed recursive estimator was $n/2$. Therefore, as we expected, we can conclude that the recursive estimate is far faster than the non recursive one when the observations are sequentally obtained. Although, this computational advantage increases with sample size, it becomes important even for moderately large sample sizes.

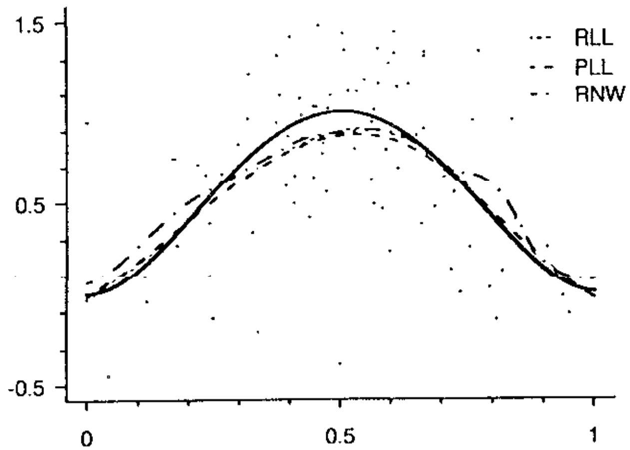

Fig. 3. Estimates based on the simulation of Model 2 for $n = 100$.

Finally, in order to observe the global behaviour of these estimators we have calculated, from simulated samples, the estimation of the regression function using the RLL, LL and RNW estimators for all models. The results for Model 2 can be seen in Fig. 3. The used bandwidths are global and have been empirically obtained. The use of variable bandwidths, such as those proposed in Fan and Gijbels (1995), provide better results, specially in the local polynomial fitting estimators. In this case, there is a large difference between the band that must be used in the boundary region and the one used in the central region.

5. Appendix: proofs

In this appendix we detail the general lines of the proofs for the theorems in Section 3. In what follows, the letter C will be used to indicate generic constants whose values are not important and may vary.

Proof of Theorem 1. Let

$$V_{t,j} = \left(\frac{X_t - x}{h_t}\right)^j K_t(X_t - x).$$

Assumptions (A.1) and (A.2) allow us to apply Bochner's lemma (Wheeden and Zygmund (1977), Theorem 9.9) to obtain

$$\lim_{t\to\infty} E\left(V_{t,j}\right) = f_X(x)\mu_j, \tag{5.1}$$

$$\lim_{t\to\infty} h_t E\left(V_{t,j}^2\right) = f_X(x)\nu_{2j}. \tag{5.2}$$

Note that, from (2.3), we have

$$\lim_{n\to\infty} E(h_n^{-j} S_j^{(n)}) - \frac{1}{n}\sum_{t=1}^{n}\left(\frac{h_t}{h_n}\right)^j E(V_{t,j}).$$

Now, the limit (5.1) and the Assumption (A.3) allow us to apply Toeplitz's lemma to obtain (3.1).

For the second moment we proceed as follows. First, we write

$$\operatorname{Var}(h_n^{-j} S_j^{(n)}) = h_n^{-2j} P_{1,j}^{(n)} + h_n^{-2j} P_{2,j}^{(n)}, \tag{5.3}$$

where

$$P_{1,j}^{(n)} = n^{-2}\sum_{t=1}^{n} h_t^{2j}\operatorname{Var}(V_{t,j}), \tag{5.4}$$

$$P_{2,j}^{(n)} = 2n^{-2}\sum_{\substack{r=1\\ r<s}}^{n}\sum_{s=1}^{n} (h_r h_s)^j \operatorname{Cov}(V_{r,j}, V_{s,j}). \tag{5.5}$$

Using (5.1), (5.2) and similar arguments to those employed to prove (3.1), we obtain

$$h_n^{-2j} P_{1,j}^{(n)} = \frac{1}{nh_n}\theta_{2j-1} f_X(x)\nu_{2j} + o\left(\frac{1}{nh_n}\right). \tag{5.6}$$

If we show that $h_n^{-2j} P_{2,j}^{(n)} = o(1/nh_n)$, from (5.3) and (5.6) it is clear that the proof is completed, namely (3.2) is established. For this, let c_n be a sequence of integer numbers such that $c_n \uparrow \infty$ and $h_n c_n \downarrow 0$ as $n \uparrow \infty$ and let us decompose $P_{2,j}^{(n)}$ in the following way

$$P_{2,j}^{(n)} = P_{21,j}^{(n)} + P_{22,j}^{(n)}, \tag{5.7}$$

where $P_{21,j}^{(n)}$ is the sum in (5.5) with r and s such that $0 < s - r \le c_n$ and where $P_{22,j}^{(n)}$ is the sum in (5.5) with r and s such that $s - r > c_n$.

From Assumption (A.5) and the stationary property of the process X, it is easy derived that $f_X^2(x)\mu_j^2 C$ is an asymptotic bound for $\operatorname{Cov}(V_{r,j}, V_{s,j})$. Hence

$$P_{21,j}^{(n)} \le Cn^{-2} \sum_{\substack{r=1\\ 0<s-r\le c_n}}^{n}\sum_{s=1}^{n} (h_r h_s)^j \le Cn^{-2} c_n \sum_{r=1}^{n} h_r^{2j}.$$

Now, applying Toeplitz's lemma again, we conclude that

$$h_n^{-2j} P_{21,j}^{(n)} = O\left(\frac{c_n}{n}\right) = o(1). \tag{5.8}$$

In order to show that $P_{22,j}^{(n)}$ is asymptotically negligible, it is neccesary to find a better bound for $\mathrm{Cov}(V_{r,j}, V_{s,j})$. For this, we will make use of the Davydov's inequality. Note that, for some $\delta > 2$, we have $E|V_{r,j}|^\delta \le h_r^{1-\delta} D_{r,j}$, being

$$D_{r,j} = \int \left|\frac{u-x}{h_r}\right|^{\delta j} \frac{1}{h_r} \left|K\left(\frac{u-x}{h_r}\right)\right|^\delta f_X(u)\, du.$$

Once again, Assumption (A.1) guarantees the applicability of Bochner's lemma to conclude that

$$D_{r,j} \longrightarrow f_X(x) \int_{-\infty}^{\infty} |v|^{\delta j} \left|K(v)\right|^\delta dv < \infty. \tag{5.9}$$

Thus, by the Davydov's inequality, we obtain

$$|\mathrm{Cov}(V_{r,j}, V_{s,j})| \le 8[\alpha(s-r)]^{1-2/\delta}(h_r h_s)^{(1-\delta)/\delta}(D_{r,j}D_{s,j})^{1/\delta}.$$

Hence

$$\begin{aligned} |P_{22,j}^{(n)}| &\le 16n^{-2} \sum_{\substack{r=1 \\ s-r>c_n}}^{n} \sum_{s=1}^{n} (h_r h_s)^{j+(1-\delta)/\delta} [\alpha(s-r)]^{1-2/\delta} (D_{r,j}D_{s,j})^{1/\delta} \\ &= 16n^{-2} \sum_{k=c_n+1}^{n-1} [\alpha(k)]^{1-2/\delta} \left[\sum_{t=1}^{n-k} (h_t^{j-1+1/\delta} D_{t,j}^{1/\delta})(h_{t+k}^{j-1+1/\delta} D_{t+k,j}^{1/\delta})\right]. \end{aligned}$$

Using the Cauchy-Schwarz inequality and since $D_{r,j} \ge 0$, the inner sum in the above expression can be bounded by $\sum_{t=1}^{n} h_t^{2(j-1+1/\delta)} D_{t,j}^{2/\delta}$. Hence

$$|P_{22,j}^{(n)}| \le 16n^{-2} \sum_{k=c_n+1}^{n-1} [\alpha(k)]^{1-2/\delta} \left[\sum_{t=1}^{n} h_t^{2(j-1+1/\delta)} D_{t,j}^{2/\delta}\right].$$

Finally, from the Assumptions (A.3) and (A.4), from the limit (5.9) and from the Toeplitz's lemma, we have

$$nh_n(h_n^{-2j} P_{22,j}^{(n)}) \longrightarrow 0, \qquad \text{as} \quad n \to \infty,$$

which jointly with (5.8) and (5.7) allow us to conclude that $h_n^{-2j} P_{2,j}^{(n)} = o(1/nh_n)$ and then the proof of (3.2) is stated. □

Proof of Theorem 2. Similar arguments to those used in proving Theorem 1 are employed again. For this reason we will only provide the main differences below.

First, it must be observed that conditioning $T_{j,(n)}^\star$ on $\vec{X}_{(n)} = (X_1, \ldots, X_n)^t$ we find that $E(T_{j,(n)}^\star / \vec{X}_{(n)}) = 0$. Hence

$$\mathrm{Cov}(h_n^{-j} T_{j,(n)}^\star, h_n^{-i} T_{i,(n)}^\star) = h_n^{-(j+i)} E(T_{j,(n)}^\star T_{i,(n)}^\star).$$

Now, in the same way as in the proof of Theorem 1, we denote

$$U_{t,j} = \left(\frac{X_t - x}{h_t}\right)^j K_t(X_t - x)(Y_t - m(X_t)) \tag{5.10}$$

and we write

$$h_n^{-(j+i)} E(T^{\star}_{j,(n)} T^{\star}_{i,(n)}) = h_n^{-(j+i)} (Q^{(n)}_{1,ji} + Q^{(n)}_{2,ji}), \tag{5.11}$$

where

$$Q^{(n)}_{1,ji} = n^{-2} \sum_{t=1}^{n} h_t^{j+i} E(U_{t,j} U_{t,i})$$

$$Q^{(n)}_{2,ji} = 2n^{-2} \sum_{\substack{r=1 \\ r<s}}^{n} \sum_{s=1}^{n} h_r^j h_s^i E(U_{r,j} U_{s,i}).$$

By conditioning on $\vec{X}_{(n)}$ and applying Bochner's lemma, we have

$$\begin{aligned} h_t E(U_{t,j} U_{t,i}) &= \int \left(\frac{u - x}{h_t}\right)^{j+i} K^2\left(\frac{u - x}{h_t}\right) \sigma_Y^2(u) f_X(u) \frac{du}{h_t} \\ &= \int z^{j+i} K^2(z) (\sigma_Y^2 f_X)(x + z h_t)\, dz \stackrel{t\to\infty}{\longrightarrow} \sigma_Y^2(x) f_X(x) \nu_{j+i}, \end{aligned} \tag{5.12}$$

at continuity points of $\sigma_Y^2 f_X$. From (5.12), Assumption (A.3) and Toeplitz's lemma we conclude that

$$\lim_{n\to\infty} n h_n h_n^{-(j+i)} Q^{(n)}_{1,ji} = \theta_{j+i-1} \sigma_Y^2(x) f_X(x) \nu_{j+i}. \tag{5.13}$$

Since (5.11) and (5.13), it is suffices to show that $h_n^{-(j+i)} Q^{(n)}_{2,ji} = o(1/nh_n)$ for completing the proof of (3.6).

Reasoning as in (5.7) we can descompose $Q^{(n)}_{2,ji}$ in the same way as $P^{(n)}_{2,j}$. Thus

$$\begin{aligned} Q^{(n)}_{2,ji} &= 2n^{-2} \left(\sum_{\substack{r=1 \\ 0<s-r\le c_n}}^{n} \sum_{s=1}^{n} h_r^j h_s^i E(U_{r,j} U_{s,i}) + \sum_{\substack{r=1 \\ s-r>c_n}}^{n} \sum_{s=1}^{n} h_r^j h_s^i E(U_{r,j} U_{s,i}) \right) \\ &= Q^{(n)}_{21,ji} + Q^{(n)}_{22,ji}. \end{aligned} \tag{5.14}$$

By conditioning on (Y_r, Y_s), we have

$$\begin{aligned} & E\left(U_{r,j} U_{s,i}\right) \\ &= \iint \left[\iint \left(v_r - m(u_r)\right)\left(v_s - m(u_s)\right) f_{Y_r Y_s}(v_r v_s)\, dv_r dv_s \right] \\ &\quad \cdot \left(\frac{u_r - x}{h_r}\right)^j K_r\left(u_r - x\right) \left(\frac{u_s - x}{h_s}\right)^i \\ &\quad \cdot K_s\left(u_s - x\right) f_{X_r X_s / Y_r Y_s}\left(u_r u_s / v_r v_s\right)\, du_r du_s \end{aligned} \tag{5.15}$$

Since K has compact support and m is continuous, there exists a constant C such that $C = \sup_{|u-x|\le h_r} |m(u)|$. On the other hand, from (A.7) and the Davydov's inequality, it follows that $E(|Y_r||Y_s|) \le C < \infty$. These considerations and assumption (A.6) allow us conclude in (5.15) that

$$|E(U_{r,j}U_{s,i})| \le C\mu_j\mu_i E\left[(|Y_r|+C)(|Y_s|+C)\right] \le C\mu_j\mu_i < \infty, \qquad \text{for} \quad r \ne s.$$

By using this bound and identical arguments to those employed for proving (5.8), we can deduce that

$$nh_n h_n^{-(j+i)} Q_{21,ji}^{(n)} = O(c_n h_n) = o(1). \tag{5.16}$$

Now, it only remains to prove that

$$h_n^{-(j+i)} Q_{22,ji}^{(n)} = o\left(\frac{1}{nh_n}\right). \tag{5.17}$$

By conditioning on Y_r and by using Assumptions (A.1), (A.6) and (A.7) again, we obtain, for $\delta > 2$, that

$$\begin{aligned} E|U_{r,j}|^\delta = & \iint \left|\frac{u_r - x}{h_r}\right|^{\delta j} K_r^\delta(u_r - x) \\ & \cdot |v_r - m(u_r)|^\delta f_{X_r/Y_r}(u_r/v_r)\, f_{Y_r}(v_r)\, du_r du_s \\ \le & \; CE(|Y_r|+C)^\delta h_r^{1-\delta} \int |z|^{\delta j} K^\delta(z)\,dz \le Ch_r^{1-\delta} \end{aligned} \tag{5.18}$$

From (5.18) and Davydov's lemma, we obtain

$$|E(U_{r,j}U_{s,j})| \le C\left[\alpha(s-r)\right]^{1-2/\delta} (h_r h_s)^{(1-\delta)/\delta}.$$

Finally, replacing this bound in the second sum in (5.14), and proceeding as in the last part of the proof of Theorem 1, we deduce (5.17) and therefore the validity of Theorem 2 is established. □

PROOF OF THEOREM 3. Let Q_n be an arbitrary linear combination of $h_n^{-j} T^\star_{j,(n)}$,

$$Q_n = \sum_{j=0}^{p} a_j h_n^{-j} T^\star_{j,(n)}.$$

By (3.5), $\sqrt{nh_n}\, Q_n$ can be also written in the form

$$\sqrt{nh_n}\, Q_n = \frac{1}{\sqrt{n}} \Sigma_{(n)} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \xi_{t,(n)}, \tag{5.19}$$

where

$$\xi_{t,(n)} = \sqrt{h_n} G_{t,(n)}(X_t - x)(Y_t - m(X_t)),$$

with

$$G_{t,(n)}(u) = \sum_{j=0}^{p} a_j \left(\frac{u}{h_n}\right)^j K_t(u).$$

If the asymptotic normality of $\sqrt{nh_n}\,Q_n$ is established, then (3.8) is followed from Cramer-Wold theorem. In order to do this, we have used Bernstein's method also known by "the big small blocks technique". In what follows, we limit ourselves to develope this procedure assuming that $p > 0$. A similar analysis is required for $p = 0$.

Step 1. Decomposition of $\Sigma_{(n)}$ into blocks.

By Assumption (A.8) there exists a sequence, $\{r_n\}$, $r_n \uparrow \infty$, such that

$$r_n s_n = o\left(\sqrt{nh_n^{3(2p+1)}}\right), \tag{5.20}$$

$$r_n (nh_n^{1-2p})^{1/2} \sum_{t=s_n}^{\infty} [\alpha(t)]^{1-2/\delta} \longrightarrow 0, \quad \text{with} \quad \delta > 2. \tag{5.21}$$

Let us define a new sequence $\{b_n\}$, in terms of $\{r_n\}$, by

$$b_n = \left[\frac{\sqrt{nh_n^{2p-1}}}{r_n}\right], \tag{5.22}$$

where $[a]$ denotes the integer part of a.

Next, for each n, let $k_n = [n/(b_n + s_n)]$ and let us split $\Sigma_{(n)}$ into $2k_n + 1$ terms in the following way,

$$\Sigma_{(n)} = \sum_{j=0}^{k_n-1} \Phi_j + \sum_{j=0}^{k_n-1} \Upsilon_j + \Lambda_{k_n} \equiv \Sigma_{b,(n)} + \Sigma_{s,(n)} + \Sigma_{r,(n)}, \tag{5.23}$$

being

$$\Phi_j = \sum_{i=1}^{b_n} \xi_{e_j+i,(n)}, \qquad \Upsilon_j = \sum_{i=b_n+1}^{b_n+s_n} \xi_{e_j+i,(n)}, \qquad \Lambda_{k_n} = \sum_{i=e_{k_n}+1}^{n} \xi_{i,(n)} \tag{5.24}$$

with $e_j = j(b_n + s_n)$, for $j = 0, \ldots, k_n - 1$. Thus, each Φ_j represents a large block summing b_n variables, each Υ_j is a small block summing s_n variables and, finally, Λ_{k_n} is a residual block.

Step 2. Both the sum of the smaller blocks, $\Sigma_{s,(n)}$, and residual block, $\Sigma_{r,(n)}$, are shown to be asymptotically negligible. That is, we need to prove that, as $n \to \infty$,

$$E\left(\frac{1}{\sqrt{n}}\Sigma_{s,(n)}\right)^2 \longrightarrow 0 \quad \text{and} \quad E\left(\frac{1}{\sqrt{n}}\Sigma_{r,(n)}\right)^2 \longrightarrow 0.$$

Since $E(\xi_{t,(n)}) = 0$, we conclude that $E(\Phi_j) = E(\Upsilon_j) = E(\Lambda_{k_n}) = 0$. Thus, from (5.23) and (5.24) we have

$$E(|\Sigma_{s,(n)}|^2) \le \sum_{j=0}^{k_n-1} E(|\Upsilon_j|^2) + 2\sum_{\substack{i=0\\ i>j}}^{k_n-1}\sum_{j=0}^{k_n-1} |E(\Upsilon_i\Upsilon_j)| \equiv M + N. \tag{5.25}$$

First, we pay attention to M. By using Davydov inequality, we have, for $\delta > 2$,

$$\begin{aligned} E|\Upsilon_j|^2 &< \sum_{t=b_n+1}^{b_n+s_n} E|\xi_{t,(n)}|^2 \\ &\quad + 16\sum_{\substack{f=b_n+1\\ f>g}}^{b_n+s_n}\sum_{g=b_n+1}^{b_n+s_n} [\alpha(f-g)]^{1-2/\delta}\left\|\xi_{e_j+f,(n)}\right\|_\delta \left\|\xi_{e_j+g,(n)}\right\|_\delta \end{aligned} \tag{5.26}$$

where $\|X\|_\delta$ denotes $(E|X|^\delta)^{1/\delta}$.

Let $U_{t,j}$ be defined by (5.10). Using Minkowsky inequality and (5.18) we can conclude

$$\begin{aligned} \|\xi_{t,(n)}\|_\delta &\le h_n^{1/2}\sum_{j=0}^{p} a_j\left(\frac{h_t}{h_n}\right)^j \|U_{t,j}\|_\delta \\ &\le C\frac{h_n^{1/2}}{h_t^{1-1/\delta}}\sum_{j=0}^{p} a_j\left(\frac{h_t}{h_n}\right)^j \le Ch_n^{1/2-p}. \end{aligned} \tag{5.27}$$

On the other hand, from (5.12) we have

$$E(|\xi_{t,(n)}|^2) \le \sum_{i=0}^{p}\sum_{j=0}^{p} a_i a_j\left(\frac{h_t}{h_n}\right)^{i+j-1} h_t\left|E\left(U_{t,i}U_{t,j}\right)\right| \le Ch_n^{1-2p}. \tag{5.28}$$

Now, since $\sum_{t=1}^{\infty}[\alpha(t)]^{1-2/\delta} < \infty$ by Assumption (A.8), we replace (5.28) and (5.27) in (5.26) and we obtain

$$\begin{aligned} M &\le C\sum_{j=0}^{k_n-1}\sum_{t=b_n+1}^{b_n+s_n} h_n^{1-2p} + C\sum_{j=0}^{k_n-1}\sum_{\substack{f=1\\ f>g}}^{s_n}\sum_{g=1}^{s_n}[\alpha(f-g)]^{1-2/\delta}h_n^{1-2p} \\ &\le C\frac{k_n s_n}{h_n^{2p-1}} + C\frac{k_n s_n}{h_n^{2p-1}}\sum_{t=1}^{s_n-1}[\alpha(t)]^{1-2/\delta} \le C\frac{k_n s_n}{h_n^{2p-1}}. \end{aligned} \tag{5.29}$$

Secondly, in order to establish bounds for N, we note that

$$N \le 2\sum_{\substack{i=0\\ i>j}}^{k_n-1}\sum_{j=0}^{k_n-1}\sum_{f=b_n+1}^{b_n+s_n}\sum_{g=b_n+1}^{b_n+s_n}\left|E(\xi_{e_i+f,(n)}\xi_{e_j+g,(n)})\right|. \tag{5.30}$$

Since $i > j$, the indices $e_i + f$ and $e_j + g$ must differ in, at least, b_n units and, therefore, if the Davydov inequality is applied, the above expression is bounded as follows

$$(5.31)\quad N \leq 4 \sum_{f=1}^{n-b_n} \sum_{g=f+b_n}^{n} |E(\xi_{f,(n)}\xi_{g,(n)})| \leq 32h_n^{1-2p} \sum_{f=1}^{n-b_n} \sum_{g=f+b_n}^{n} [\alpha(g-f)]^{1-2/\delta} \leq C\frac{n}{h_n^{2p-1}} \sum_{t=b_n}^{\infty} [\alpha(t)]^{1-2/\delta}.$$

From (5.22) and (5.20), we deduce that, as $n \to \infty$,

$$(5.32)\quad \frac{s_n}{b_n h_n^{2p-1}} \approx \frac{s_n r_n}{(nh_n^{3(2p-1)})^{1/2}} \longrightarrow 0 \Longrightarrow \frac{s_n}{b_n} \longrightarrow 0.$$

Hence, we can replace $\sum_{t=b_n}^{\infty}[\alpha(t)]^{1-2/\delta}$ by $\sum_{t=s_n}^{\infty}[\alpha(t)]^{1-2/\delta}$ in (5.31). From this and (5.29), we conclude in (5.25) that

$$(5.33)\quad \frac{1}{n}E(|\Sigma_{s,(n)}|^2) \leq C\left(\frac{k_n s_n}{nh_n^{2p-1}} + \frac{1}{h_n^{2p-1}} \sum_{t=s_n}^{\infty} [\alpha(t)]^{1-2/\delta}\right).$$

The second summand in (5.33) tends to zero by (A.8) and, from definition of k_n, it follows that

$$\frac{k_n s_n}{nh_n^{2p-1}} \approx \frac{n}{b_n + s_n} \frac{s_n}{nh_n^{2p-1}} \approx \frac{s_n}{b_n h_n^{2p-1}} \to 0,$$

as $n \to \infty$, by (5.32). Hence it is deduced that $1/nE(|\Sigma_{s,(n)}|^2) \to 0$ as $n \to \infty$.

Taking into account that $k_n < b_n + s_n < 2b_n$, the same arguments employed for bounding $n^{-1}E(|\Sigma_{s,(n)}|^2)$ lead to

$$\frac{1}{n}E(|\Sigma_{r,(n)}|^2) \leq C\frac{b_n}{(nh_n^{2p-1})} \approx \frac{C}{r_n} \longrightarrow 0.$$

Step 3. The summands in the large blocks, $\Sigma_{b,(n)}$, are shown to be asymptotically independent. Namely, it is required to prove that, as $n \to \infty$,

$$(5.34)\quad \left| E(e^{iu\Sigma_{b,(n)}}) - \prod_{j=0}^{k_n-1} E(e^{iu\Phi_j}) \right| \longrightarrow 0.$$

From Volkonskii and Rozanov (1959) lemma we have

$$\left| E(e^{iu\Sigma_{b,(n)}}) - \prod_{j=0}^{k_n-1} E(e^{iu\Phi_j}) \right| \leq 16(k_n - 1)\alpha(s_n) \approx \frac{n}{b_n}\alpha(s_n),$$

and now (5.22) and (5.21) lead to

$$(5.35)\quad \frac{n}{b_n}\alpha(s_n) \leq \frac{n}{b_n} \sum_{t=s_n}^{\infty} [\alpha(t)]^{1-2/\delta} \approx \frac{r_n}{\sqrt{nh_n^{2p-1}}} \sum_{t=s_n}^{\infty} [\alpha(t)]^{1-2/\delta} \to 0$$

and (5.34) is stated.

Step 4. We prove that

$$\frac{1}{n}\sum_{j=0}^{k_n-1} E\left(\Phi_j\right)^2 \longrightarrow \sigma_Q^2, \qquad \text{as} \quad n\to\infty, \tag{5.36}$$

where $\sigma_Q^2=\sum_{j=0}^{p}\sum_{i=0}^{p} a_j a_i \theta_{j+i-1}\nu_{j+i} f_X(x)\sigma_Y^2(x)$.

First, (3.6) yields $\mathrm{Var}(\sqrt{nh_n}\,Q_n)\longrightarrow\sigma_Q^2$, namely, $n^{-1}E(\Sigma_{(n)})^2$ tends to σ_Q^2. But this fact together with established results in Step 2 lead to

$$\frac{1}{n}E(\Sigma_{b,(n)})^2 \longrightarrow \sigma_Q^2, \qquad \text{as} \quad n\to\infty. \tag{5.37}$$

Then, since

$$\frac{1}{n}E(\Sigma_{b,(n)})^2 = \frac{1}{n}\sum_{j=0}^{k_n-1}E(\Phi_j)^2+\frac{2}{n}\sum_{\substack{i=0\\ i>j}}^{k_n-1}\sum_{j=0}^{k_n-1}E(\Phi_i\Phi_j), \tag{5.38}$$

the limit (5.36) follows from (5.37) and (5.38) once we have shown that the second summand in (5.38) tends to zero as $n\to\infty$. For this, by using the arguments employed before for bounding N, we find

$$\frac{1}{n}\sum_{\substack{i=0\\ i>j}}^{k_n-1}\sum_{j=0}^{k_n-1}E(\Phi_i\Phi_j)\le C\frac{1}{h_n^{2p-1}}\sum_{t=s_n}^{\infty}[\alpha(t)]^{1-2/\delta}\longrightarrow 0, \qquad \text{as} \quad n\to\infty.$$

Step 5. Let M be a fixed truncation point and let us denote

$$Y_{t,M}=Y_t I\{|Y_t|\le M\}. \tag{5.39}$$

We have that $Y_t=Y_{t,M}+\overline{Y}_{t,M}$, with $\overline{Y}_{t,M}=Y_tI\{|Y_t|>M\}$. By replacing Y_t by $Y_{t,M}$ we can write

$$\begin{aligned}
&m_M(x)=E(Y_{t,M}/X_t=x),\\
&\sigma_{Y,M}^2(x)=E((Y_{t,M}-m_M(X_t))^2/X_t=x),\\
&\xi_{t,M,(n)}=\sqrt{h_n}G_{t,(n)}(X_t-x)(Y_{t,M}-m_M(X_t)),\\
&\sqrt{nh_n}\,Q_{n,M}=\frac{1}{\sqrt{n}}\Sigma_{M,(n)}=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\xi_{t,M,(n)},\\
&\overline{Q}_{n,M}=Q_n-Q_{n,M}.
\end{aligned}$$

Then, reasoning as in the proof of Theorem 2, we can obtain

$$\lim_{n\to\infty}\mathrm{Var}\left(\sqrt{nh_n}\,Q_{n,M}\right)-\sigma_{Q,M}^2(x)=\sum_{i=0}^{p}\sum_{j=0}^{p}a_ia_j\theta_{i+j-1}\nu_{i+j}f_X(x)\sigma_{Y,M}^2(x)$$

Step 6. The asymptotic normality of $Q_{n,M}$ is established.

From Steps 1 to 4, it suffices to shown that $\Phi_{j,b,M}$ (the big blocks with the truncated variables $\xi_{t,M,(n)}$) satisfy the standard Lindeberg-Feller condition for asymptotic normality under independence. Here, this condition takes the form

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{k_n-1}E\left(\Phi_{j,b,M}^2 I\left\{|\Phi_{j,b,M}|\geq\varepsilon\sigma_{Q,M}\sqrt{n}\right\}\right)\longrightarrow 0, \tag{5.40}$$

$$\text{as}\quad n\to\infty,\quad \forall\varepsilon>0,\quad \varepsilon\in R.$$

By Assumption (A.1), we have

$$|\Phi_{j,b,M}|\leq\sum_{i=1}^{b_n}|\xi_{e_j+i,M,(n)}|\leq Cb_n h_n^{1/2-p},$$

hence

$$\max_{0\leq j\leq k_n-1}|n^{-1/2}\Phi_{j,b,M}|\leq C\frac{b_n}{\sqrt{nh_n^{2p-1}}}\approx\frac{1}{r_n}\longrightarrow 0,\qquad\text{as}\quad n\to\infty.$$

Therefore, $\{|\Phi_{j,b,M}|\geq\varepsilon\sigma_{Q,M}\sqrt{n}\}$ is an empty set when n is large enough, and therefore (5.40) holds and

$$\sqrt{nh_n}\,Q_{n,M}\xrightarrow{d}N(0,\sigma_{Q,M}^2). \tag{5.41}$$

is concluded.

Step 7. It is proved that

$$\varphi_{Q_n}(t)\longrightarrow\varphi_Z^{\sigma_Q^2}(t),\qquad\text{as}\quad n\to\infty, \tag{5.42}$$

where $\varphi_{Q_n}(t)$ and $\varphi_Z^{\sigma_Q^2}(t)$ denote the characteristic functions of $\sqrt{nh_n}\,Q_n$ and of a random variable $N(0,\sigma_Q^2(x))$ respectively.

In order to show (5.42), we proceed as follows. By using the same notation that in (5.42), we have

$$|\varphi_{Q_n}(t)-\varphi_Z^{\sigma_Q^2}(t)|\leq|\varphi_{Q_{n,M}}(t)||\varphi_{\overline{Q}_{n,M}}(t)-1|+|\varphi_Z^{\sigma_{Q,M}^2}(t)-\varphi_Z^{\sigma_Q^2}(t)|$$
$$+|\varphi_{Q_{n,M}}(t)-\varphi_Z^{\sigma_{Q,M}^2}(t)|\equiv S_1+S_2+S_3.$$

Next, the convergence to zero of each $S_i, i=1,2,3$, as $n\to\infty$ is shown. First, in the same way that in proof of Theorem 2, it can be proved that

$$\lim_{n\to\infty}\mathrm{Var}(\sqrt{nh_n}\,\overline{Q}_{n,M})=\sum_{i,j=0}^{p}a_ia_j\theta_{i+j-1}\nu_{i+j}f_X(x)\mathrm{Var}(\overline{Y}_{t,M}/X_t=x).$$

Nevertheless, if $M\uparrow\infty$ then $\mathrm{Var}(\overline{Y}_{t,M}/X_t=x)$ tends to zero by the dominated convergence theorem. Therefore S_1 converges to zero.

The dominated convergence theorem again and assumption (A.9) lead to the convergence to zero of the second term (S_2) as $M \uparrow \infty$.

Finally, the convergence to zero of S_3 follows from (5.41) and the Levy theorem, for every $M > 0$.

Thus, (5.42) has been stated and then the Cramer theorem allow us to conclude that $\sqrt{nh_n}\, Q_n$ converges in distribution to $N(0, \sigma_Q^2)$ and the proof of Theorem 3 is completed. □

Acknowledgements

The authors wish to thank an Associated Editor and two referees for their helpful comments and suggestions.

References

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots, *J. Amer. Statist. Assoc.*, **74**, 829–836.

Cleveland, W. S. and Devlin, S. J. (1988). Locally weighted regression: an approach to regression anlysis by local fitting, *J. Amer. Statist. Assoc.*, **83**, 597–610.

Devroye, L. and Wagner, T. J. (1980). On the L_1 convergence of kernel estimators of regression function with application in discrimination, *Z. Wahrsch. Verw. Gebiete*, **51**, 15–25.

Doukhan, P. (1995). *Mixing, Properties and Examples*, Lecture Notes in Statistics, Vol. 85, Springer, Berlin.

Eubank, R. L. (1988). *Spline Smoothing and Nonparametric Regression*, Marcel Dekker, New York.

Fan, J. (1992). Design-adaptive nonparametric regression, *J. Amer. Statist. Assoc.*, **87**, 998–1004.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiency, *Ann. Statist.*, **21**, 196–216.

Fan, J. and Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers, *Ann. Statist.*, **20**, 2008–2036.

Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaption, *J. Roy. Stat. Soc. Ser. B*, **57**(2), 371–394.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*, Chapman and Hall, London.

Fan, J., Gasser, T., Gijbels, I., Brockmann, M. and Engel, J. (1997). Local polynomial regression: optimal kernels and asymptotic minimax efficiency, *Ann. Inst. Statist. Math.*, **49** (1), 79–99.

Greblicky, W. and Pawlak, M. (1987). Necessary and sufficient consistency conditions for a recursive kernel regression estimate, *J. Multivariate Anal.*, **23**, 67–76.

Härdle, W. (1990). Applied nonparametric regression, *Econometric Society Monographs*, No. 19, Cambridge University Press, Cambridge.

Hastie, T. and Loader, C. (1993). Local regression: automatic kernel carpentry, *Statist. Sci.*, **8** (2), 120–143.

Krzyzak, A. and Pawlak, M. (1984). Almost everywhere convergence of recursive regression function estimate and classification, *IEEE Trans. Inform. Theory*, **30**, 91–93.

Lejeune, M. (1985). Estimation non-paramétrique par noyaux: régression polynomiale mobile, *Re. Statist. Appli.*, **33**, 43–68.

Masry, E. (1996a). Multivariate regression estimation: local polynomial fitting for time series, *Stochastic Process. Appl.*, **65**, 81–101.

Masry, E. (1996b). Multivariate local polynomial regression for time series: uniform strong consistency and rates, *J. Time Ser. Anal.*, **17**, 571–599.

Masry, E. and Fan, J. (1997). Local polynomial estimation of regression function for mixing processes, *Scand. J. Statist.*, **24**, 165–179.

Müller, H. G. (1988). *Nonparametric Regression Analysis of Longitudinal Data*, Lecture Notes in Statistics, Vol. 46, Springer, Berlin.

Rosenblatt, M. (1956). A central limit theorem and strong mixing condition, *Proc. Nat. Acad. Sci.*, **42**, 43–47.

Roussas, G. G. (1992). Exact rates of almost sure convergence of a recursive kernel estimate of a probability density function: application to regression and hazard rate estimation, *Nonparametric Statistics*, **1**, 171–195.

Roussas, G. G and Tran, L. T. (1992). Asymptotic normality of the recursive kernel regression estimate under dependence conditions, *Ann. Statist.*, **20** (1), 98–120.

Ruppert, D. and Wand, P. (1994). Multivariate locally weighted least squares regression, *Ann. Statist.*, **22**, 1346–1370.

Seifert, B. and Gasser, T. (1996). Finite sample variance of local polynomials: analysis and solutions, *J. Amer. Statist. Assoc.*, **91**, 267–275.

Stone, C. J. (1977). Consistent nonparametric regression, *Ann. Statist.*, **5**, 595–620.

Volkonskii, V. A. and Rozanov, Yu. A. (1959). Some limit theorems for random functions, *Theory Probab. Appl.*, **4**, 178–197.

Wheeden, R. L. and Zygmund, A. (1977). *Measure and Integral*, Marcel Dekker, New York.