# Recursive local polynomial regression under dependence conditions

**Juan M. Vilar-Fernández**[*] **and José A. Vilar-Fernández**
*Departamento de Matemáticas*
*Universidade de A Coruña, Spain.*

## Abstract

In the case of the random design nonparametric regression, one recursive local polynomial smoother is considered. Expressions for the bias and the variance matrix of the estimators of the regression function and its derivatives are obtained under dependence conditions (strongly mixing processes). The obtained Mean Squared Error is shown to be larger than those of the analogous nonrecursive regression estimators, although retaining the same convergence rate. The properties of strong consistency with convergence rates are established for the proposed estimators. Finally, in order to analyse the influence of both the sample size and the dependence in the behaviour of the proposed recursive estimator, a simulation study is performed.

**Key Words:** Local polynomial fitting, recursive nonparametric estimation, strongly mixing processes.
**AMS subject classification:** 62G07, 62H12, 62M09.

## 1 Introduction

In recent years, the local polynomial regression estimator has received increasing attention and it has gained acceptance as an attractive method of nonparametric estimation of a regression function and its derivatives. This estimator is obtained by locally fitting a $p$-th degree polynomial to the data via weighted least squares, and it has favorable asymptotic properties compared with other classical kernel regression estimators. Other additional advantages of local polynomial fitting are its improved boundary behaviour, its adaptation to estimate regression function derivatives and its nice minimax properties.

Since earlier work on local least squares regression estimators by Stone (1977) and Cleveland (1979), further relevant contributions to the theory and computation of these estimators have appeared. Some significant recent references are Lejeune (1985), Tsybakov (1990), Fan (1992, 1993), Hastie and Loader (1993), Fan and Gijbels (1992, 1995), Ruppert and Wand (1994) and Ruppert *et al.* (1995). A motivation and study of this smoothing method can also be found in the recent monograph by Fan and Gijbels (1996).

This paper is concerned with recursive regression estimation based on local polynomial fitting. As is well known, the recursivity property is particularly interesting when the sample data are obtained by means of some observational mechanism that allows an increase in the sample size over time. This situation is usual in many control and supervision problems and, above all, in time series analysis. In the above cases, the recursive estimates allow us to update the estimations as additional observations are obtained, unlike nonrecursive methods where estimates must be completely recalculated when each additional item of data is received. From a practical point of view, this iterative procedure provides an important saving in computational time and memory, so that the updating of the estimates is independent of the previous sample size. Works studying recursive nonparametric regression estimators include those by Revesz (1977), Devroye and Wagner (1980), Krzyzak and Pawlak (1984), and Greblicky and Pawlak (1987) among others.

As indicated above the use of recursive estimates is particularly appropriate in time series analysis and that is why the present study of this is performed for models of dependent observations. The statistical properties of local polynomial regression for dependent data have been studied in recent papers by Masry and Fan (1997) and Masry (1996a,b) and relevant papers about kernel-type recursive estimators with dependent data include those by Rousas (1992) and Roussas and Tran (1992).

In a recent paper, Vilar and Vilar (1998) proposed a recursive local polynomial kernel estimator of the regression function and its derivatives and studied its asymptotic behaviour. Now, in the current, a more general recursive local polynomial estimator is formulated and both quadratic-mean and strong consistency for this estimator are established along with the corresponding rates of convergence.

The organization of this paper is as follows. In the second section, an

estimator for the regression function and its derivatives is defined and a recursive algorithm is derived. In the subsequent two sections asymptotic properties of the proposed estimator are presented. Thus, in Section 3, expressions for the bias and variance/covariance matrix are obtained. In Section 4, almost sure convergence is proved and the rates of convergence are established. In Section 5, a simulation study is presented which allows us to achieve two important goals: firstly, to illustrate the finite sample behaviour of the proposed estimator and secondly, to make a comparative analysis between this estimator and other analogous nonrecursive and recursive estimators. Finally, the appendix is devoted to the proofs of the results.

## 2 Recursive local polynomial regression

Let $(X, Y)$ be a stationary stochastic process in $\mathbb{R}^2$ with unknown joint density function $f(x, y)$. Our goal is to estimate the regression function, $m(x) = E(Y \mid X = x)$, and its derivatives based on an observed sample $\{(X_t, Y_t)\}_{t=1}^n$. If we assume that the $(p+1)$-th derivative of the regression function at the point $x$ exists, then we can estimate $\beta_j = m^{(j)}(x)/(j!)$ for $j = 0, 1, \ldots, p$, by minimizing the function

$$\Psi_{(n)}(\vec{\beta}) = \sum_{t=1}^n \left( Y_t - \sum_{j=0}^p \beta_j (X_t - x)^j \right)^2 \omega_{n,t}, \qquad (2.1)$$

where $\vec{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$ and the weights are $\omega_{n,t} = n^{-1} K_n (X_t - x)$ and $K_n(u) = h_n^{-1} K\left(h_n^{-1} u\right)$, $K$ is a kernel function and $h_n$ the bandwidth. The estimators of $\vec{\beta}$ are obtained as the solution to the weighted least squares problem given in (2.1) and are called local polynomial kernel estimators. It is interesting to observe that this class of estimators includes the classical Nadaraya-Watson estimator, which minimize (2.1) when $p = 0$. Of special interest also is the local linear kernel estimator corresponding to $p = 1$.

In this paper we study this class of estimators, modifying the weights as follows

$$\omega_{n,t} = \frac{h_t^\eta}{H_n(\eta)} K_t \left( X_t - x \right), \qquad (2.2)$$

where $h_t$ is a sequence of bandwidths and $H_n(\eta) = \sum_{t=1}^n h_t^\eta$ with $\eta \in [0, 1]$.

These weights are a particular case of those used by Deheuvels (1974) in a context of nonparametric density estimation. More generally, Deheuvels suggests the choice

$$
\omega_{n,t} = \frac{h_t H(h_t)}{\displaystyle\sum_{i=1}^{n} H(h_i) h_i} K_t \left( X_t - x \right),
$$

where $H : (0, \infty) \to (0, \infty)$ is an arbitrary function. Thus, the weights defined in (2.2) are obtained when $H(h) = h^{\eta-1}$. As will be shown in section 3, $\eta$ can be seen as another smoothing parameter, although its effect over the variance-bias trade-off is very much smaller than the one due to the bandwidth sequence $h_t$. Thus, the selection of an optimal value for $\eta$ is not relevant and this parameter is considered here in order to obtain a more general estimator, which includes such recursive estimators as those defined by Devroye and Wagner (1980) and Masry (1996a).

By minimizing (2.1) with the sequence of weights proposed in (2.2), we obtain the estimator $\widehat{\vec{\beta}}_{(n)}$ which is recursive, as will be shown later. We first introduce the minimization problem in matrix notation for a more concise presentation of the results. Let us denote

$$
\vec{Y}_{(n)} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \qquad X_{(n)} = \begin{pmatrix} 1 & (X_1 - x) & \cdots & (X_1 - x)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (X_n - x) & \cdots & (X_n - x)^p \end{pmatrix}
$$

and let $W_{(n)} = \mathrm{diag}\left(\omega_{n,1}, \ldots, \omega_{n,n}\right)$ be the diagonal array of weights given in (2.2).

Then, by assuming the invertibility of $X_{(n)}^t W_{(n)} X_{(n)}$, standard weighted least squares theory leads to the solution

$$
\widehat{\vec{\beta}}_{(n)} = \left( X_{(n)}^t W_{(n)} X_{(n)} \right)^{-1} X_{(n)}^t W_{(n)} \vec{Y}_{(n)} = S_{(n)}^{-1} \vec{T}_{(n)}, \qquad (2.3)
$$

where $S_{(n)}$ is the $(p+1) \times (p+1)$ array whose $(i, j)$-th element is $s_{n,i+j-2}$ with

$$
s_{n,i} = \frac{1}{H_n(\eta)} \sum_{t=1}^{n} h_t^{\eta} \left( X_t - x \right)^i K_t \left( X_t - x \right), \qquad \text{for } 0 \le i \le 2p, \qquad (2.4)
$$

and $\vec{T}_{(n)} = (t_{n,0}, t_{n,1}, \ldots, t_{n,p})^t$ with

$$t_{n,i} = \frac{1}{H_n(\eta)} \sum_{t=1}^{n} h_t^{\eta} \left( X_t - x \right)^i K_t \left( X_t - x \right) Y_t, \qquad \text{for } 0 \leq i \leq p. \quad (2.5)$$

The case of $\eta = 0$ in (2.3), giving rise to $\omega_{n,t} = n^{-1} K_t \left( X_t - x \right)$, has recently been studied by Vilar and Vilar (1998). Another case of particular importance is that corresponding to constant local fitting ($p = 0$). In this case, when $\eta = 0$, a recursive version of the Nadaraya-Watson kernel estimator is obtained, whose explicit expression is given by

$$\widehat{m}_n(x) = \widehat{\beta}_{n,0}(x) = \frac{\sum_{t=1}^{n} K_t(X_t - x)Y_t}{\sum_{t=1}^{n} K_t(X_t - x)}. \quad (2.6)$$

Some results about the asymptotic behaviour of (2.6) can be found in Greblicky and Pawlak (1987) using independent observations, and in Roussas and Tran (1992) under dependence conditions.

Note that both $S_{(n)}$ and $\vec{T}_{(n)}$ can be computed recursively. In fact, it follows directly from (2.4) and (2.5) that

$$S_{(n+1)} \;=\; \frac{H_n(\eta)}{H_{n+1}(\eta)} S_{(n)} + \omega_{n+1,n+1} \vec{x}_{(n+1)} \vec{x}_{(n+1)}^t, \qquad (2.7)$$

$$\vec{T}_{(n+1)} \;=\; \frac{H_n(\eta)}{H_{n+1}(\eta)} \vec{T}_{(n)} + \omega_{n+1,n+1} Y_{n+1} \vec{x}_{(n+1)}, \qquad (2.8)$$

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

If we now substitute (2.7) and (2.8) in (2.3) it is straightforward to derive that

$$\widehat{\vec{\beta}}_{(n+1)} = \widehat{\vec{\beta}}_{(n)} + \omega_{n+1,n+1} \left( Y_{n+1} - \vec{x}_{(n+1)}^t \widehat{\vec{\beta}}_{(n)} \right) S_{(n+1)}^{-1} \vec{x}_{(n+1)}. \quad (2.9)$$

This identity allows us to conclude that the estimator (2.3) can be computed recursively. Moreover, in order to update the matrix $S_{(n+1)}^{-1}$ for each additional observation, matrix algebra can be used to obtain

$$S_{(n+1)}^{-1} = \left( 1 + \tilde{h} \right) \left( S_{(n)}^{-1} - \frac{\tilde{h} K_{n+1} \left( X_{n+1} - x \right) S_{(n)}^{-1} \vec{x}_{(n+1)} \vec{x}_{(n+1)}^t S_{(n)}^{-1}}{1 + \tilde{h} K_{n+1} \left( X_{n+1} - x \right) \vec{x}_{(n+1)}^t S_{(n)}^{-1} \vec{x}_{(n+1)}} \right),$$

$$(2.10)$$

where $\tilde{h} = h_{n+1}^{\eta} / H_n(\eta)$.

From expressions (2.9) and (2.10) a recursive algorithm for obtaining the estimator $\widehat{\vec{\beta}}_{(n+1)}$ can be easily deduced by following the lines of the Plackett-Kalman iterative algorithm for regression models. Figure 1 summarizes its flowchart.

## 3   Mean squared error

In this section we establish precise asymptotic expressions for both the bias and the variance/covariance matrix of the regression function estimators and their derivatives as defined in (2.3). In order to do this, the following assumptions will be employed in our analysis:

**(A1)** The kernel function $K(u)$ is bounded with a bounded support.

**(A2)** The sequence of bandwidths $\{h_n\}$ satisfies $h_n > 0$ , $h_n \downarrow 0$, $nh_n \uparrow \infty$ and $\theta_\gamma = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \left( \frac{h_i}{h_n} \right)^{\gamma} < \infty$ for $0 \le \gamma \le 2\eta + 2p - 1$.

**(A3)** Both stationary processes $(X_t, Y_t)$ are strongly mixing ($\alpha$-mixing) and $\sum_{k=1}^{\infty} k^{\epsilon} \left[ \alpha(k) \right]^{1-2/\delta} < \infty$, for some $\delta > 2$ and $\epsilon > 1 - 2/\delta$.

**(A4)** i) The joint probability density of $X_t$ and $X_{t+s}$, $f(x, y; s)$ satisfies

$$|f(x, y; s) - f_X(x)f_X(y)| \le cst. < \infty, \qquad \text{for all } x, y \text{ and } s \ge 1,$$

where $f_X$ denotes the marginal density of $X_t$.

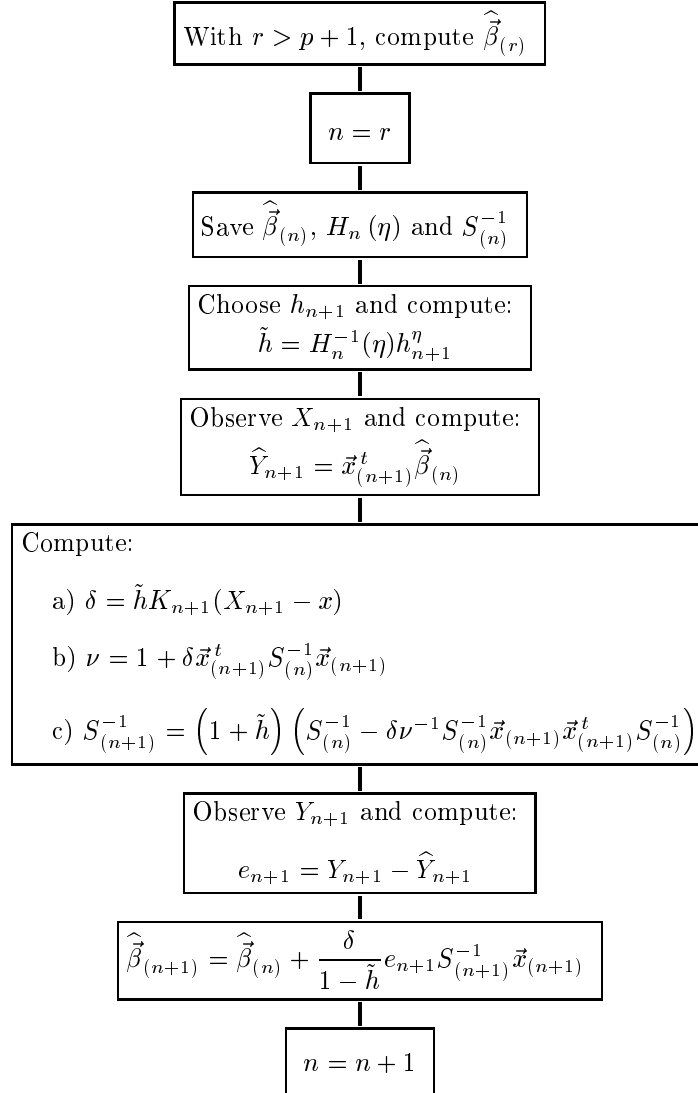ii) The conditional probability density $f_{X_t|Y_t}$ verifies

$$f_{X_t|Y_t}(x_t|y_t) \le cst. < \infty.$$

iii) For all $s \ge 1$, the conditional probability density $f_{X_t X_{t+s}|Y_t Y_{t+s}}$ verifies

$$f_{X_t X_{t+s}|Y_t Y_{t+s}}(x_t x_{t+s}|y_t y_{t+s}) \le cst. < \infty.$$

**(A5)** $E|Y_t|^{\delta} < \infty$, for some $\delta > 2$.

From now on we use the following notation. Let $\mu_j = \int u^j K(u) \, du$ and $\nu_j = \int u^j K^2(u) \, du$ and let $\vec{\mu}_\theta = (\theta_{\eta+p+1}\mu_{p+1}, \ldots, \theta_{\eta+2p+1}\mu_{2p+1})^t$.

$$\boxed{\text{With } r > p+1, \text{ compute } \widehat{\vec{\beta}}_{(r)}}$$

$$\boxed{n = r}$$

$$\boxed{\text{Save } \widehat{\vec{\beta}}_{(n)}, \ H_n\left(\eta\right) \text{ and } S_{(n)}^{-1}}$$

$$\boxed{\begin{array}{c}\text{Choose } h_{n+1} \text{ and compute:} \\ \tilde{h} = H_n^{-1}(\eta)h_{n+1}^{\eta}\end{array}}$$

$$\boxed{\begin{array}{c}\text{Observe } X_{n+1} \text{ and compute:} \\ \widehat{Y}_{n+1} = \vec{x}_{(n+1)}^{t}\widehat{\vec{\beta}}_{(n)}\end{array}}$$

Compute:

a) $\delta = \tilde{h}K_{n+1}(X_{n+1} - x)$

b) $\nu = 1 + \delta\vec{x}_{(n+1)}^{t}S_{(n)}^{-1}\vec{x}_{(n+1)}$

c) $S_{(n+1)}^{-1} = \left(1 + \tilde{h}\right)\left(S_{(n)}^{-1} - \delta\nu^{-1}S_{(n)}^{-1}\vec{x}_{(n+1)}\vec{x}_{(n+1)}^{t}S_{(n)}^{-1}\right)$

$$\boxed{\begin{array}{c}\text{Observe } Y_{n+1} \text{ and compute:} \\ e_{n+1} = Y_{n+1} - \widehat{Y}_{n+1}\end{array}}$$

$$\boxed{\widehat{\vec{\beta}}_{(n+1)} = \widehat{\vec{\beta}}_{(n)} + \frac{\delta}{1-\tilde{h}}e_{n+1}S_{(n+1)}^{-1}\vec{x}_{(n+1)}}$$

$$\boxed{n = n+1}$$

Figure 1: *Recursive algorithm for updating* $\widehat{\vec{\beta}}_{(n)}$.

**Theorem 3.1.** *Let us assume that conditions* (A1)-(A5) *hold. Then*

(i) *For every continuity point $x$ of $f_X$, we have*

$$H_{(n)} E\left[\left.\widehat{\vec{\beta}}_{(n)} - \vec{\beta}\,\right|(X_1, \ldots, X_n)\right] =$$

$$\frac{m^{(p+1)}(x)}{(p+1)!} h_n^{p+1} S^{-1} \vec{\mu}_\theta + o_P\left(h_n^{p+1}(1, \ldots, 1)^t\right),$$

*where $H_{(n)} = diag\left(1, h_n, \ldots, h_n^p\right)$ and $S$ denotes the $(p+1) \times (p+1)$ matrix whose $(i, j)$-th element is such that $s_{i,j} = s_{i+j-2}$ with $s_k = \theta_{\eta+k} \mu_k / \theta_\eta$ for $0 \le k \le 2p$.*

(ii) *If $\sigma_Y^2(x)$ denotes the $Var\left(Y \,|\, X = x\right)$, then for every continuity point $x$ of $\sigma_Y^2 f_X$ whenever $f_X(x) > 0$, we have*

$$Var\left[\left. H_{(n)} \widehat{\vec{\beta}}_{(n)} \,\right|(X_1, \ldots, X_n)\right] = \frac{1}{nh_n} \frac{\sigma_Y^2(x)}{f_X(x)} S^{-1} \tilde{S} S^{-1} + o_P\left(\frac{1}{nh_n}\right),$$

*where $\tilde{S}$ denotes the matrix whose $(i, j)$-th element is such that $\widetilde{s}_{i,j} = \widetilde{s}_{i+j-2}$ with $\widetilde{s}_k = \theta_{2\eta+k-1} \nu_k / \theta_\eta^2$, for $0 \le k \le 2p$.*

Asymptotic expressions for the conditional bias and variance of the recursive local polynomial fit proposed in this paper are directly derived from Theorem 3.1. For $j = 0, \ldots, p$, these are given by:

$$\text{Bias}\left[\left. \widehat{m}^{(j)}(x) \,\right|(X_1, \ldots, X_n)\right] = h_n^{p+1-j} \frac{m^{(p+1)}(x)}{(p+1)!} j! B_j\left(1 + o_P(1)\right), \quad (3.1)$$

$$\text{Var}\left[\left. \widehat{m}^{(j)}(x) \,\right|(X_1, \ldots, X_n)\right] = \frac{1}{nh_n^{2j+1}} \frac{\sigma_Y^2(x)}{f_X(x)} (j!)^2 V_j\left(1 + o_P(1)\right), \quad (3.2)$$

where the terms $B_j$ and $V_j$ denote the $j$-th element of $S^{-1} \vec{\mu}_\theta$ and the $j$-th diagonal element of $S^{-1} \tilde{S} S^{-1}$ respectively.

The expressions (3.1) and (3.2) generalize those derived by Vilar and Vilar (1998) for the particular case of $\eta = 0$ and coincide with those obtained in Masry and Fan (1997) for the nonrecursive local polynomial estimator under similar dependence conditions but with different values for $B_j$ and $V_j$. Therefore, the most important consequence of Theorem 3.1 is that both the recursive estimate and the nonrecursive one exhibit the same convergence rate with respect to their mean squared errors.

In order to compare several kernel regression smoothers, we have listed in Table 1 the asymptotic expressions for the conditional bias and the conditional variance of the following estimators: the Nadaraya-Watson estimator (NW), the estimator obtained by local linear fitting (LL), and their recursive versions (RNW and RLL), which are obtained from (2.3) letting $p = 0$ and $p = 1$ respectively. These asymptotic expressions are derived from Theorem 3.1 for the estimators based on local linear fitting and from Theorem 4.1 in Vilar and Vilar (1993) for the recursive version of the Nadaraya-Watson estimator.

| Estimator | Bias | Variance |
|---|---|---|
| NW ($p = 0$) | $h_n^2 \left( \dfrac{1}{2} m''(x) + \dfrac{m'(x) f_X'(x)}{f_X(x)} \right) \mu_2$ | $\dfrac{1}{n h_n} \dfrac{\sigma_Y^2(x)}{f_X(x)} \nu_0$ |
| RNW ($p = 0$) | $h_n^2 \left( \dfrac{1}{2} m''(x) + \dfrac{m'(x) f_X'(x)}{f_X(x)} \right) \dfrac{\theta_{\eta+2}}{\theta_\eta} \mu_2$ | $\dfrac{1}{n h_n} \dfrac{\sigma_Y^2(x)}{f_X(x)} \dfrac{\theta_{2\eta-1}}{\theta_\eta^2} \nu_0$ |
| LL ($p = 1$) | $h_n^2 \dfrac{m''(x)}{2} \mu_2$ | $\dfrac{1}{n h_n} \dfrac{\sigma_Y^2(x)}{f_X(x)} \nu_0$ |
| RLL ($p = 1$) | $h_n^2 \dfrac{m''(x)}{2} \dfrac{\theta_{\eta+2}}{\theta_\eta} \mu_2$ | $\dfrac{1}{n h_n} \dfrac{\sigma_Y^2(x)}{f_X(x)} \dfrac{\theta_{2\eta-1}}{\theta_\eta^2} \nu_0$ |

*Table 1: Pointwise bias and variance of local polynomial regression smoothers.*

Table 1 shows clearly that the bounds obtained for the pointwise bias in both the recursive case ($\text{Bias}_R$) and the nonrecursive case ($\text{Bias}_{NR}$) satisfy the relationship

$$\text{Bias}_R = B(\eta) \, \text{Bias}_{NR}, \ \text{ with } B(\eta) = \frac{\theta_{\eta+2}}{\theta_\eta},$$

whereas the corresponding relationship for the variance is given by

$$\text{Var}_R = V(\eta) \, \text{Var}_{NR}, \ \text{ with } V(\eta) = \frac{\theta_{2\eta-1}}{\theta_\eta^2}.$$

Thus, if the usual selection of bandwidth $h_n = Cn^{-\rho}$ is considered, then for all $\rho \in [0, 1]$ we have

$$B(\eta) = \frac{1 - \rho\eta}{1 - (2 + \eta)\rho} \ \text{ and } \ V(\eta) = \frac{(1 - \rho\eta)^2}{1 - \rho(2\eta - 1)},$$

with $B(\eta) > 1$ and $V(\eta) < 1$, in this case. Therefore, it is clear that the recursive estimators exhibit larger bias and smaller variance than their analogous nonrecursive estimators. If the variance-bias trade-off that performs

the mean squared error is analysed, it can be derived that the mean squared error is larger for the recursive estimator than for the non-recursive one.

Note that if $\eta \in [0, 1]$, then $B(\eta)$ increases with $\eta$, unlike $V(\eta)$ which decreases with $\eta$. Thus, values for $\eta$ close to 1 lead to smoother estimation and values for $\eta$ close to 0 produce a less biased estimate.

## 4   Almost sure convergence

Next we establish the almost sure consistency for the recursive estimators $\widehat{\vec{\beta}}_{(n)}$. The employed technique makes use of one result of almost sure consistency for strongly mixing processes according to Masry (1987) (see Theorem 3 in Masry's paper). The proof of this result is based on the notion and properties of mixingales due to McLeish (1975) and it was used to obtain rates of almost sure convergence for kernel-type recursive estimators of the probability density function.

In addition to the conditions (A1)-(A5), the following assumptions will be needed in order to show the almost sure convergence of $\widehat{\vec{\beta}}_{(n)}$.

**(A6)** The function $m^{(p+1)}(x)$ is bounded and is uniformly continuous on $\mathbb{R}$.

**(A7)** The sequence $\{h_n\}$ is such that $nh_n^{j+\eta} \uparrow \infty$, for $j = 0, \ldots, 2p + 1$, and $\sum_{t=1}^{\infty}(th_t^{1-1/\delta})^{-2} < \infty$, for some $\delta > 2$.

**(A8)** For some $\delta > 2$ and $\varepsilon > 0$, the mixing coefficients $\alpha(k)$ satisfy

$$\sum_{n=1}^{\infty} (\log n) (\log \log n)^{1+\varepsilon} [\alpha(n)]^{1-2/\delta} \left[ \sum_{t=n}^{\infty} \left( th_t^{1-1/\delta} \right)^{-2} \right] < \infty.$$

**Theorem 4.1.** *If the assumptions* (A1), (A2), (A4.ii), (A5), (A6), (A7) *and* (A8) *hold, then*

$$H_{(n)} \left( \widehat{\vec{\beta}}_{(n)} - \vec{\beta} \right) \longrightarrow 0, \ \ almost \ surely \ as \ n \to \infty. \qquad (4.1)$$

The next task is to derive a rate of almost sure convergence for $\widehat{\vec{\beta}}_{(n)}$. For this, we need to change assumption (A8) for the stronger condition:

**(A9)** For some $\delta > 2$ and $\varepsilon > 0$, the mixing coefficients $\alpha(k)$ satisfy

$$\sum_{n=1}^{\infty} (\log n)(\log\log n)^{1+\varepsilon} [\alpha(n)]^{1-2/\delta} < \infty.$$

**Theorem 4.2.** *If the assumptions of Theorem 4.1 and (A9) hold, then*

$$H_{(n)}\left(\widehat{\vec{\beta}}_{(n)} - \vec{\beta}\right) = o\left(\frac{(\log n)(\log\log n)^{1+\varepsilon}}{nh_n^{2(1-1/\delta)}}\right)^{1/2} + O(h_n^{p+1}) \quad a.s. \quad (4.2)$$

*So, if the smoothing parameter is chosen in the way*

$$h_n = cst. \left(\frac{(\log n)(\log\log n)}{n}\right)^{\frac{1}{4+2(p-1/\delta)}} \tag{4.3}$$

*then*

$$\left(\left(\frac{n}{(\log n)(\log\log n)}\right)^{\frac{p+1}{4+2(p-1/\delta)}} \frac{1}{(\log\log n)^{\varepsilon/2}}\right) H_{(n)}\left(\widehat{\vec{\beta}}_{(n)} - \vec{\beta}\right) \longrightarrow 0 \ a.s.$$

The rates of almost sure convergence for the individual components $\widehat{\beta}_{n,j}$ are directly derived from Theorem 4.2.

**Corollary 4.1.** *Under assumptions of Theorem 4.2, we have*

$$\left(\widehat{m}_n^{(j)}(x) - m^{(j)}(x)\right) = o\left(\frac{(\log n)(\log\log n)^{1+\varepsilon}}{nh_n^{2(1+j-1/\delta)}}\right)^{1/2} + O(h_n^{p+1-j})$$

*almost surely, for $j = 0, 1, \ldots, p$.*

*By choosing $h_n$ as in (4.3) we obtain that*

$$\left(\left(\frac{n}{(\log n)(\log\log n)}\right)^{\frac{p+1-j}{4+2(p-1/\delta)}} \frac{1}{(\log\log n)^{\varepsilon/2}}\right) \left(\widehat{m}_n^{(j)}(x) - m^{(j)}(x)\right) \longrightarrow 0,$$

$$\tag{4.4}$$

*almost surely, for $j = 0, 1, \ldots, p$.*

Note that the rate of almost sure convergence for $\widehat{m}_n^{(j)}(x)$ in (4.4) is faster when $\delta$ approaches to 2. Nevertheless, in this case assumptions (A8) and (A9) become more restrictive in the sense that the mixing coefficients $\alpha(k)$ decay more quickly and, therefore, the dependence condition is stronger.

## 5   Simulation study

With the purpose of comparing the proposed nonparametric estimation method with other classical methods, a numerical study with simulated samples is performed in this section.

The regression model is of the form $Y_t = m(X_t) + \varepsilon_t$, with $t = 1, \ldots, n$, where $m(x)$ and $\varepsilon_t$ denote respectively the theoretical regression function and the error of the model. In particular, we have chosen $m(x) = sin(5\pi x)$, the design points $X_t$ have been drawn from a uniform density on $[0, 1]$ and, finally, the responses have been generated using an AR(1) structure for $\varepsilon_t$, $\varepsilon_t = \rho\varepsilon_{t-1} + e_t$, $\{e_t\}$ being a sequence of i.i.d. random variables with a common normal distribution of mean 0 and standard deviation 0.4. The described model therefore satisfies the $\alpha$-mixing dependence condition.

The regression function is estimated at $N = 200$ equally spaced points in the interval $[0, 1]$ and four nonparametric estimators are evaluated: the Nadaraya-Watson estimator (NW), the recursive Nadaraya-Watson given in (2.6) (RNW), the local linear fitting (LL) and the recursive local linear fitting (RLL). As the kernel function we chose the quartic kernel, $K(u) = 15(1-u^2)^2/16$ if $|u| \leq 1$. Concerning the selected bandwidth in the recursive estimators, we have used values in the form $h_t = Ct^{-1/5}$, where $C$ is a parameter to be empirically evaluated by minimizing the average mean squared error (AMSE) for each estimator.

In order to study the behaviour of the RLL estimator and the influence of the sample size we carried out the study for $n = 200$, $300$, $400$ and $500$. In Figure 2, we display the quotient between the mean squared error of the RLL estimator proposed above and the NW, RNW and LL estimators.

The good performance of the recursive estimator by local polynomial fitting (RLL) (which has been computed with $\eta = 0$ in all cases) is clearly shown in Figure 2. Moreover, it is interesting to note that the polynomial estimators have smaller mean squared error than the classical Nadaraya-Watson estimators for all employed sample sizes. On the other hand, although the results provided by nonrecursive estimators present a slight improvement with respect to those obtained by the recursive estimators, the differences are very small. Similar graphs to Figure 2 are obtained when the study is performed with other regression functions.

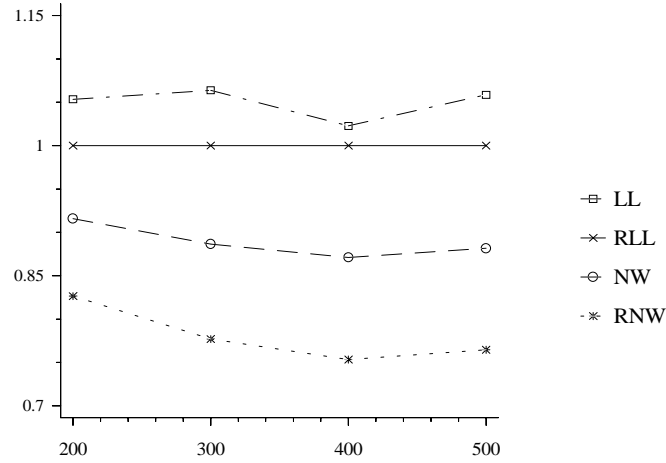In the second step of our study, the performance of the four estimators

*Figure 2: Relative efficiency, as a function of n, of the estimators LL, NW and RNW with respect to the RLL estimator, with $\rho = 0.6$.*

is analysed in terms of the amount of dependence, and similar conclusions can be deduced. In Figure 3, we show the graph of the variance for the four estimators as a function of $\rho$, and Figure 4 displays the graph of the relative efficiency of all estimators with respect to the RLL estimator for each correlation value.

Once more it is clear from Figure 3 and Figure 4 that, independently of the correlation, the proposed recursive estimator exhibits a better performance than both the RNW and the NW estimators, being a little bit worse than the LL estimator.

The above considerations, together with the computational efficiency of recursive estimators with respect to nonrecursive ones when the observations are sequentially received, allows us to state that the proposed estimator is quite competitive.

## Appendix: proofs

The proof of Theorem 3.1 is obtained by following a reasoning scheme analogous to the one carried out in proving Theorems 1-2 in Vilar and Vilar (1998) and is therefore omitted.
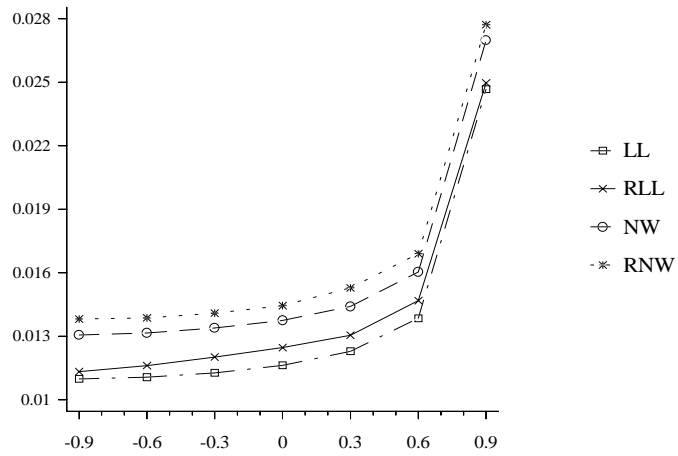
*Figure 3*: *Variance of the four analysed estimators as a function of $\rho$, with $n = 200$.*
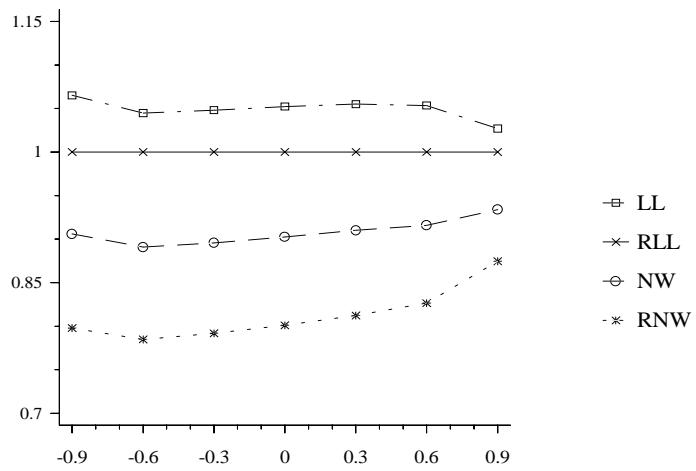


*Figure 4*: *Relative efficiency, as a function of $\rho$, of LL, NW and RNW estimators with respect to the RLL estimator, with $n = 200$.*

## A.1  Proof of Theorem 4.1

In order to establish the validity of (4.1), the following steps are performed:

STEP 1: By the continuity of the first $p + 1$ derivatives of $m$ at the point $x$, we may perform a $p$-th order Taylor series expansion in a neighborhood of $x$, so that for $|X_t - x| \leq h_t$, $t = 1, \ldots, n$, we have

$$m(X_t) = \sum_{j=0}^{p} \frac{m^{(j)}(x)}{j!} (X_t - x)^j + \varphi_t. \qquad (A.1)$$

If an integral remainder $\varphi_t$ is considered, then we find

$$
\begin{aligned}
\varphi_t &= \frac{1}{p!} (X_t - x)^{p+1} \int_0^1 u^p m^{(p+1)} (X_t + u(x - X_t))\, du \\
&= p_t + \frac{m^{(p+1)}(x)}{(p+1)!} (X_t - x)^{p+1},
\end{aligned}
$$

where

$$p_t = \frac{(X_t - x)^{p+1}}{p!} \int_0^1 (1 - w)^p \left[ m^{(p+1)}(x + w(X_t - x)) - m^{(p+1)}(x) \right] dw,$$

and so we can replace (A.1) by

$$m(X_t) = \sum_{j=0}^{p+1} \frac{m^{(j)}(x)}{j!} (X_t - x)^j + p_t. \qquad (A.2)$$

Equation (A.2) can be also written in matrix notation as follows

$$\vec{M}_{(n)} = X_{(n)} \vec{\beta} + \frac{m^{(p+1)}(x)}{(p+1)!} \left( (X_1 - x)^{p+1}, \ldots, (X_n - x)^{p+1} \right)^t + \vec{P}_{(n)}, \quad (A.3)$$

where $\vec{M}_{(n)}$ and $\vec{P}_{(n)}$ denote respectively the vectors $(m(X_1), \ldots, m(X_n))^t$ and $(p_1, \ldots, p_n)^t$.

On the other hand, let us consider the vector $\vec{T}_{(n)}^{\star} = \left( t_{n,0}^{\star}, \ldots, t_{n,p}^{\star} \right)^t$, whose $j$-th component is given by

$$t_{n,j}^{\star} = (H_n(\eta))^{-1} \sum_{t=1}^{n} h_t^{\eta} (X_t - x)^j K_t (X_t - x) (Y_t - m(X_t)). \qquad (A.4)$$

From (A.3) and (A.4) it follows that

$$S_{(n)}^{-1}\vec{T}_{(n)}^{\star} = S_{(n)}^{-1}\vec{T}_{(n)} - \vec{\beta} - \frac{m^{(p+1)}(x)}{(p+1)!}S_{(n)}^{-1}\vec{B}_{(n)} - S_{(n)}^{-1}\vec{R}_{(n)}, \qquad (A.5)$$

where

$$\vec{B}_{(n)} = (s_{n,p+1}, \ldots, s_{n,2p+1})^t,$$

$$\vec{R}_{(n)} = X_{(n)}^t W_{(n)} \vec{P}_{(n)} = (r_{n,0}, \ldots, r_{n,p})^t. \qquad (A.6)$$

Equation (A.5) together with (2.3) lead to

$$\widehat{\vec{\beta}}_{(n)} - \vec{\beta} = S_{(n)}^{-1}\vec{T}_{(n)}^{\star} + \frac{m^{(p+1)}(x)}{(p+1)!}S_{(n)}^{-1}\vec{B}_{(n)} + S_{(n)}^{-1}\vec{R}_{(n)}. \qquad (A.7)$$

The previous equation is analogous to the one deduced by Masry (1996b) for the estimate obtained by local polynomial fitting in the nonrecursive setting. In the same way, in the work mentioned above, expression (A.7) provides us the starting point for establishing the proof of Theorem 4.1. In fact, the next three steps consist of proving the strong consistency of each one of the elements of the matrices on the right-hand side of equation (A.7). In order to do this, the following almost sure convergence result for strong mixing processes, jointly with Kronecker's lemma, will be used.

**Theorem A.1.** (Masry (1987)) *Let* $\{X_t\}$ *be a strong mixing process and let* $\{g_t\}$ *be a sequence of Borel measurable functions on* $\mathbb{R}^d$. *Let* $Z_t = g_t(X_t) - E(g_t(X_t))$ *and put* $S_n = \sum_{t=1}^n Z_t$. *If*

$$\sum_{t=1}^{\infty} \left[ E\left( |Z_t|^\delta \right) \right]^{2/\delta} < \infty, \qquad (A.8)$$

*and*

$$\sum_{n=1}^{\infty} (\log n)(\log\log n)^{1+\varepsilon} [\alpha(n)]^{1-2/\delta} \sum_{t=n}^{\infty} \left[ E\left( |Z_t|^\delta \right) \right]^{2/\delta} < \infty \qquad (A.9)$$

*for some* $\varepsilon > 0$ *and* $\delta > 2$, *then* $\sum_{t=1}^n Z_t$ *converges almost surely to a finite limit as* $n \to \infty$.

STEP 2: We prove that under assumptions (A1), (A2), (A6), (A7) and (A8), we have

$$h_n^{-(p+j+1)} r_{n,j} \longrightarrow 0, \text{ almost surely as } n \to \infty, \text{ for } j = 0, \ldots, p, \quad \text{(A.10)}$$

and therefore we can write

$$h_n^{-(p+1)} H_{(n)}^{-1} \vec{R}_{(n)} \longrightarrow 0, \text{ almost surely as } n \to \infty, \quad \text{(A.11)}$$

where it is important to note that the convergence result (A.11) must be interpreted in the sense that each element of the matrix converges almost surely.

In order to establish (A.10), we proceed as follows. Let us put

$$\frac{1}{h_n^{p+j+1}} r_{n,j} = \frac{1}{h_n^{p+j+1}} \left[ r_{n,j} - E\left(r_{n,j}\right) \right] + \frac{1}{h_n^{p+j+1}} E\left(r_{n,j}\right) = \Delta_r + \Delta_d.$$

First, we will show that the deterministic part, $\Delta_d$, tends to zero. Let

$$V_{t,j} = (X_t - x)^j K_t (X_t - x) p_t.$$

By making a change of variable, we obtain

$$E\left(V_{t,j}\right) = \frac{h_t^{p+j+1}}{p!} \int v^{p+j+1} K(v) I_t(x,v) f\left(x + vh_t\right) dv,$$

where $I_t(x,v) = \int_0^1 (1-w)^p [m^{(p+1)}(x + wvh_t) - m^{(p+1)}(x)] dw$. Under (A6), $|I_t(x,v)| \leq (\int_0^1 (1-w)^p w \, dw)|v| h_t$ and hence

$$\left| E\left(V_{t,j}\right) \right| \leq \frac{h_t^{p+j+2}}{p!} \left( \int_0^1 (1-w)^p w \, dw \right) \left( \int_R \left| v^{p+j+2} K(v) \right| f_X(x + vh_t) dv \right) \quad \text{(A.12)}$$

From the application of Bochner's lemma to the last integral in (A.12), we conclude that

$$E\left(V_{t,j}\right) = O\left(h_t^{p+j+2}\right). \quad \text{(A.13)}$$

From expression (A.6) it follows that

$$r_{n,j} = \sum_{t=1}^n \frac{h_t^\eta}{H_n(\eta)} V_{t,j}, \quad \text{(A.14)}$$

and we can use (A.13) and (A.14) to obtain

$$\frac{1}{h_n}|\Delta_d| = \frac{1}{h_n^{p+j+2}}\left|E\left(r_{n,j}\right)\right| \le \frac{1}{h_n^{p+j+2}}\sum_{t=1}^{n}\frac{h_t^{\eta}}{H_n(\eta)}\left|E\left(V_{t,j}\right)\right| \le$$

$$\text{cst.}\left[\frac{1}{n}\sum_{t=1}^{n}\left(\frac{h_t}{h_n}\right)^{\eta}\right]^{-1}\left[\frac{1}{n}\sum_{t=1}^{n}\left(\frac{h_t}{h_n}\right)^{p+j+2+\eta}\right].$$

Now, under assumption (A2), we can use Toeplitz's lemma to conclude that

$$\lim_{n\to\infty}\frac{1}{h_n}\Delta_d = \text{cst.}\frac{\theta_{p+j+2+\eta}}{\theta_{\eta}} \;\Rightarrow\; \Delta_d = O\left(h_n\right) = o\left(1\right). \tag{A.15}$$

Next, in order to prove that $\Delta_r$ tends to zero almost surely, it is useful to introduce the following notation:

For $j = 0, \ldots, p$ and $t = 1, 2, \ldots,$ let

$$a_{t,j} = h_t^{p+j+1}H_t(\eta) \quad \text{and} \quad Z_{t,j} = \frac{h_t^{\eta}}{a_{t,j}}\left[V_{t,j} - E\left(V_{t,j}\right)\right]. \tag{A.16}$$

From (A.14) and (A.16), we can write

$$\Delta_r = h_n^{-(p+j+1)}\left[r_{n,j} - E\left(r_{n,j}\right)\right] = \frac{1}{a_{n,j}}\sum_{t=1}^{n}a_{t,j}Z_{t,j}. \tag{A.17}$$

Now we show that the sequence $\{Z_{t,j}\}_t$ satisfies conditions (A.8) and (A.9) for some $\varepsilon > 0$ and $\delta > 2$.

In a similar way to that followed to obtain (A.13), we find

$$E\left(\left|V_{t,j}\right|^{\delta}\right) \le \text{cst.}\, h_t^{(p+j+1)\delta+1}. \tag{A.18}$$

From (A.16) and (A.18) it follows that

$$\left[E\left(\left|Z_{t,j}\right|^{\delta}\right)\right]^{2/\delta} \le \text{cst.}\left(\frac{h_t^{\eta}h_t^{1/\delta}}{H_t(\eta)}\right)^2,$$

and, by assumption (A2), we conclude that $H_t(\eta) = O\left(th_t^{\eta}\right)$. Therefore, we have

$$\left[E\left(\left|Z_{t,j}\right|^{\delta}\right)\right]^{2/\delta} \le \text{cst.}\left(\frac{h_t^{1/\delta}}{t}\right)^2. \tag{A.19}$$

Now, the conditions (A.8) and (A.9) are straightforwardly derived from (A.19) and the assumptions (A7) and (A8).

Hence Theorem A.1 is applicable and $\sum_{t=1}^{n} Z_{t,j}$ converges almost surely to a finite limit as $n \to \infty$. From this, and since the sequence $a_{n,j}$ is increasing to infinity by (A7), we can use Kronecker's lemma to obtain the convergence almost surely to zero of $\Delta_r$. This fact, jointly with (A.15), leads to the proof of (A.10).

STEP 3: We prove that under assumptions (A1), (A2), (A7) and (A8), we have

$$h_n^{-j} s_{n,j} \longrightarrow \frac{\theta_{\eta+j}}{\theta_\eta} f_X(x) \mu_j, \text{ almost surely as } n \to \infty, \text{ for } j = 0, \ldots, 2p+1$$

and therefore we can write

$$H_{(n)}^{-1} S_{(n)} H_{(n)}^{-1} \longrightarrow f_X(x) S, \text{ almost surely as } n \to \infty,$$

and

$$h_n^{-(p+1)} H_{(n)}^{-1} \vec{B}_{(n)} \longrightarrow f_X(x) \vec{B}, \text{ almost surely as } n \to \infty,$$

where $\vec{B} = (s_{p+1}, \ldots, s_{2p+1})^t$.

We will start with the identity

$$\frac{1}{h_n^j} s_{n,j} - \frac{\theta_{\eta+j}}{\theta_\eta} f_X(x) \mu_j = \frac{1}{h_n^j} \left[ s_{n,j} - E\left(s_{n,j}\right) \right] +$$

$$\left[ \frac{1}{h_n^j} E\left(s_{n,j}\right) - \frac{\theta_{\eta+j}}{\theta_\eta} f_X(x) \mu_j \right] = \Delta_r' + \Delta_d'.$$

As in the previous proof, it follows that $\Delta_d'$ tends to zero from (A1) and (A2).

If, as in (A.17), we write $\Delta_r' = \left( \sum_{t=1}^{n} a_{t,j}' Z_{t,j}' \right) / a_{n,j}'$, where the sequences $a_{t,j}'$ and $Z_{t,j}'$ are now given by

$$a_{t,j}' = h_t^j H_t(\eta) \quad \text{and} \quad Z_{t,j}' = \frac{h_t^\eta}{a_{t,j}'} \left[ V_{t,j}' - E\left(V_{t,j}'\right) \right], \tag{A.20}$$

with

$$V_{t,j}' = \left(X_t - x\right)^j K_t\left(X_t - x\right),$$

then analogous arguments to those used in step 2 lead to

$$E\left(\left|V'_{t,j}\right|^{\delta}\right) \leq \text{cst.} \frac{h_t^{j\delta+1}}{h_t^{\delta}} \quad \text{and} \quad \left[E\left(\left|Z'_{t,j}\right|^{\delta}\right)\right]^{2/\delta} \leq \text{cst.} \left(\frac{h_t^{1/\delta}}{th_t}\right)^2.$$
$$(A.21)$$

From (A.21) and the assumptions (A7) and (A8), we can show that the sequence $\{Z'_{t,j}\}_t$ satisfies conditions (A.8) and (A.9). Therefore, the almost sure convergence of $\Delta'_r$ is concluded from Theorem A.1 and the Kronecker's lemma.

STEP 4: We prove that under assumptions (A1), (A2), (A4.ii), (A5), (A7) and (A8), we have

$$h_n^{-j} t_{n,j}^{\star} \longrightarrow 0, \text{ almost surely as } n \to \infty, \text{ for } j = 0, \ldots, p,$$

and therefore we can write

$$H_{(n)}^{-1} \vec{T}_{(n)}^{\star} \longrightarrow 0, \text{ almost surely as } n \to \infty.$$

Note that

$$\frac{1}{h_n^j} t_{n,j}^* = \Delta''_r = \frac{1}{a'_{n,j}} \sum_{t=1}^{n} a'_{t,j} Z''_{t,j}.$$

where $a'_{t,j}$ is given as in (A.20) and $Z''_{t,j} = [V''_{t,j} - E(V''_{t,j})]h_t^{\eta}/a'_{t,j}$, being

$$V''_{t,j} = (X_t - x)^j K_t (X_t - x) [Y_t - m(X_t)]. \qquad (A.22)$$

It is clear that conditioning in (A.22) on $Y_t$ we find, for $\delta > 2$,

$$E\left(\left|V''_{t,j}\right|^{\delta}\right) = \int |x_t - x|^{\delta j} K_t^{\delta}(x_t - x) \left[\int |y_t - m(x_t)|^{\delta} f_{Y_t}(y_t) \, dy_t\right] \cdot$$

$$f_{X_t|Y_t}(x_t|y_t) \, dx_t.$$

Therefore, in this case, an asymptotic bound for $[E(|Z_{t,j}|^{\delta})]^{2/\delta}$ may be derived as follows.

Since $K$ has compact support and $m$ is continuous, there exists a constant $C_t$ such that $C_t = \sup_{|x_t - x| \leq h_t} |m(x_t)|$. This fact, together with assumption (A4.ii), allows us to conclude that

$$E\left(\left|V''_{t,j}\right|^{\delta}\right) \leq \int h_t^{\delta j} |v|^{\delta j} \frac{1}{h_t^{\delta-1}} K^{\delta}(v) \left[\int (|y_t| + C_t)^{\delta} f_{Y_t}(y_t) \, dy_t\right] dv$$

$$\leq \frac{h_t^{\delta j}}{h_t^{\delta-1}} \left(\int |v|^{\delta j} K^{\delta}(v) \, dv\right) E\left(|Y_t|^{\delta}\right).$$

Finally, assumptions (A1) and (A5) lead to

$$
E\left(\left|V_{t,j}''\right|^{\delta}\right) = O\left(h_t^{(j-1)\delta+1}\right) \quad \text{and} \quad \left[E\left(\left|Z_{t,j}''\right|^{\delta}\right)\right]^{2/\delta} \leq \text{cst.}\left(\frac{h_t^{1/\delta}}{th_t}\right)^2
\tag{A.23}
$$

and the result is achieved by using (A.23) in the same way as in the two previous steps.

STEP 5: The decomposition of the error $\widehat{\vec{\beta}}_{(n)} - \vec{\beta}$ given in equation (A.7) allows us to write

$$
\begin{aligned}
H_{(n)}\left(\widehat{\vec{\beta}}_{(n)} - \vec{\beta}\right) &= \left(H_{(n)}^{-1}S_{(n)}H_{(n)}^{-1}\right)^{-1}\left(H_{(n)}^{-1}\vec{T}_{(n)}^{\star}\right) \\
&+ \frac{m^{(p+1)}(x)}{(p+1)!}\left(H_{(n)}^{-1}S_{(n)}H_{(n)}^{-1}\right)^{-1}\left(\frac{1}{h_n^{p+1}}H_{(n)}^{-1}\vec{B}_{(n)}\right)h_n^{p+1} \\
&+ \left(H_{(n)}^{-1}S_{(n)}H_{(n)}^{-1}\right)^{-1}\left(\frac{1}{h_n^{p+1}}H_{(n)}^{-1}\vec{R}_{(n)}\right)h_n^{p+1} \\
&= \Gamma_1 + \Gamma_2 + \Gamma_3.
\end{aligned}
\tag{A.24}
$$

From the convergence results derived in steps 2, 3 and 4, we obtain that $\Gamma_1 = o(1)$, $\Gamma_2 = O(h_n^{p+1})$ and $\Gamma_3 = o(h_n^{p+1})$. The strong consistency for the estimator $\vec{\beta}_{(n)}$ is therefore established.

## A.2    Proof of Theorem 4.2

Let

$$
\phi_n = \left(\frac{nh_n^{2(1-1/\delta)}}{\left(\left(\log n\right)\left(\log\log n\right)^{1+\varepsilon}\right)}\right)^{1/2}.
$$

The sequence of real numbers $\phi_n$ is such that $\phi_n > 0$ and $\phi_n \uparrow \infty$. Therefore, if it is proved that $\phi_n\Gamma_i = o(1)$ a.s. for $i = 1, 2, 3$ then (4.2) follows from equation (A.24). But, in view of the found orders for $\Gamma_1$, $\Gamma_2$ and $\Gamma_3$ in the above proof, it is enough to show that

$$
\phi_n h_n^{-j} t_{n,j}^{\star} \longrightarrow 0, \text{ almost surely as } n \to \infty, \text{ for } j = 0, \ldots, p.
$$

For $j = 0, \ldots, p$ and $t = 1, 2, \ldots$ let us put

$$\phi_n h_n^{-j} t_{n,j}^{\star} = \frac{1}{\psi_{n,j}} \sum_{t=1}^{n} \psi_{t,j} W_{t,j},$$

where

$$\psi_{t,j} = \frac{h_t^{j} H_t(\eta)}{\phi_t}$$

and

$$W_{t,j} = \frac{\phi_t h_t^{\eta}}{H_t(\eta)} U_{t,j} \text{ with } U_{t,j} = \left(\frac{X_t - x}{h_t}\right)^{j} K_t(X_t - x)(Y_t - m(X_t)).$$

Proceeding in a similar way to that employed to bound $V_{t,j}''$ in step 4 of the proof of Theorem 4.1, it is obtained that $E(|U_{t,j}|^{\delta}) = o(h_t^{1-\delta})$. Therefore

$$\left[E\left(|W_{t,j}|^{\delta}\right)\right]^{2/\delta} \leq \text{cst.} \left(\frac{1}{t(\log t)(\log\log t)^{1+\varepsilon}}\right)^2. \qquad \text{(A.25)}$$

From (A.25) and (A9) it is deduced that the sequence of random variables $\{W_{t,j}\}_t$ for $j = 0, 1, \ldots, p$, satisfies (A.8) and (A.9) of Theorem A.1. On the other hand, Kronecker's lemma is applicable since $\psi_{n,j} > \text{cst.} \, n^{1/2} h_n^{j+1/2}$ and so $\psi_{n,j}$ tends to infinity by assumption (A7). Thus, the almost sure convergence of $\{\phi_n h_n^{-j} t_{n,j}^{\star}\}$ follows from Theorem A.1 and Kronecker's lemma and the proof is therefore complete.

### References

Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, **74**, 829-836.

Deheuvels, P. (1974). Conditions nécessaires et sufissantes de convergence ponetuelle presque sûre et uniforme presque sûre des estimateurs de la densité. *C. R. Acad. Sci. Paris. Sér. A.*, **278**, 1217-1220.

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

Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, **87**, 998-1004.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiency. *Annals of Statistics*, **21**, 196-216.

Fan, J. and I. Gijbels (1992). Variable bandwidth and local linear regression smoothers. *Annals of Statistics*, **20**, 2008-2036.

Fan, J. and I. Gijbels (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaption. *Journal of the Royal Statistical Society, B*, **57**, 2, 371-394.

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

Greblicky, W. and M. Pawlak (1987). Necessary and sufficient consistency conditions for a recursive kernel regression estimate. *Journal of the Multivariate Analisys*, **23**, 67-76.

Hastie, T. and C. Loader (1993). Local regression: automatic kernel carpentry. *Statistical Science*, **8**, 2, 120-143.

Krzyzak, A. and M. Pawlak (1984). Almost everywhere convergence of recursive regression function estimate and classification. *IEEE Transactions in Information Theory*, **30**, 91-93.

Lejeune, M. (1985). Estimation non-paramétrique par noyaux: régression polynomiale mobile. *Revue de Statist. Appliq.*, **33**, 43-68.

Masry, E. (1987). Almost sure convergence of recursive density estimators for stationary mixing processes. *Statistics and Probability Letters*, **5**, 249-254.

Masry, E. (1996a). Multivariate regression estimation: local polynomial fitting for time series. *Stochastic Processes and Applications*, **65**, 81-101.

Masry, E. (1996b). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *Journal of Time Series Analisys*, **17**, 571-599.

Masry, E. and J. Fan (1997). Local polynomial estimation of regression function for mixing processes. *Scandinavian Journal of Statistics*, **24**, 165-179.

McLeish, D.L. (1975). A maximal inequality and dependent strong law. *Annals of Probability*, **3**, 829-839.

Revesz, P. (1977). How to apply the method of stochastic approximation in the nonparametric estimation of a regression function. *Mathematische Operationsforschung, Serie Statistics*, **8**, 119-126.

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 L.T. Tran (1992). Asymptotic normality of the recursive kernel regression estimate under dependence conditions. *Annals of Statistics*, **20**, 1, 98-120.

Ruppert, D. and P. Wand (1994). Multivariate locally weighted least squares regression. *Annals of Statistics*, **22**, 1346-1370.

Ruppert, D., S.J. Sheather and M.P. Wand (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, **90**, 1257-1270.

Stone, C.J. (1977). Consistent nonparametric regression. *Annals of Statistics*, **5**, 595-620.

Tsybakov, A.B. (1990). Locally-polynomial algorithms of passive stochastic approximation. *Problems of Control and Information Theory*, 19, 3, 181-195.

Vilar, J.M. and Vilar, J.A. (1993). Estimación no paramétrica, recursiva, de curvas. *Estadística Española*, **35**, 134, 579-616.

Vilar, J.A. and Vilar, J.M. (1998). Recursive estimation of regression functions by local polynomial fitting. *Annals of the Institute of Statistical Mathematics*, **50**, 4, 729-754.