ON THE UNIFORM STRONG CONSISTENCY OF LOCAL POLYNOMIAL REGRESSION UNDER DEPENDENCE CONDITIONS

Francisco-Fernández, M., Vilar-Fernández J.M. and Vilar-Fernández, J.A.

Departamento de Matemáticas, Facultad de Informática

Universidad de A Coruña 15071, SPAIN

mariofr@udc.es, eijvilar@udc.es and eijoseba@udc.es

Key Words: nonparametric regression estimation; local polynomial fitting; dependent data; strong consistency.

ABSTRACT

In this paper, nonparametric estimators of the regression function, and its derivatives, obtained by means of weighted local polynomial fitting are studied. Consider the fixed regression model where the error random variables are coming from a stationary stochastic process satisfying a mixing condition. Uniform strong consistency, along with rates, are established for these estimators.

Furthermore, when the errors follow an AR(1) correlation structure, strong consistency

properties are also derived for a modified version of the local polynomial estimators proposed by Vilar-Fernández and Francisco-Fernández in (1).

1. INTRODUCTION AND DEFINITIONS

Let us consider the fixed regression model where the functional relationship between the design points, $x_{t,n}$, and the responses, $Y_{t,n}$, can be expressed as

$$Y_{t,n} = m(x_{t,n}) + \varepsilon_{t,n}, \quad 1 \leq t \leq n, \tag{1.1}$$

where $m(x)$ is a regression function and $\varepsilon_{t,n}$, $1 \leq t \leq n$, is a sequence of unobserved random variables with zero mean and finite variance σ^2. For each n, it is assumed that $\varepsilon_{1,n}, \varepsilon_{2,n}, \ldots, \varepsilon_{n,n}$ have the same joint distribution as $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$, where $\{\epsilon_t,\, t \in Z\}$ is a strictly stationary stochastic process. Finally, it is considered that the design points, $x_{t,n}$, $1 \leq t \leq n$, follow a regular design generated by a design density f; that is, for each n, the design points are defined by

$$\int_0^{x_{t,n}} f(x)d(x) = \frac{t-1}{n-1}, \quad 1 \leq t \leq n, \tag{1.2}$$

f being a positive function (see (2)).

In this setting, the goal is to estimate the regression function, $m(x) = E\left(Y|X = x\right)$, and its derivatives when the errors are dependent, satisfying the strong mixing condition. To do this, the local polynomial regression (LPR) estimator will be considered in the present study. Assuming that the $(p+1)$th derivative of the regression function at point x exists, then $\vec{\beta}(x) = (\beta_0(x), \beta_1(x), \cdots, \beta_p(x))^t$, where $\beta_j(x) = m^{(j)}(x)/(j!)$, with $j = 0, 1, \ldots, p$, can be estimated by minimizing the function

$$\Psi(\vec{\beta}(x)) = \sum_{t=1}^{n} \left(Y_{t,n} - \sum_{j=0}^{p} \beta_j(x)(x_{t,n} - x)^j \right)^2 \omega_{t,n}, \tag{1.3}$$

where $\omega_{t,n} = n^{-1}K_n(x_{t,n} - x)$ are the weights and $K_n(u) = h_n^{-1}K\left(h_n^{-1}u\right)$, K being a kernel function and h_n the bandwidth or smoothing parameter that controls the size of the local neighborhood and so the amount of smoothing.

Now, commonly used matrix notation is introduced for concise presentation of results. Denote

$$\vec{Y}_{(n)} = \begin{pmatrix} Y_{1,n} \\ \vdots \\ Y_{n,n} \end{pmatrix}, \qquad X_{(n)} = \begin{pmatrix} 1 & (x_{1,n} - x) & \cdots & (x_{1,n} - x)^p \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (x_{n,n} - x) & \cdots & (x_{n,n} - x)^p \end{pmatrix},$$

and let $W_{(n)} = \operatorname{diag}\left(\omega_{1,n}, \ldots, \omega_{n,n}\right)$ be the diagonal array of weights. Then, by assuming the invertibility of $X_{(n)}^t W_{(n)} X_{(n)}$, the LPR estimator of $\vec{\beta}(x)$ is given by

$$\hat{\vec{\beta}}_{(n)}(x) = \left(X_{(n)}^t W_{(n)} X_{(n)}\right)^{-1} X_{(n)}^t W_{(n)} \vec{Y}_{(n)} = S_{(n)}^{-1}(x) \vec{T}_{(n)}(x), \tag{1.4}$$

where $S_{(n)}(x)$ is the $(p+1) \times (p+1)$ array whose $(i+1, j+1)$th element is $s_{i,j,n}(x) = s_{i+j,n}(x),\ i, j = 0, 1, \ldots, p$, with

$$s_{j,n}(x) = \frac{1}{n} \sum_{t=1}^{n} (x_{t,n} - x)^j K_n\left(x_{t,n} - x\right), \quad 0 \leq j \leq 2p, \tag{1.5}$$

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

$$t_{j,n}(x) = \frac{1}{n} \sum_{t=1}^{n} (x_{t,n} - x)^j K_n\left(x_{t,n} - x\right) Y_{t,n}, \quad 0 \leq j \leq p. \tag{1.6}$$

So, the LPR estimator of $m^{(j)}(x)$ is given by $\hat{m}_n^{(j)}(x) = j!\hat{\beta}_j(x)$, $\hat{\beta}_j(x)$ being the jth component of $\hat{\vec{\beta}}_{(n)}(x)$, $j = 0, 1, \ldots, p$.

Since early papers on LPR, (3) and (4), many other relevant papers on this smoothing method have appeared, showing that the LPR estimator presents several good properties. See, for example, (5), (6), (7), (8), (9), and the references within. In these papers the independence of the observations is assumed. The statistical properties of LPR estimator with dependent data have been studied in recent works ((10), (11), (12), (13), (14), (15) and (16)). In these works the regression model with random design is considered and the assumption of the data satisfying some mixing condition is used. In the present context of fixed design, the asymptotic normality of the LPR estimator was studied in (17) when the random error, ε_t, has absolutely summable autocovariances. A complete study of this smoothing method can also be found in monograph (18).

The present paper is devoted to establishing strong uniform consistency and obtaining sharp rates of almost sure convergence over a compact set of R of two estimators of $m(x)$ and its derivatives. The first of these estimators is the LPR estimator given in (1.4), considering that the random errors satisfy an $\alpha-$mixing condition. The second estimator considered is a modified version of the LPR estimator, which has been proposed by Vilar-Fernández and Francisco-Fernández in (1). This last estimator is studied in the particular case where the stochastic process ϵ_t follows an AR(1) type correlation structure

$$\epsilon_t = \rho\epsilon_{t-1} + e_t, \ t \in \mathsf{Z}, \tag{1.7}$$

with $|\rho| < 1$ and $\{e_t\}_{t\in\mathsf{Z}}$, a noise process with mean zero and finite variance σ_e^2. The proposed estimator is obtained in two steps. In the first step, a matrix $P_{(n)}$, defined by

$$P_{(n)} = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 \\ -\rho & 1 & 0 & \dots & 0 \\ 0 & -\rho & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & -\rho & 1 \end{pmatrix},$$

in the case of AR(1) errors, is used to transform the regression model and to get a new model where the errors are uncorrelated.

For this purpose, performing a Taylor series expansion, the regression model (1.1) can be approximated by

$$\vec{Y}_{(n)} \approx X_{(n)}\vec{\beta}(x) + \vec{\varepsilon}_{(n)}, \tag{1.8}$$

where $\vec{\varepsilon}_{(n)} = (\varepsilon_{1,n}, \varepsilon_{2,n}, \dots, \varepsilon_{n,n})$.

Then, the errors of the following regression model

$$P_{(n)}\vec{Y}_{(n)} = P_{(n)}X_{(n)}\vec{\beta}(x) + P_{(n)}\vec{\varepsilon}_{(n)} \tag{1.9}$$

are uncorrelated.

Now, in the second step, assuming that $X^t_{(n)}P^t_{(n)}W_{(n)}P_{(n)}X_{(n)}$ is nonsingular, an estimator of $\vec{\beta}(x)$, generalized least squares estimator (GLPR), is obtained by using weighted least squares

$$\tilde{\vec{\beta}}_{G,n}(x) = \left(X^t_{(n)}P^t_{(n)}W_{(n)}P_{(n)}X_{(n)}\right)^{-1} X^t_{(n)}P^t_{(n)}W_{(n)}P_{(n)}\vec{Y}_{(n)} = \tilde{C}^{-1}_{(n)}(x)\tilde{\vec{G}}_{(n)}(x), \tag{1.10}$$

where $\tilde{C}_{(n)}(x) = X^t_{(n)}P^t_{(n)}W_{(n)}P_{(n)}X_{(n)}$ and $\tilde{\vec{G}}_{(n)}(x) = X^t_{(n)}P^t_{(n)}W_{(n)}P_{(n)}\vec{Y}_{(n)}$.

As matrix $P_{(n)}$ is unknown, the new estimator of $\vec{\beta}(x)$ is obtained by changing $P_{(n)}$ to an estimator of it, $\hat{P}_{(n)}$. This new estimator is called feasible least squares estimator (FLPR) and it is given by

$$\hat{\vec{\beta}}_{F,n}(x) = \left(X^t_{(n)}\hat{P}^t_{(n)}W_{(n)}\hat{P}_{(n)}X_{(n)}\right)^{-1} X^t_{(n)}\hat{P}^t_{(n)}W_{(n)}\hat{P}_{(n)}\vec{Y}_{(n)} = \hat{C}^{-1}_{(n)}(x)\hat{\vec{G}}_{(n)}(x), \tag{1.11}$$

where $\hat{C}^{-1}_{(n)}(x)$ is assumed to exist.

In the case of AR(1) errors considered here, the matrix $P_{(n)}$ is estimated on the basis of a previous consistent estimation of ρ. In (1), ρ was estimated by

$$\hat{\rho}_n = \frac{\sum_{t=1}^{n-1} \hat{\varepsilon}_{t,n}\,\hat{\varepsilon}_{t+1,n}}{\sum_{t=1}^{n} \hat{\varepsilon}^2_{t,n}}, \tag{1.12}$$

where $\hat{\varepsilon}_{t,n} = Y_{t,n} - \hat{m}_n(x_{t,n})$, $1 \leq t \leq n$, are nonparametric residuals and $\hat{m}_n(x)$ is a consistent estimator of $m(x)$, for example, the LPR estimator. A natural estimator for $P_{(n)}$ is then obtained by replacing ρ with $\hat{\rho}_n$.

Vilar-Fernández and Francisco-Fernández in (1) proved that the estimators $\hat{\vec{\beta}}_{(n)}(x)$ and $\hat{\vec{\beta}}_{F,n}(x)$ have the same asymptotic distribution. However, in a simulation study a better behavior was observed for the mean integrated squared error of estimator $\hat{\vec{\beta}}_{F,n}(x)$ with respect to $\hat{\vec{\beta}}_{(n)}(x)$ when correlation of the observations was large.

Before continuing, it is worth mentioning some works concerned with the study of strong consistency properties for kernel nonparametric estimators of the regression function under dependence conditions. For the regression model with random design, Collomb and Härdle in (19) obtained the strong uniform convergence of a robust nonparametric estimator under $\phi-$mixing conditions, Gyorfi et al. in (20) and Roussas in (21) proved strong consistency of the Nadaraya-Watson kernel estimator under several mixing conditions, Troung and Stone in (22) obtained weak convergence rates under α-mixing assumption and Masry and Tjøstheim in (23) established strong convergence rates and asymptotic normality under $\alpha-$mixing conditions. Recently, Lu and Cheng in (24) proved the distribution-free strong consistency under α-mixing and quite mild conditions, Ango Nze and Douckan in (25) considered delta-sequence estimators and established uniform convergence in the mean and almost surely under $\alpha-$mixing conditions and absolute regularity. In (26), Masry employed the local polynomial fitting for the estimation of the multivariate regression function and obtained uniform strong consistency with rates for strong mixing processes. Vilar-Fernández and Vilar-Fernández in (16) studied a recursive local polynomial smoother under α-mixing dependence and established properties of strong consistency.

In the case of nonparametric regression with fixed design, Roussas in (27) studied a general linear smoother of the regression function and obtained consistency in quadratic mean and strong consistency under several mixing conditions of the errors. Roussas et al. in (28) established the asymptotic normality of this estimator when the errors are $\alpha-$mixing and in (29), Tran et al. generalized this result dispensing of mixing assumptions and encompassing models with discrete noise.

The organization of this paper is as follows: in Section 2, the uniform strong convergence of the estimator $\hat{\vec{\beta}}_{(n)}(x)$ over compact subsets of R is proven and the rates of convergence are established. In Section 3, the uniform strong convergence for the new estimator $\hat{\vec{\beta}}_{F,n}(x)$ is obtained. Finally, Section 4 is devoted to the proofs of the results.

2. UNIFORM STRONG CONSISTENCY OF THE LPR ESTIMATOR

To establish the uniform strong convergence of the LPR estimator, $\hat{\vec{\beta}}_{(n)}(x)$, defined in (1.4), a standard proof technique based on using a Berstein's type inequality for strongly mixing sequences joint to a coupling argument due to Rio (30) is considered. First of all, it is necessary to split the error estimation. For this purpose, assuming the continuity of the first $(p+1)$ derivatives of $m(x)$, a Taylor series expansion around x with an integral remainder is performed, so that

$$m(x_{t,n}) = \sum_{j=0}^{p} \frac{m^{(j)}(x)}{j!}(x_{t,n}-x)^j + \frac{m^{(p+1)}(x)}{(p+1)!}(x_{t,n}-x)^{p+1} + R_{t,n}(x), \ t=1,\ldots,n, \quad (2.1)$$

where

$$R_{t,n}(x) = \frac{(x_{t,n}-x)^{p+1}}{p!} \int_0^1 (1-w)^p \left[m^{(p+1)}(x+w(x_{t,n}-x)) - m^{(p+1)}(x)\right] dw. \quad (2.2)$$

In matrix form,

$$\vec{M}_{(n)} = X_{(n)}\vec{\beta}(x) + \frac{m^{(p+1)}(x)}{(p+1)!} \begin{pmatrix} (x_{1,n}-x)^{p+1} \\ \vdots \\ (x_{n,n}-x)^{p+1} \end{pmatrix} + \vec{R}_{(n)}(x), \quad (2.3)$$

with $\vec{M}_{(n)} = (m(x_{1,n}),\ldots,m(x_{n,n}))^t$ and $\vec{R}_{(n)}(x) = (R_{1,n}(x),\ldots,R_{n,n}(x))^t$.

Using (1.4) and (2.3), the following is obtained:

$$E\left(\hat{\vec{\beta}}_{(n)}(x)\right) = \vec{\beta}(x) + \frac{m^{(p+1)}(x)}{(p+1)!} S_{(n)}^{-1}(x)\vec{U}_{(n)}(x) + S_{(n)}^{-1}(x)\vec{V}_{(n)}(x), \quad (2.4)$$

where $\vec{U}_{(n)}(x)$ and $\vec{V}_{(n)}(x)$ are $(p+1)$-dimensional vectors whose $(j+1)$th components $U_{j,n}(x)$ and $V_{j,n}(x)$, $j = 0, 1, \ldots, p$, are given by

$$U_{j,n}(x) = s_{j+p+1,n}(x),$$

$$V_{j,n}(x) = \frac{s_{j+p+1,n}(x)}{p!} \int_0^1 (1-w)^p \left[m^{(p+1)}(x + w(x_{t,n} - x)) - m^{(p+1)}(x) \right] dw.$$

From (2.4), the following decomposition for the error

$$\begin{aligned} \hat{\vec{\beta}}_{(n)}(x) - \vec{\beta}(x) &= S_{(n)}^{-1}(x)\vec{T}^{\star}_{(n)}(x) + \frac{m^{(p+1)}(x)}{(p+1)!} S_{(n)}^{-1}(x)\vec{U}_{(n)}(x) + S_{(n)}^{-1}(x)\vec{V}_{(n)}(x) \\ &= \vec{\Delta}_{1,n}(x) + \vec{\Delta}_{2,n}(x) + \vec{\Delta}_{3,n}(x), \end{aligned} \tag{2.5}$$

is directly derived, where it has been denoted that

$$\vec{T}^{\star}_{(n)}(x) = \vec{T}_{(n)}(x) - E\left(\vec{T}_{(n)}(x)\right).$$

The vector $\vec{\Delta}_{1,n}(x)$ is random and the vectors $\vec{\Delta}_{2,n}(x)$ and $\vec{\Delta}_{3,n}(x)$ are deterministic. Now, to obtain the strong consistency of the LPR estimator, the convergence of these three vectors is studied next.

The following assumptions are required:

A.1. The kernel function $K(\cdot)$ is symmetric, positive, Lipschitz continuous and with a bounded support.

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

A.3. The matrix $S_{(n)}^{-1}(x)$ exists for $x \in A$, with A a compact subset of R, and the design density function f satisfies $0 < C \leq f(x) \leq C' < \infty$, with C and C' real numbers.

A.4. The function $m^{(p+1)}(x)$ is uniformly continuous on A.

A.5. $E\left(|\varepsilon_t|^{\delta}\right) < \infty$, for some $\delta > 2$.

A.6. The stationary process ε_t is α-mixing, with mixing coefficients $\alpha(k)$ such that

$$\sum_{k=1}^{\infty} k\alpha(k)^{1-2/\delta} < \infty.$$

A.7. Define the sequence $M_n = \left(n \ln n \left(\ln \ln n\right)^{1+\gamma}\right)^{1/\delta}$, for some $0 < \gamma < 1$. Then h_n is chosen in such a way that

$$\gamma_n = \left(\frac{nM_n^2}{h_n^3 \ln n}\right)^{1/2} \to \infty \quad \text{and} \quad b_n = \left(\frac{nh_n}{M_n^2 \ln n}\right)^{1/2} \to \infty \quad \text{as } n \to \infty.$$

In addition, the α-mixing sequence $\alpha(k)$ satisfies

$$\sum_{n=1}^{\infty} \frac{n\gamma_n}{b_n} \left(\frac{nM_n^2}{h_n \ln n}\right)^{1/2} \alpha\left(b_n\right) < \infty.$$

Assumptions A.1, A.2, A.3, A.4 and A.5 are not very restrictive regularity conditions and are quite usual in the context of local polynomial regression. While assumption A.6 is also a usual summability requirement on the mixing coefficients, assumption A.7 is a more complex condition involving both the bandwidth and the α-mixing sequence. It is imposed to determine an appropriate truncation sequence to obtain a precise block size when the Bernstein's block technique is employed. It has also been required by Masry in (26). It is quite usual to find complex conditions on the α-mixing coefficients in the literature concerning strong consistency. Interesting comments about this topic can be seen in Lu and Cheng in (24) (Remark 2.3).

In particular, it can be easily seen that if

$$h_n = O\left(\frac{\ln n}{n}\right)^{\phi}, \text{ with } \phi < 1 - 2/\delta,$$

then the conditions $\gamma_n \to \infty$ and $b_n \to \infty$ in assumption A.7 are satisfied. If, in addition, the mixing coefficients are assumed to decay exponentially fast, that is, $\alpha(k) = O\left(\exp\{-\theta k\}\right)$, then both assumption A.6 and the summability restriction in assumption A.7 are also satisfied for all $\theta > 0$.

Another mixing case of interest is $\alpha(k) = O\left(k^{-\theta}\right)$. In such a case, straightforward calculations allow us to conclude that the summability restriction in assumption A.7 is satisfied provided that

$$\theta > \frac{(1+\phi)5/2 + 3/\delta}{(1-\phi)1/2 - 1/\delta} = L\left(\phi, \delta\right).$$

Note that for fixed ϕ, the function $L\left(\phi, \delta\right)$ is monotonically decreasing in δ. Therefore, assumption A.5 and the summability condition in assumption A.7 move in opposite directions since the larger δ is chosen, the smaller θ can be selected.

In what follows $H_{(n)}$ denotes the diagonal array $\operatorname{diag}\left(1, h_n, h_n^2, \cdots, h_n^p\right)$ and S is the $(p+1) \times (p+1)$ array whose $(i+1, j+1)$th element is $s_{i,j} = \mu_{i+j}$, $i, j = 0, 1, \ldots, p$, where $\mu_j = \int u^j K(u)\, du$ and $\vec{\mu} = \left(\mu_{p+1}, \ldots, \mu_{2p+1}\right)^t$.

The following result of Francisco-Fernández and Vilar-Fernández given in (17) will be used.

PROPOSITION 1. If assumptions A.1 and A.2 hold, then

$$\lim_{n\to\infty} h_n^{-j} s_{j,n}(x) = f(x)\mu_j, \quad 0 \le j \le 2p+1. \tag{2.6}$$

Result (2.6) can be written in matrix form as

$$\lim_{n\to\infty} H_{(n)}^{-1} S_{(n)}(x) H_{(n)}^{-1} = f(x)S. \tag{2.7}$$

Proposition 1 is then used to obtain the uniform convergence for the deterministic terms $\vec{\Delta}_{2,n}(x)$ and $\vec{\Delta}_{3,n}(x)$.

PROPOSITION 2. If assumptions A.1-A.4 hold, then

$$\sup_{x\in A} H_{(n)}\vec{\Delta}_{2,n}(x) = h_n^{p+1}\left(\sup_{x\in A} m^{(p+1)}(x)\right)\frac{1}{(p+1)!}S^{-1}\vec{\mu}\,(1+o(1)) = O\left(h_n^{p+1}\right) \tag{2.8}$$

and

$$\sup_{x\in A} H_{(n)}\vec{\Delta}_{3,n}(x) = O\left(h_n^{p+2}\right). \tag{2.9}$$

Next, the strong consistency for the random term $\vec{\Delta}_{1,n}(x)$ is established.

PROPOSITION 3. If assumptions A.1-A.7 are fulfilled, then

$$\sup_{x\in A} H_{(n)}\vec{\Delta}_{1,n}(x) = O\left(\frac{\ln n}{nh_n}\right)^{1/2} \quad \text{almost sure.} \tag{2.10}$$

The proof of Proposition 3 is the most complex step. The basic tool is the Bernstein's blocks technique which consists in approximating mixing sequences by independent sequences, so that a Bernstein-type exponential inequality can be applied. After the blocks are determined, an independence approximation argument must be used. In this paper a coupling theorem due to Rio in (30) is adopted, which has also been used by others (see for example Carbon, Tran and and Wu in (31)). Alternatively, other authors have considered a weaker argument from Bradley (32).

The following theorem follows from equation (2.5) and Propositions 2 and 3.

THEOREM 1. If assumptions A.1-A.7 are fulfilled, then

$$\sup_{x\in A} H_{(n)}\left(\hat{\vec{\beta}}_{(n)}(x)-\vec{\beta}(x)\right) = O\left(h_n^{p+1}\right)+O\left(\frac{\ln n}{nh_n}\right)^{1/2} \text{ almost sure.} \tag{2.11}$$

In the next result, derived directly from Theorem 1, the uniform strong convergence and rates for the local polynomial estimator of the regression function and its derivatives are established.

COROLLARY 1. Under the hypothesis of Theorem 1, we have, for $j=0,1,\ldots,p$, that

$$\sup_{x\in A}\left(\hat{m}_n^{(j)}(x)-m^{(j)}(x)\right) = O\left(h_n^{p+1-j}\right)+O\left(\frac{\ln n}{nh_n^{1+2j}}\right)^{1/2} \text{ almost sure.} \tag{2.12}$$

3. UNIFORM STRONG CONSISTENCY OF THE FLPR ESTIMATOR

This section is devoted to establishing the uniform strong convergence of the FLPR estimator, $\hat{\vec{\beta}}_{F,n}(x)$, defined in (1.11). Therefore, throughout this section it will be assumed that the stochastic process ϵ_t follows an AR(1) model with $|\rho| < 1$.

Our attention will first focus on $\tilde{\vec{\beta}}_{G,n}(x)$, the estimator defined in (1.10). In particular, conditions to ensure the uniform strong convergence of $\tilde{\vec{\beta}}_{G,n}(x)$ will be established.

Approximating the vector $\vec{M}_{(n)}$ with a Taylor series in a neighborhood of x and employing similar arguments as those used in the previous section, the following decomposition of the error is obtained:

$$\begin{aligned}\tilde{\vec{\beta}}_{G,n}(x) - \vec{\beta}(x) &= \tilde{C}^{-1}_{(n)}(x)\tilde{\vec{G}}^{*}_{(n)}(x) + \frac{m^{(p+1)}(x)}{(p+1)!}\tilde{C}^{-1}_{(n)}(x)\tilde{\vec{U}}_{(n)}(x) + \tilde{C}^{-1}_{(n)}(x)\tilde{\vec{V}}_{(n)}(x) \\ &= \tilde{\vec{\Delta}}_{1,n}(x) + \tilde{\vec{\Delta}}_{2,n}(x) + \tilde{\vec{\Delta}}_{3,n}(x), \qquad (3.1)\end{aligned}$$

where

$$\tilde{\vec{G}}^{*}_{(n)}(x) = \tilde{\vec{G}}_{(n)}(x) - E\left(\tilde{\vec{G}}_{(n)}(x)\right),$$

and $\tilde{\vec{U}}_{(n)}(x)$ and $\tilde{\vec{V}}_{(n)}(x)$ are $(p+1)$-dimensional vectors analogous to $\vec{U}_{(n)}(x)$ and $\vec{V}_{(n)}(x)$ in (2.4), respectively. Here, they are given by

$$\tilde{\vec{U}}_{(n)}(x) = X^{t}_{(n)}P^{t}_{(n)}W_{(n)}P_{(n)}\left((x_{1,n}-x)^{p+1}, \ldots, (x_{n,n}-x)^{p+1}\right)^{t}, \qquad (3.2)$$

and

$$\tilde{\vec{V}}_{(n)}(x) = X^{t}_{(n)}P^{t}_{(n)}W_{(n)}P_{(n)}\vec{R}_{(n)}(x), \qquad (3.3)$$

with $\vec{R}_{(n)}(x)$ given in (2.3).

Using basic algebra allows to derive the particular form for the $(j+1)$th component of $\tilde{\vec{U}}_{(n)}(x)$, for $j = 0, 1, \ldots, p$:

$$\begin{aligned}
\tilde{U}_{j,n}(x) &= \tilde{c}_{j,p+1,n}(x) \\
&= \frac{1}{n}\sum_{t=1}^{n}(x_{t,n}-x)^{j+p+1}K_n(x_{t,n}-x) - \frac{\rho^2}{n}(x_{1,n}-x)^{j+p+1}K_n(x_{1,n}-x) \\
&\quad -\frac{\rho}{n}\sum_{t=2}^{n}(x_{t,n}-x)^{j}(x_{t-1,n}-x)^{p+1}K_n(x_{t,n}-x) \\
&\quad -\frac{\rho}{n}\sum_{t=2}^{n}(x_{t,n}-x)^{p+1}(x_{t-1,n}-x)^{j}K_n(x_{t,n}-x) \\
&\quad +\frac{\rho^2}{n}\sum_{t=2}^{n}(x_{t-1,n}-x)^{j+p+1}K_h(x_{t,n}-x).
\end{aligned}$$

To show the convergence to zero of terms $\tilde{\vec{\Delta}}_{i,n}(x)$, $i = 1, 2, 3$, assumptions A.3, A.4 and A.7 are modified as follows:

A.3′. The matrix $\tilde{C}^{-1}_{(n)}(x)$ exists for $x \in A$, with A a compact subset of R, and the design density f verifies $0 < C \leq f(x) < C'$, with C and C' real numbers.

A.4′. The functions f' and $m^{(p+1)}$ are continuous on A.

A.7′. Denote $M_n = \left(n \ln n \,(\ln\ln n)^{1+\gamma}\right)^{1/\delta}$ for some $0 < \gamma < 1$. The bandwidth h_n is such that the sequences

$$\gamma_n = \left(\frac{nM_n^2}{h_n^3 \ln n}\right)^{1/2} \to \infty \quad \text{and} \quad b_n = \left(\frac{nh_n}{M_n^2 \ln n}\right)^{1/2} \to \infty \quad \text{as } n \to \infty.$$

The next result, obtained by Vilar-Fernández and Francisco-Fernández in (1), is here used to show the convergence to zero of the deterministic terms $\tilde{\vec{\Delta}}_{2,n}(x)$ and $\tilde{\vec{\Delta}}_{3,n}(x)$.

PROPOSITION 4. If assumptions A.1, A.2, A.3′ and A.4′ hold, then

$$\lim_{n\to\infty} h_n^{-(i+j)} \tilde{c}_{i,j,n}(x) = (1-\rho)^2 f(x)\mu_{i+j}, \quad 0 \le i,j \le p+1. \tag{3.4}$$

Expression (3.4) can be rewritten in matrix form as follows:

$$\lim_{n\to\infty} H_{(n)}^{-1} \tilde{C}_{(n)}(x) H_{(n)}^{-1} = (1-\rho)^2 f(x) S. \tag{3.5}$$

The strong consistency of both $\tilde{\vec{\Delta}}_{2,n}(x)$ and $\tilde{\vec{\Delta}}_{3,n}(x)$ is now a direct consequence of Proposition 4 and is established in the following result.

PROPOSITION 5. If assumptions A.1, A.2, A.3′ and A.4′ hold, then

$$\sup_{x\in A} H_{(n)} \tilde{\vec{\Delta}}_{2,n}(x) = O\left(h_n^{p+1}\right) \tag{3.6}$$

and

$$\sup_{x\in A} H_{(n)} \tilde{\vec{\Delta}}_{3,n}(x) = O\left(h_n^{p+2}\right). \tag{3.7}$$

The strong convergence of the random term $\tilde{\vec{\Delta}}_{1,n}(x)$ is established in the following Proposition.

PROPOSITION 6. If assumptions A.1, A.2, A.3′, A.4′, A.5 and A.7′ are fulfilled, then

$$\sup_{x\in A} H_{(n)} \tilde{\vec{\Delta}}_{1,n}(x) = O\left(\frac{\ln n}{n h_n}\right)^{1/2} \text{ almost sure.} \tag{3.8}$$

In accordance with (3.1), the uniform strong convergence of the estimator $\tilde{\vec{\beta}}_{G,n}(x)$ follows from Propositions 5 and 6.

THEOREM 2. Under assumptions of Proposition 6, it is verified that

$$\sup_{x\in A} H_{(n)}\left(\tilde{\vec{\beta}}_{G,n}(x) - \vec{\beta}(x)\right) = O\left(h_n^{p+1}\right) + O\left(\frac{\ln n}{nh_n}\right)^{1/2} \quad \text{almost sure.} \tag{3.9}$$

On the other hand, on the basis of the convergence in probability of the estimator $\hat{\rho}_n$, defined in (1.12), to the autoregressive coefficient ρ, the uniform strong convergence to zero of the term $H_{(n)}\left(\hat{\vec{\beta}}_{F,n}(x) - \tilde{\vec{\beta}}_{G,n}(x)\right)$ is obtained in the next result.

THEOREM 3. If assumptions A.1, A.2, A.3′ and A.4′ hold, then

$$\sup_{x\in A} H_{(n)}\left(\hat{\vec{\beta}}_{F,n}(x) - \tilde{\vec{\beta}}_{G,n}(x)\right) = o\left(h_n^{p+1}\right) + o\left(\frac{\ln n}{nh_n}\right)^{1/2} \quad \text{almost sure.} \tag{3.10}$$

From both, Theorems 2 and 3, the uniform strong convergence of the FLPR estimator $\hat{\beta}_{F,n}(x)$ is deduced.

THEOREM 4. Under assumptions of Theorem 2 it is verified that

$$\sup_{x\in A} H_{(n)}\left(\hat{\vec{\beta}}_{F,n}(x) - \vec{\beta}(x)\right) = O\left(h_n^{p+1}\right) + O\left(\frac{\ln n}{nh_n}\right)^{1/2} \quad \text{almost sure.} \tag{3.11}$$

If $\hat{m}_{F,n}^{(j)}(x)$ denotes the nonparametric feasible least squares estimator of $m^{(j)}(x)$, for $j = 0, 1, \ldots, p$, then $\hat{m}_{F,n}^{(j)}(x) = (j!)\,\hat{\beta}_{j,F,n}(x)$, being

$$\hat{\vec{\beta}}_{F,n}(x) = \left(\hat{\beta}_{0,F,n}(x), \hat{\beta}_{1,F,n}(x), \ldots, \hat{\beta}_{p,F,n}(x)\right)^t.$$

Result (3.11) can be then reformulated in terms of $\hat{m}_{F,n}^{(j)}(x)$, as shown in Corollary 2.

COROLLARY 2. Under assumptions of Theorem 4, it is verified that

$$\sup_{x \in A} \left(\hat{m}_{F,n}^{(j)}(x) - m^{(j)}(x)\right) = O\left(h_n^{p+1-j}\right) + O\left(\frac{\ln n}{n h_n^{1+2j}}\right)^{1/2} \text{ almost sure.} \tag{3.12}$$

The extension of these results to regression models with more general correlation structures, for example, $ARMA(p,q)$ models, is conceptually straightforward but with the drawback that the $P_{(n)}$ matrix depends on more parameters and these need to be estimated.

4. PROOFS

Throughout this section, the proofs of the results presented in sections 2 and 3 are outlined.

PROOF OF PROPOSITION 2.

According to (2.5),

$$\begin{aligned}\vec{\Delta}_{2,n}(x) &= \frac{m^{(p+1)}(x)}{(p+1)!} S_{(n)}^{-1}(x) \vec{U}_{(n)}(x) \\ &= \frac{m^{(p+1)}(x)}{(p+1)!} h_n^{p+1} H_{(n)}^{-1} \left(H_{(n)}^{-1} S_{(n)}(x) H_{(n)}^{-1}\right)^{-1} \left(\frac{1}{h_n^{p+1}} H_{(n)}^{-1} \vec{U}_{(n)}(x)\right).\end{aligned}$$

Then, using Proposition 1 it is obtained that

$$\vec{\Delta}_{2,n}(x) = h_n^{p+1} \frac{m^{(p+1)}(x)}{(p+1)!} H_{(n)}^{-1} S^{-1} \vec{\mu} \left(1 + o(1)\right)$$

and thus (2.8) is derived from the boundedness of $m^{(p+1)}(x)$.

The proof of (2.9) follows for similar arguments and it has been omitted.

PROOF OF PROPOSITION 3.

From (2.5)

$$H_{(n)} \vec{\Delta}_{1,n}(x) = \left(H_{(n)}^{-1} S_{(n)}(x) H_{(n)}^{-1} \right)^{-1} H_{(n)}^{-1} \vec{T}_{(n)}^{\star}(x).$$

The limit of $H_{(n)}^{-1} S_{(n)}(x) H_{(n)}^{-1}$ is given in (2.7), so that, it is sufficient to establish the almost sure convergence of $H_{(n)}^{-1} \vec{T}_{(n)}^{\star}(x)$, that is, of its components $h_n^{-j} t_{j,n}^{\star}(x)$, for $j = 0, 1, \ldots, p$.

Let $\{M_n\}$ be the sequence of positive numbers defined in A.7 and let

$$\varepsilon_{t,M_n} = \varepsilon_{t,n} \; I \left(|\varepsilon_{t,n}| \leq M_n \right). \tag{4.1}$$

Replacing ε_t with ε_{t,M_n}, new terms $t_{j,n}^{\star B}(x)$ are defined by

$$t_{j,n}^{\star B}(x) = \frac{1}{n} \sum_{t=1}^{n} \left(x_{t,n} - x \right)^j K_n \left(x_{t,n} - x \right) \varepsilon_{t,M_n}, \quad 0 \leq j \leq p, \tag{4.2}$$

so that, for $j = 0, 1, \ldots, p$,

$$\sup_{x \in A} h_n^{-j} t_{j,n}^{\star}(x) \leq \sup_{x \in A} h_n^{-j} \left(t_{j,n}^{\star}(x) - t_{j,n}^{\star B}(x) \right) + \sup_{x \in A} h_n^{-j} t_{j,n}^{\star B}(x) = Q_{j,1,n} + Q_{j,2,n}.$$

Now, each term on the right-hand side above is examined. As far as term $Q_{j,1,n}$ is concerned, it is observed that

$$h_n^{-j} \left(t_{j,n}^{\star}(x) - t_{j,n}^{\star B}(x) \right) = \frac{1}{n} \sum_{t=1}^{n} \varepsilon_{t,n} I \left(|\varepsilon_{t,n}| > M_n \right) h_n^{-j} \left(x_{t,n} - x \right)^j K_n \left(x_{t,n} - x \right).$$

Under A.5 and using Markov's inequality, it is obtained that

$$P\left(|\varepsilon_{n,n}| > M_n\right) \leq M_n^{-\delta} E\left(|\varepsilon_{n,n}|^{\delta}\right) < \infty,$$

for sufficiently large n. In addition, Proposition 1 ensures the summability of the factor on the right-hand side. Therefore, Borel-Cantelli Lemma gives that $|\varepsilon_{n,n}| \leq M_n$ almost surely for all sufficiently large n. On the other hand, since M_n is increasing, there exists n such that $|\varepsilon_{t,n}| \leq M_n$ almost surely for $t \leq n$. From the above conclusions follows the almost sure convergence to zero of $Q_{j,1,n}$, for $j = 0, 1, \ldots, p$.

Attention is now concentrated on the term $Q_{j,2,n}$. Since A is compact, it can be covered with γ_n intervals of length $2l_n$ and center $x'_{k,n}$. Denote $I_{k,n} = \left[x'_{k,n} - l_n, x'_{k,n} + l_n\right]$, $1 \leq k \leq \gamma_n$, with $\gamma_n \to \infty$ as $n \to \infty$. Note that necessarily $l_n = O\left(\gamma_n^{-1}\right)$. Then

$$Q_{j,2,n} = \sup_{x \in A} h_n^{-j} t_{j,n}^{\star B}(x)$$

$$\leq \max_{1 \leq k \leq \gamma_n} \sup_{x \in A \cap I_{k,n}} h_n^{-j} \left|t_{j,n}^{\star B}(x) - t_{j,n}^{\star B}(x'_{k,n})\right| + \max_{1 \leq k \leq \gamma_n} h_n^{-j} \left|t_{j,n}^{\star B}(x'_{k,n})\right| = P_{j,1,n} + P_{j,2,n}. \quad (4.3)$$

Under assumption A.1, for each $x \in I_{k,n}$ one has

$$h_n^{-j} \left|t_{j,n}^{\star B}(x) - t_{j,n}^{\star B}(x'_{k,n})\right| \leq C_j \frac{l_n M_n}{h_n^2} = C \frac{M_n}{\gamma_n h_n^2}, \quad (4.4)$$

where C is a positive real number. Taking into account (4.3), (4.4) and the definitions of the sequences M_n and γ_n given in A.7, it is concluded that

$$P_{j,1,n} \leq C \left(\frac{\ln n}{n h_n}\right)^{1/2} \to 0 \quad \text{as } n \to \infty. \quad (4.5)$$

Thus, only the following remains to be proven:

$$P_{j,2,n} = O\left(\frac{\ln n}{n h_n}\right)^{1/2} \text{ almost sure.} \quad (4.6)$$

The two following lemmas are necessary to show (4.6). The first of these lemmas is a coupling theorem for strongly mixing real-valued random variables.

LEMMA 1. (Theorem 4 of (30)) Let $\mathcal{A}$ be a $\sigma-$field of (Ω, z, P) and let X be a real-valued random variable taking almost sure values in $[a, b]$. Suppose furthermore that there exists a random variable β with uniform distribution over $[0, 1]$, independent of $\mathcal{A} \vee \sigma(X)$. Then, there exists some random variable X^* independent of $\mathcal{A}$ and with the same distribution as X such that

$$E\left(|X - X^*|\right) \leq 2(b-a)\alpha\left(\mathcal{A}, \sigma(X)\right).$$

Moreover, X^* is a $\mathcal{A} \vee \sigma(X) \vee \sigma(\beta)-$measurable random variable.

The second lemma is the Bernstein inequality. The proof can be seen in (33).

LEMMA 2. Let $Y_1, Y_2, \ldots, Y_n$ be independent bounded random variables with zero mean and $|Y_i| \leq M$. Denoting σ_i^2 as the variance of Y_i, and supposing $\sum_{i=1}^n \sigma_i^2 \leq V$. Then, for each $\eta > 0$,

$$P\left(\left|\sum_{i=1}^n Y_i\right| \geq \eta\right) \leq 2\exp\left(-\frac{1}{2}\eta^2 \Big/ \left(V + \frac{1}{3}M\eta\right)\right).$$

From (4.2), $h_n^{-j} t_{j,n}^{\star B}(x)$ can be written in the form

$$h_n^{-j} t_{j,n}^{\star B}(x) = \frac{1}{n}\sum_{t=1}^n \xi_{t,n}(x) = \Gamma_n(x), \tag{4.7}$$

where

$$\xi_{t,n}(x) = \left(\frac{x_{t,n} - x}{h_n}\right)^j K_n\left(x_{t,n} - x\right)\varepsilon_{t,M_n}. \tag{4.8}$$

The Bernstein's blocks technique is used next. Set $n = 2s_n b_n + v_n$, where s_n, b_n and v_n are integer numbers satisfying $s_n \to \infty$, $b_n \to \infty$ as $n \to \infty$ and $0 \leq v_n < b_n$. Then, $\Gamma_n(x)$ is split into $2s_n$ blocks of size b_n plus a residual block of size v_n as follows:

Define the blocks

$$B_{k,n}(x) = \frac{1}{n} \sum_{t=(k-1)b_n+1}^{kb_n} \xi_{t,n}(x), \quad k = 1, 2, \ldots, 2s_n. \tag{4.9}$$

So the partition of $\Gamma_n(x)$ is given by

$$\Gamma_n(x) = \frac{1}{n} \sum_{t=1}^{n} \xi_{t,n}(x) = \Gamma_{1,n}(x) + \Gamma_{2,n}(x) + \Gamma_{3,n}(x), \tag{4.10}$$

where

$$\Gamma_{1,n}(x) = \sum_{j=1}^{s_n} B_{2j-1,n}(x), \ \Gamma_{2,n}(x) = \sum_{j=1}^{s_n} B_{2j,n}(x) \text{ and } \Gamma_{3,n}(x) = \sum_{t=2s_n b_n+1}^{n} \xi_{t,n}(x). \tag{4.11}$$

Since the contribution of the residual term $\Gamma_{3,n}(x)$ is negligible, we have for each $\varepsilon > 0$

$$\begin{aligned} P\left(|P_{j,2,n}| > \varepsilon\right) &\leq P\left(\max_{1\leq k\leq \gamma_n} \left|\Gamma_{1,n}(x'_{k,n})\right| > \varepsilon/2\right) + P\left(\max_{1\leq k\leq \gamma_n} \left|\Gamma_{2,n}(x'_{k,n})\right| > \varepsilon/2\right) \\ &\leq 2P\left(\max_{1\leq k\leq \gamma_n} \left|\Gamma_{1,n}(x'_{k,n})\right| > \frac{\varepsilon}{2}\right) \leq 2\gamma_n \sup_{x\in A} P\left(\left|\Gamma_{1,n}(x)\right| > \frac{\varepsilon}{2}\right). \end{aligned} \tag{4.12}$$

To bound the last term in (4.12) the independence approximation argument given in Lemma 1 is used next.

Let $\{U_j\}_{j>0}$ be a sequence of independent random variables with uniform distribution over $[0,1]$, independent of $\{B_{2j-1,n}(x)\}_{j=1}^{s_n}$. By Lemma 1, for any positive j, there exists a measurable function F_j such that $B^*_{2j-1,n}(x) = F_j\left(B_{1,n}(x), B_{3,n}(x), \ldots, B_{2j-1,n}(x), U_j\right)$ satisfies the conditions of Lemma 1 (here $\mathcal{A} = \sigma\left(B_{2k-1,n}(x),\, 1 \leq k < j\right)$). Therefore, for each

j, $B^*_{2j-1,n}(x)$ is independent of $B^*_{1,n}(x), B^*_{3,n}(x), \ldots, B^*_{2j-3,n}(x)$, has the same distribution as $B_{2j-1,n}(x)$ and satisfies

$$E\left(\left|B_{2j-1,n}(x) - B^*_{2j-1,n}(x)\right|\right) \leq 2\left\|B_{2j-1,n}(x)\right\|_\infty \alpha\left(b_n\right). \tag{4.13}$$

From (4.9) and (4.11), it can be written that

$$P\left(\left|\Gamma_{1,n}(x)\right| > \frac{\varepsilon}{2}\right) \leq P\left(\left|\sum_{j=1}^{s_n}\left(B_{2j-1,n}(x) - B^*_{2j-1,n}(x)\right)\right| > \frac{\varepsilon}{4}\right)$$
$$+P\left(\left|\sum_{j=1}^{s_n} B^*_{2j-1,n}(x)\right| > \frac{\varepsilon}{4}\right) = \Delta_{1,n}(x) + \Delta_{2,n}(x). \tag{4.14}$$

By assumption A.1,

$$\left\|B_{2j-1,n}(x)\right\|_\infty = \left\|\frac{1}{n}\sum_{t=(2j-2)b_n+1}^{(2j-1)b_n}\left(\frac{x_{t,n}-x}{h_n}\right)^j \frac{1}{h_n}K\left(\frac{x_{t,n}-x}{h_n}\right)\varepsilon_{t,M_n}\right\|_\infty \leq C_1\frac{b_nM_n}{nh_n}, \tag{4.15}$$

with C_1 a positive constant. Then, by using Markov's inequality, (4.13) and (4.15), one has

$$\Delta_{1,n}(x) \leq \sum_{j=1}^{s_n} P\left(\left|B_{2j-1,n}(x) - B^*_{2j-1,n}(x)\right| > \frac{\varepsilon}{4s_n}\right) \leq \frac{8s_n^2}{\varepsilon}C_1\frac{b_nM_n}{nh_n}\alpha\left(b_n\right). \tag{4.16}$$

Concerning $\Delta_{2,n}(x)$, if σ_j^2 denotes $Var\left(B^*_{2j-1,n}(x)\right) = Var\left(B_{2j-1,n}(x)\right)$, then Davidov's Lemma (see (34)) and assumption A.6 lead to $\sum_{k=1}^{\infty} k\left|c(k)\right| < \infty$, where $Cov\left(\epsilon_t, \epsilon_{t+k}\right) = \sigma^2 c(k)$. Therefore Proposition 2 in (17) can be applied to obtain that

$$\sum_{j=1}^{s_n}\sigma_j^2 \leq E\left(h_n^{-j}t^\star_{j,n}(x)\right)^2 = \frac{1}{nh_n}\nu_{2j}f(x)c(\varepsilon) \leq \frac{1}{nh_n}\nu_{2j}C_2c(\varepsilon) = \frac{C_v}{nh_n}, \tag{4.17}$$

where $\nu_{2j} = \int u^{2j}K^2(u)du$, $c(\varepsilon) = \sigma^2\left(c(0) + 2\sum_{k=1}^{\infty}\left|c(k)\right|\right)$ and C_2 and C_v are positive constants.

Now, Lemma 2, (4.15) and (4.17) lead to

$$\Delta_{2,n}(x) \leq 2\exp\left(-\frac{1}{2}\frac{\varepsilon^2}{16} \Big/ \left(\frac{C_v}{nh_n} + \frac{\varepsilon}{12}C_1\frac{b_n M_n}{nh_n}\right)\right) = 2\exp\left(-\frac{1}{32}\frac{\varepsilon^2 nh_n}{C_v + C_1\varepsilon b_n M_n \frac{1}{12}}\right). \tag{4.18}$$

It follows from (4.12), (4.14), (4.16) and (4.18) that

$$P\left(|P_{j,2,n}| > \varepsilon\right) \leq 2\gamma_n\left[\left(\frac{8s_n^2}{\varepsilon}C_1\frac{b_n M_n}{nh_n}\alpha\left(b_n\right)\right) + 2\exp\left(-\frac{1}{32}\frac{\varepsilon^2 nh_n}{C_v + C_1\varepsilon b_n M_n \frac{1}{12}}\right)\right]. \tag{4.19}$$

By choosing $\varepsilon = \varepsilon_n = C_\varepsilon\left(\frac{\ln n}{nh_n}\right)^{1/2}$, expression (4.19) can be written in the form

$$P\left(|P_{2,j,n}| > \varepsilon\right) \leq a_n + b_n,$$

where

$$a_n = \frac{16C_1}{C_\varepsilon}\gamma_n\frac{n}{b_n}\left(\frac{M_n^2 n}{h_n \ln n}\right)^{1/2}\alpha\left(b_n\right)$$

and

$$b_n = 4\gamma_n n^{\frac{-c_\varepsilon^2}{32\left(C_v + \frac{C_1}{12}C_\varepsilon\right)}}.$$

From assumption A.7, the sequence a_n is summable and by selecting a large enough C_ε, one can obtain that b_n is also summable. Then, the Borel-Cantelli Lemma leads to (4.6) and the proof of Proposition 3 is completed.

PROOF OF PROPOSITION 5.

This proof follows in the same way as Proposition 2 but using Proposition 4 instead of Proposition 1 and, therefore, it has been omitted.

PROOF OF PROPOSITION 6.

Similar arguments to those used in Proposition 3 are employed here and only the main differences are presented below.

According to (3.1), $H_{(n)}\tilde{\vec{\Delta}}_{1,n}(x) = \left(H_{(n)}^{-1}\tilde{C}_{(n)}(x)H_{(n)}^{-1}\right)^{-1} H_{(n)}^{-1}\tilde{\vec{G}}_{(n)}^{*}(x)$. Since Proposition 4 ensures the convergence of the first term, our attention is focused on the components of $H_{(n)}^{-1}\tilde{\vec{G}}_{(n)}^{*}(x)$, which are given by

$$h_n^{-j}\tilde{g}_{j,n}^{*}(x) = \frac{1}{n}\sum_{t=1}^{n}\chi_{t,n}(x),$$

with

$$\chi_{1,n}(x) = \left(1-\rho^2\right)\left(\frac{x_{1,n}-x}{h_n}\right)^j K_n\left(x_{1,n}-x\right)\varepsilon_{1,n}$$

and

$$\chi_{t,n}(x) = \left[\left(\frac{x_{t,n}-x}{h_n}\right)^j - \rho\left(\frac{x_{t-1,n}-x}{h_n}\right)^j\right] K_n\left(x_{t,n}-x\right)e_{t,n},\ t=2,3,\ldots,n.$$

Note that the random variables $\{\chi_{t,n}(x)\}_{t=1}^{n}$ are independent.

Next, a truncation argument is again used. Let ε_{t,M_n} be as in (4.1) and consider the error variables defined by $e_{t,M_n} = \varepsilon_{t,M_n} - \rho\varepsilon_{t-1,M_n}$. Then terms $\tilde{g}_{j,n}^{*B}(x)$ are constructed as $\tilde{g}_{j,n}^{*}(x)$, but replacing $\varepsilon_{1,n}$ with ε_{1,M_n} and $e_{t,n}$ with e_{t,M_n}, $t=2,\ldots,n$. Hence, for $j=0,\ldots,p$,

$$\sup_{x\in A} h_n^{-j}\tilde{g}_{j,n}^{*}(x) \le \sup_{x\in A} h_n^{-j}\left(\tilde{g}_{j,n}^{*}(x) - \tilde{g}_{j,n}^{*B}(x)\right) + \sup_{x\in A} h_n^{-j}\tilde{g}_{j,n}^{*B}(x) = \tilde{Q}_{j,1,n} + \tilde{Q}_{j,2,n}.$$

For $\eta > 0$, the Tchebyshev and Cauchy-Schwartz inequalities lead to

$$P\left(\tilde{Q}_{j,1,n} > \eta\right) \le \sup_{x\in A}\left[\frac{1}{n\eta}\left(1-\rho^2\right)\left(\frac{x_{1,n}-x}{h_n}\right)^j K_n\left(x_{1,n}-x\right)\left[P\left(|\varepsilon_{1,n}| > M_n\right)\right]^{1/2}\sigma_e\right.$$

$$+\frac{1}{n\eta}\sum_{t=2}^{n}\left(\left(\frac{x_{t,n}-x}{h_n}\right)^j-\rho\left(\frac{x_{t-1,n}-x}{h_n}\right)^j\right)K_n\left(x_{t,n}-x\right)\left[P\left(|e_t|>2M_n\right)\right]^{1/2}\sigma_e\Bigg].$$

Since M_n is increasing, $|\varepsilon_{1,n}|<M_n$ and $|e_{t,n}|<2M_n$ almost surely. Hence, using the Borel-Cantelli Lemma, it is concluded that $\tilde{Q}_{j,1,n}\to 0$ almost sure as $n\to\infty$.

With regard to the almost sure convergence of $\tilde{Q}_{j,2,n}$ we proceed as in (4.3). Consider γ_n intervals $I_{k,n}=\left[x'_{k,n}-l_n,x'_{k,n}+l_n\right]$ covering A. Then

$$\tilde{Q}_{j,2,n}=\max_{1\le k\le\gamma_n}\sup_{x\in A\cap I_{k,n}}h_n^{-j}\left|\tilde{g}^{*B}_{j,n}(x)-\tilde{g}^{*B}_{j,n}(x'_{k,n})\right|+\max_{1\le k\le\gamma_n}h_n^{-j}\left|\tilde{g}^{*B}_{j,n}(x'_{k,n})\right|=\tilde{P}_{j,1,n}+\tilde{P}_{j,2,n}.$$

From assumption A.1,

$$\tilde{P}_{j,1,n}\le\frac{M_n\left(1-\rho^2\right)C_j}{nh_n}\left|\frac{x'_{k,n}-x}{h_n}\right|+\frac{2M_nC_j}{nh_n}\sum_{t=2}^{n}\left(\left|\frac{x'_{k,n}-x}{h_n}\right|(1-\rho)+O\left(\frac{1}{nh_n}\right)\right)$$

$$=O\left(\frac{M_n}{h_n}\left(\frac{l_n}{h_n}+\frac{1}{nh_n}\right)\right)=O\left(\frac{M_n}{\gamma_nh_n^2}\right),$$

and definitions of M_n and γ_n in A.7$'$ allow us to conclude that $\tilde{P}_{j,1,n}\le O\left(\frac{\ln n}{nh_n}\right)^{1/2}$.

Thus, only the convergence of $\tilde{P}_{j,2,n}$ remains to be studied. For each $\varepsilon>0$,

$$P\left(\max_{1\le k\le\gamma_n}h_n^{-j}\left|\tilde{g}^{*B}_{j,n}(x'_{k,n})\right|>\varepsilon\right)=\gamma_n\sup_{x\in A}P\left(\left|\sum_{t=1}^{n}\frac{1}{n}\chi^B_{t,n}(x)\right|>\varepsilon\right),\tag{4.20}$$

where $\chi^B_{t,n}(x)$ denotes independent random variables defined as $\chi_{t,n}(x)$, but replacing $\varepsilon_{1,n}$ with ε_{1,M_n} and $e_{t,n}$ with e_{t,M_n}.

From assumptions A.1 and A.7$'$, it is obtained that

$$\left|\frac{1}{n}\chi^B_{t,n}(x)\right|\le C_e\frac{M_n}{nh_n},\quad t=1,2,\dots,n,\tag{4.21}$$

with C_e being a positive constant.

On the other hand, taking into account the independence of random variables $\chi_{t,n}(x)$ and using Proposition 3.2 of (1), one has that

$$\sum_{t=1}^{n} Var\left(\frac{1}{n}\chi_{t,n}^{B}(x)\right) = Var\left(h_n^{-j}\tilde{g}_{j,n}^{*B}(x)\right) \leq \frac{1}{nh_n}\nu_{2j}f(x)(1-\rho)^2\sigma_e^2 = \frac{C_v}{nh_n}, \tag{4.22}$$

where σ_e^2 is the variance of $e_{t,n}$.

Using (4.21), (4.22) and the Berstein inequality (Lemma 2), as in (4.18), it is concluded that

$$P\left(\left|\sum_{t=1}^{n}\frac{1}{n}\chi_{t,n}^{B}(x)\right| > \varepsilon\right) \leq 2\exp\left(\frac{-3nh_n\varepsilon^2}{6C_v + 2C_eM_n\varepsilon}\right). \tag{4.23}$$

If the bound in (4.23) is considered in (4.20) and $\varepsilon = \varepsilon_n = C_\varepsilon\left(\frac{\ln n}{nh_n}\right)^{1/2}$ is taken with a large enough C_ε, then Borel-Cantelli's Lemma leads to

$$\tilde{P}_{j,2,n} = O\left(\frac{\ln n}{nh_n}\right)^{1/2} \quad \text{almost sure,}$$

and the proof of Proposition 6 is completed.

PROOF OF THEOREM 3.

From definitions (1.10) and (1.11),

$$H_{(n)}\left|\hat{\vec{\beta}}_{F,n}(x) - \tilde{\vec{\beta}}_{G,n}(x)\right|$$

$$= H_{(n)}\left|\left(X_{(n)}^t\hat{\Omega}_n^{-1}X_{(n)}\right)^{-1}X_{(n)}^t\hat{\Omega}_n^{-1}\vec{Y}_{(n)} - \left(X_{(n)}^t\Omega_n^{-1}X_{(n)}\right)^{-1}X_{(n)}^t\Omega_n^{-1}\vec{Y}_{(n)}\right|, \tag{4.24}$$

where the matrices $\hat{P}_{(n)}^tW_{(n)}\hat{P}_{(n)}$ and $P_{(n)}^tW_{(n)}P_{(n)}$ have been denoted by $\hat{\Omega}_n^{-1}$ and Ω_n^{-1}, respectively. In addition, the dependence on x of all the matrices in (4.24) has been omitted for simplicity in notation.

Replacing $\vec{Y}_{(n)}$ with $\vec{M}_{(n)} + \vec{\varepsilon}_{(n)}$ and using (2.3), one obtains that

$$\sup_{x \in A} H_{(n)} \left| \hat{\vec{\beta}}_{F,n}(x) - \tilde{\vec{\beta}}_{G,n}(x) \right| \leq \Phi_{1,n} + \Phi_{2,n} + \Phi_{3,n}, \tag{4.25}$$

where

$$\begin{aligned}
\Phi_{1,n} &= \sup_{x \in A} \left\{ \frac{1}{(p+1)!} m^{(p+1)}(x) H_{(n)} \left| \hat{C}_{(n)}^{-1} X_{(n)}^t \hat{\Omega}_n^{-1} \vec{U}_x - \tilde{C}_{(n)}^{-1} X_{(n)}^t \Omega_n^{-1} \vec{U}_x \right| \right\}, \\
\Phi_{2,n} &= \sup_{x \in A} \left\{ H_{(n)} \left| \hat{C}_{(n)}^{-1} X_{(n)}^t \hat{\Omega}_n^{-1} \vec{R}_{(n)} - \tilde{C}_{(n)}^{-1} X_{(n)}^t \Omega_n^{-1} \vec{R}_{(n)} \right| \right\}, \\
\Phi_{3,n} &= \sup_{x \in A} \left\{ H_{(n)} \left| \hat{C}_{(n)}^{-1} X_{(n)}^t \hat{\Omega}_n^{-1} \vec{\varepsilon}_{(n)} - \tilde{C}_{(n)}^{-1} X_{(n)}^t \Omega_n^{-1} \vec{\varepsilon}_{(n)} \right| \right\},
\end{aligned}$$

with $\vec{U}_x$ denoting the vector $\left((x_{1,n} - x)^{p+1}, \ldots, (x_{n,n} - x)^{p+1}\right)^t$.

The three terms on the right hand-side of (4.25) are studied next.

Let $X_{(n)}^t \hat{\Omega}_n^{-1} \vec{U}_x$ and $X_{(n)}^t \Omega_n^{-1} \vec{U}_x$ be denoted by $\hat{C}_{u,n}$ and $\tilde{C}_{u,n}$, respectively. From assumption A.4$'$ follows that

$$\Phi_{1,n} \leq C_M H_{(n)} \sup_{x \in A} \left| \hat{C}_{(n)}^{-1} \hat{C}_{u,n} - \tilde{C}_{(n)}^{-1} \tilde{C}_{u,n} \right|$$

$$\leq C_M H_{(n)} \sup_{x \in A} \left| \hat{C}_{(n)}^{-1} \left(\hat{C}_{u,n} - \tilde{C}_{u,n} \right) \right| + C_M H_{(n)} \sup_{x \in A} \left| \left(\hat{C}_{(n)}^{-1} - \tilde{C}_{(n)}^{-1} \right) \tilde{C}_{u,n} \right| = \Phi_{1,1,n} + \Phi_{1,2,n},$$

where C_M is a positive real number.

Proposition 4 together with the almost sure convergence of the estimator $\hat{\rho}_n$ (Proposition 3.4 of (1)) lead to

$$\Phi_{1,1,n} = C_M \sup_{x \in A} \left| \left(H_{(n)}^{-1} \hat{C}_{(n)} H_{(n)}^{-1} \right)^{-1} H_{(n)}^{-1} \left(\hat{C}_{u,n} - \tilde{C}_{u,n} \right) \right|$$

$$= C_M \sup_{x \in A} \left(\left((1 - \hat{\rho}_n)^2 f(x) S + o(1) \right)^{-1} \left((1 - \rho)^2 f(x) \vec{\mu} + o(1) \right) h_n^{p+1} \right) o(1) = o\left(h_n^{p+1} \right). \tag{4.26}$$

On the other hand, it can be written that

$$\Phi_{1,2,n} = C_M \sup_{x\in A}\left|\left[\left(H_{(n)}^{-1}X_{(n)}^t\left(\hat{\Omega}_n^{-1}-\Omega_n^{-1}\right)X_{(n)}H_{(n)}^{-1}+H_{(n)}^{-1}X_{(n)}^t\Omega_n^{-1}X_{(n)}H_{(n)}^{-1}\right)^{-1} - \left(H_{(n)}^{-1}X_{(n)}^t\Omega_n^{-1}X_{(n)}H_{(n)}^{-1}\right)^{-1}\right]H_{(n)}^{-1}X_{(n)}^t\Omega_n^{-1}\vec{U}_x\right|.$$

Then, Proposition 4 and Lemma 1 of (1) allow us to deduce that $\Phi_{1,2,n} \leq o\left(h_n^{p+1}\right)$, which jointly with (4.26) lead to $\Phi_{1,n} \leq o\left(h_n^{p+1}\right)$.

Taking into account the expression of $\vec{R}_{(n)}$ given in (2.2) and using the same arguments employed previously, it is directly deduced that $\Phi_{2,n} \leq o\left(h_n^{p+1}\right)$.

Finally, the strong consistency of $\Phi_{3,n}$ is established as follows:

$$\begin{aligned}\Phi_{3,n} &\leq \sup_{x\in A}\left(H_{(n)}\left|\hat{C}_{(n)}^{-1}X_{(n)}^t\left(\hat{\Omega}_n^{-1}-\Omega_n^{-1}\right)\vec{\varepsilon}_{(n)}\right|\right)+\sup_{x\in A}H_{(n)}\left(\left|\left(\hat{C}_{(n)}^{-1}-\tilde{C}_{(n)}^{-1}\right)X_{(n)}^t\Omega_n^{-1}\vec{\varepsilon}_{(n)}\right|\right)\\ &\leq \sup_{x\in A}\left(\left|H_{(n)}\hat{C}_{(n)}^{-1}H_{(n)}\right|\left|H_{(n)}^{-1}X_{(n)}^t\left(\hat{\Omega}_n^{-1}-\Omega_n^{-1}\right)\vec{\varepsilon}_{(n)}\right|\right)\\ &+\sup_{x\in A}\left(\left|H_{(n)}\left(\hat{C}_{(n)}^{-1}-\tilde{C}_{(n)}^{-1}\right)H_{(n)}\right|\left|H_{(n)}^{-1}X_{(n)}^t\Omega_n^{-1}\vec{\varepsilon}_{(n)}\right|\right)=\Phi_{3,1,n}+\Phi_{3,2,n}.\end{aligned}$$

Now, convergence of both $\Phi_{3,1,n}$ and $\Phi_{3,2,n}$ are derived by using Propositions 4 and 6, the strong convergence of $\hat{\rho}_n$ and similar arguments to those employed to establish the consistency of $\Phi_{1,1,n}$ and $\Phi_{1,2,n}$. In particular, it is concluded that $\Phi_{3,n} \leq o\left(\frac{\ln n}{nh_n}\right)^{1/2}$ almost sure and the proof of (3.10) is stated.

ACKNOWLEDGMENTS

This work was partially supported by Grants PGIDT01PXI10505PR, PB98-0182-C02-01 and MCyT Grant BFM2002-00265 (European FEDER support included). The authors are pleased to acknowledge the helpful comments of the referee.

REFERENCES

(1) Vilar-Fernández, J.M.; Francisco-Fernández, M. Local polynomial regression smoothers with AR-error structure. TEST 2002, 11 (2), 439–464.

(2) Sacks, J ; Ylvisaker, D. Designs for regression problems with correlated errors III. Ann. Math. Statist. 1970, 41, 2057–2074.

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

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

(5) Tsybakov, A. B. Robust reconstruction of functions by the local-approximation method. Problems Inform. Transmission 1986, 22, 133–146.

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

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

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

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

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

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

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

(13) Härdle, W.; Tsybakov, A. Local polynomial estimators of the volatility function in nonparametric autoregression. J. Econometrics **1997**, *81*, 223–242.

(14) Opsomer, J.D. Nonparametric regression in the presence of correlated errors. In *Modelling Longitudinal and Spatially Correlated Data: Methods, Applications and Future Directions*; Gregoire, T.G., Brillinger, D.R., Diggle, P.J., Russek-Cohen, E., Warren, W.G., Wolfinger, R.D., Eds.; Springer: New York, **1997**, 339-348.

(15) Vilar-Fernández, J.A.; Vilar-Fernández, J.M. Recursive estimation of regression functions by local polynomial fitting. Ann. Ins. Statist. Math. **1998**, *50* (4), 729–754.

(16) Vilar-Fernández, J.M.; Vilar-Fernández, J.A. Recursive local polynomial regression under dependence conditions. TEST **2000**, *9* (1), 209–233.

(17) Francisco-Fernández, M.; Vilar-Fernández, J.M. Local polynomial regression estimation with correlated errors. Comm. Statist. Theory and Methods **2001**, *30* (7), 1271–2293.

(18) Fan, J.; Gijbels, I. *Local polynomial modelling and its applications*; Chapman and Hall:

London, 1996.

(19) Collomb, G.; Härdle, W. Strong uniform convergence rates in robust nonparametric time series analysis and prediction: Kernel regression estimation from dependent observations. Stochastic Process. Appl. 1986, 23, 77–89.

(20) Györfi, L.; Härdle, W.; Sarda, P.; Vieu, P. Nonparametric curve estimation from time series; Lectures Notes in Statistics, Springer: Berlin, 1989; Vol 60.

(21) Roussas, G.G. Nonparametric regression estimation under mixing conditions. Stochastic Process. Appl. 1990, 36, 107–116.

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

(23) Masry, E.; Tjøstheim, D. Nonparametric estimation and identification of nonlinear ARCH time series. Econometric Theory 1995, 65, 258–289.

(24) Lu, Z.; Cheng, P. Distribution-free strong consistency for nonparametric kernel regression involving nonlinear time series. J. Statist. Plann. Inference 1997, 65, 67–86.

(25) Nze, P.A; Doukhan, P. Functional estimation for time series: Uniform convergence properties. J. Statist. Plann. Inference 1998, 68, 5–29.

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

(27) Roussas, G.G. Consistent regression with fixed design points under dependence condi-

tions. Statist. Probab. Lett. 1989, 8, 41–50.

(28) Roussas, G.; Tran, L.; Ionannides, D. Fixed design regression for time series: asymptotic normality. J. Multivariate Anal. 1992, 40, 262–291.

(29) Tran, L.; Roussas, G.; Yakowitz, S.; Truong, V. Fixed-design regression for linear time series. Ann. Statist. 1996, 24 (3), 975–991.

(30) Rio, E. The functional law of the iterated logarithm for stationary strongly mixing sequences. Ann. Prob. 1995, 23 (3), 1188–1203.

(31) Carbon, M.; Tran, L.T.; Wu, B. Kernel density estimation for random fields. Statist. Probab. Lett. 1997, 36, 115-125.

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

(33) Pollard, D. Convergence of Stochastic Processes; Springer: New York, 1984.

(34) Hall, P.; Heyde, C.C. Martingale Limit Theory and its Applications; Academic Press: New York, 1980.