*Sociedad Española de Estadística
e Investigación Operativa*

Test

Volume 11, Number 2. December 2002

Local Polynomial Regression Smoothers with AR-error Structure

Juan M. Vilar Fernándezs and Mario Francisco Fernández
Departamento de Matemáticas
Universidad de A Coruña, Spain

Sociedad de Estadística e Investigación Operativa
Test (2002) *Vol.* **11**, *No.* 2, pp. 439–464

Sociedad de Estadística e Investigación Operativa
Test (2002) *Vol.* **11**, *No.* 2, pp. 439–464

Local Polynomial Regression Smoothers with AR-error Structure

Juan M. Vilar Fernández* **and Mario Francisco Fernández**
Departamento de Matemáticas
Universidad de A Coruña, Spain

Abstract

Consider the fixed regression model with random observation error that follows an AR(1) correlation structure. In this paper, we study the nonparametric estimation of the regression function and its derivatives using a modified version of estimators obtained by weighted local polynomial fitting. The asymptotic properties of the proposed estimators are studied; expressions for the bias and the variance/covariance matrix of the estimators are obtained and the joint asymptotic normality is established. In a simulation study, a better behavior of the Mean Integrated Squared Error of the proposed regression estimator with respect to that of the classical local polynomial estimator is observed when the correlation of the observations is large.

Key Words: Nonparametric estimators; local polynomial fitting; autoregressive process.

AMS subject classification: 62G07, 62H12, 62M09.

1 Introduction

Local polynomial (LP) fitting has gained acceptance as an attractive method for estimating the regression function and its derivatives. The advantages of this nonparametric estimation method include its simplicity, it is highly intuitive, easy to compute, it achieves automatic boundary corrections and possesses important minimax properties.

Since early papers on LP, Stone (1977) and Cleveland (1979), many relevant contributions of this method have appeared in statistics literature,

This work has been partially supported by grants PB98-0182-C02-01, PGIDT01PXI10505PR and MCyT Grant BFM2002-00265 (European FEDER support included).

*Correspondence to: Juan M. Vilar-Fernández, Departamento de Matemáticas, Universidad de A Coruña, 15071 A Coruña, Spain, Email: eijvilar@udc.es

Received: May 2001; Accepted: January 2002

such as Tsybakov (1986), Fan (1993), Hastie and Loader (1993), Ruppert and Wand (1994), Fan and Gijbels (1995) and Ruppert et al. (1995). In these papers, the independence of the observations was assumed. Masry (1996b), Masry (1996a), Masry and Fan (1997), Härdle and Tsybakov (1997), Härdle et al. (1998) and Vilar and Vilar (1998) studied the asymptotic properties of LP regression estimator in a context of dependence. A broad study of this estimation method can be found in the monograph by Fan and Gijbels (1996).

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 smooth regression function which is defined in $[0,1]$. Without loss of generality, we can assume that the $\varepsilon_{t,n}$, $1 \leq t \leq n$, are unobserved random variables with zero mean and finite variance, σ_ε^2. We assume, for each n, that $\{\varepsilon_{1,n}, \varepsilon_{2,n}, ..., \varepsilon_{n,n}\}$ have the same joint distribution as $\{\epsilon_1, \epsilon_2, ..., \epsilon_n\}$, where $\{\epsilon_t,\, t \in \mathbb{Z}\}$ is a strictly stationary stochastic process. The design points $x_{t,n}$, $1 \leq t \leq n$, follow a regular design generated by a density f. So, for each n, the design points are defined by

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

where f is a positive function defined in $[0,1]$.

For simplicity, the subindex n in the sample data and in the errors notation will be avoided, that is, we are going to write x_i, Y_i and ε_i.

Francisco-Fernández and Vilar-Fernández (2001) studied the properties of the LP estimator of the regression function and its derivatives for the model given in (1.1), when the random error ε_t has absolutely sumable autocovariances. The asymptotic normality of a general linear smoother of the regression function, $m(x)$, for model (1.1), under dependence conditions imposed on the $\varepsilon_t's$, has been established in Roussas et al. (1992) and Tran et al. (1996).

In this paper, it is supposed that the stochastic process $\{\varepsilon_t, t \in \mathbb{Z}\}$ follows an AR(1) type correlation structure. A new estimator based on transforming the statistical model to get uncorrelated errors and then use LP fitting is proposed. That is, the well-studied feature of modelling the correlation structure of the data when a global polynomial model is fitted to the case of LP fitting, is extended.

The organization of the article is as follows: In Section 2, the proposed estimator is defined. In Section 3, expressions for the bias and the variance/covariance matrix of the proposed estimators of the regression function and its derivatives are obtained and joint asymptotic normality is established. In Section 4, a simulation study is presented, where an improved behavior of the proposed estimator with respect to the LP estimator, under the Mean Integrated Squared Error criterion can be observed, when the correlation coefficient ρ is large. Finally, the last section is devoted to the proofs of the obtained results.

2 Definition of the estimators

Consider the regression model given in (1.1) and assume that the $(p+1)$th derivative of $m(\cdot)$ at the point x exists. The parameter vector $\vec{\beta}(x) = (\beta_0(x), \beta_1(x), \cdots, \beta_p(x))^t$, where $\beta_j(x) = m^{(j)}(x)/(j!)$, with $j = 0, 1, \ldots, p$, using a weighted LP fitting, can be estimated by minimizing the function

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

where $\omega_{n,t} = n^{-1}K_h(x_t - x)$ are the weights, $K_h(u) = h_n^{-1}K\left(h_n^{-1}u\right)$, K being a kernel function and h_n the bandwidth that controls the degree of smoothing. Then, assuming the invertibility of $X_{(n)}^t W_{(n)} X_{(n)}$ (for this, at least $p+1$ points with positive weights, $\omega_{n,t}$, are required), the estimator of $\vec{\beta}(x)$, obtained as a solution to the weighted least squares problem given in (2.1), is

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

where we have introduced the following matrix notation

$$\vec{Y}_{(n)} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \qquad X_{(n)} = \begin{pmatrix} 1 & (x_1 - x) & \cdots & (x_1 - x)^p \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (x_n - x) & \cdots & (x_n - x)^p \end{pmatrix}, \tag{2.3}$$

$W_{(n)} = \text{diag}\left(\omega_{n,1}, \ldots, \omega_{n,n}\right)$ is the diagonal matrix of weights, $S_{(n)}$ is the matrix $(p+1) \times (p+1)$ whose $(i+1, j+1)$th element is $s_{i,j}^{(n)} = s_{i+j}^{(n)}$, $i, j = 0, \ldots, p$, with

$$s_j^{(n)} = \frac{1}{n} \sum_{t=1}^{n} \left(x_t - x\right)^j K_h \left(x_t - x\right), \quad 0 \leq j \leq 2p, \tag{2.4}$$

and $\vec{T}_{(n)}$ is the vector$\left(t_{0,(n)}, t_{1,(n)}, ..., t_{p,(n)}\right)^t$, with

$$t_{i,(n)} = \frac{1}{n} \sum_{t=1}^{n} \left(x_t - x\right)^i K_h \left(x_t - x\right) Y_t, \quad 0 \leq i \leq p. \tag{2.5}$$

The asymptotic properties of the LP estimator, $\hat{\beta}_L(x)$, were studied by Francisco-Fernández and Vilar-Fernández (2001) when the random errors are correlated. In this paper, it is assumed that the stochastic process ϵ_t follows an AR(1) type correlation structure: $\epsilon_t = \rho\epsilon_{t-1} + e_t$, $t \in \mathbb{Z}$, with $|\rho| < 1$ and $\{e_t\}_{t \in \mathbb{Z}}$, a noise process, with mean zero and finite variance, σ_e^2. The variance/covariance matrix of this process is $E\left(\vec{\epsilon}\vec{\epsilon}^t\right) = \sigma_e^2 \Omega_{(n)}$, where $\vec{\epsilon}^t = (\epsilon_1, \epsilon_2, ..., \epsilon_n)$ and $\Omega_{(n)}$ is a nonsingular matrix and positive definite, given by

$$\Omega_{(n)} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{n-1} \\ \rho & 1 & \rho & \ldots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \ldots & \rho^{n-3} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \ldots & 1 \end{pmatrix}.$$

Since $\Omega_{(n)}$ is positive definite, a nonsingular matrix $P_{(n)}$, with the property $P_{(n)}^t P_{(n)} = \Omega_{(n)}^{-1}$, always exists. This matrix is

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

To calculate the new nonparametric estimator of the regression function and its derivatives, in a first step, the observations are transformed to obtain a model with uncorrelated errors. For this purpose, performing a Taylor series expansion in a neighborhood of x, one gets

$$\begin{aligned} m\left(x_t\right) &= \sum_{j=0}^{p} \frac{m^{(j)}(x)}{j!}\left(x_t - x\right)^j + \frac{m^{(p+1)}(x)}{(p+1)!}\left(x_t - x\right)^{p+1} + o\left(x_t - x\right)^{p+1}, \\ & \quad t = 1, \ldots, n, \end{aligned} \tag{2.6}$$

or, in matrix form,

$$\vec{M}_{(n)} = X_{(n)}\vec{\beta}(x) + \frac{m^{(p+1)}(x)}{(p+1)!}\begin{pmatrix} (x_1 - x)^{p+1} \\ \vdots \\ (x_n - x)^{p+1} \end{pmatrix} + \begin{pmatrix} o(x_1 - x)^{p+1} \\ \vdots \\ o(x_n - x)^{p+1} \end{pmatrix}, \tag{2.7}$$

where $\vec{M}_{(n)} = \left(m\left(x_1\right), ..., m\left(x_n\right)\right)^t$. So, model (1.1) can be approximated by

$$\vec{Y}_{(n)} \approx X_{(n)}\vec{\beta}(x) + \vec{\varepsilon}_{(n)}, \tag{2.8}$$

and the errors of the following regression model are uncorrelated

$$P_{(n)}\vec{Y}_{(n)} = P_{(n)}X_{(n)}\vec{\beta}(x) + P_{(n)}\vec{\varepsilon}_{(n)}. \tag{2.9}$$

Now, assuming that $X_{(n)}^t P_{(n)}^t W_{(n)} P_{(n)} X_{(n)}$ is nonsingular, an estimator of $\vec{\beta}(x)$ is obtained using weighted least squares

$$\begin{aligned} \tilde{\beta}_G(x) &= \left(X_{(n)}^t P_{(n)}^t W_{(n)} P_{(n)} X_{(n)}\right)^{-1} X_{(n)}^t P_{(n)}^t W_{(n)} P_{(n)} \vec{Y}_{(n)} \\ &= \tilde{C}_{(n)}^{-1} \tilde{G}_{(n)}, \end{aligned} \tag{2.10}$$

where $\tilde{C}_{(n)} = X_{(n)}^t P_{(n)}^t W_{(n)} P_{(n)} X_{(n)}$ and $\tilde{G}_{(n)} = X_{(n)}^t P_{(n)}^t W_{(n)} P_{(n)} Y_{(n)}$.

The drawback of this estimator is that in most practical situations it cannot be computed because the matrix $P_{(n)}$ is unknown but should be estimated. A natural estimator for $P_{(n)}$ is obtained substituting ρ by the estimator $\hat{\rho}_{(n)}$,

$$\hat{\rho}_{(n)} = \frac{\sum_{t=1}^{n-1} \hat{\varepsilon}_t \hat{\varepsilon}_{t+1}}{\sum_{t=1}^{n} \hat{\varepsilon}_t^2}, \tag{2.11}$$

where $\hat{\varepsilon}_t = Y_t - \hat{m}_n(x_t)$, $1 \leq t \leq n$, are nonparametric residuals. These are calculated using a consistent estimator of $m(x_t)$, for example, the LP estimator given in (2.2). Now, the new estimator of $\vec{\beta}(x)$ is

$$\begin{aligned} \hat{\beta}_F(x) &= \left(X^t_{(n)} \hat{P}^t_{(n)} W_{(n)} \hat{P}_{(n)} X_{(n)} \right)^{-1} X^t_{(n)} \hat{P}^t_{(n)} W_{(n)} \hat{P}_{(n)} \vec{Y}_{(n)} \\ &= \hat{C}^{-1}_{(n)} \hat{G}_{(n)}, \end{aligned} \tag{2.12}$$

where it is assumed that $\hat{C}^{-1}_{(n)}$ exists.

The extension of $\hat{\beta}_F(x)$ 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.

3 Asymptotic analysis

In this section, asymptotic expressions for the bias and variance/covariance matrix of $\hat{\beta}_F(x)$ are obtained and the joint asymptotic normality is established. Firstly, the asymptotic properties of $\tilde{\beta}_G(x)$ defined in (2.10) are studied. We follow a similar approach to that employed by Francisco-Fernández and Vilar-Fernández (2001) to obtain the asymptotic normality of the LP estimator $\hat{\beta}_L(x)$.

The following assumptions will be made in our analysis:

A.1. The functions f' and $m^{(p+1)}$ are continuous on $[0, 1]$.

A.2. The kernel K is symmetric, with support $[-1, 1]$, Lipschitz continuous and $K > 0$.

A.3. The point x at which the estimation is taking place satisfies $h_n < x < 1 - h_n$, for all $n \geq n_0$ where n_0 is fixed.

A.4. The sequence of bandwidths, $\{h_n\}$, satisfies that $h_n > 0$, $h_n \to 0$, $nh_n \to \infty$ as $n \to \infty$.

First, the convergence for the entries of matrices $\tilde{C}_{(n)} = \left(\tilde{c}^{(n)}_{i,j} \right)^p_{i,j=0}$ and $\tilde{G}_{(n)} = \left(\tilde{g}_{j,(n)} \right)^p_{j=0}$ is studied.

Proposition 3.1. *Under assumptions A1-A4, we have*

$$\lim_{n\to\infty} h_n^{-(i+j)}\tilde{c}_{i,j}^{(n)} = (1-\rho)^2 f(x)\mu_{i+j}, \quad 0 \le i,j \le p+1. \tag{3.1}$$

This result can be expressed in matrix form as

$$\lim_{n\to\infty} H_{(n)}^{-1}\tilde{C}_{(n)}H_{(n)}^{-1} = (1-\rho)^2 f(x)S, \tag{3.2}$$

where $H_{(n)} = \text{diag}\left(1, h_n, h_n^2, \cdots, h_n^p\right)$ *and* S *is the* $(p+1)\times(p+1)$ *matrix whose* $(i+1, j+1)$*th element is* $s_{i,j} = \mu_{i+j}$, $i,j = 0,\ldots,p$, *with* $\mu_r = \int u^r K(u)\,du$.

By condition A2, S is positive definite, see Lemma 1 of Tsybakov (1986), and therefore nonsingular.

The next proposition establishes the asymptotic variance/covariance matrix of vector $\tilde{G}_{(n)}^* = X_{(n)}^t P_{(n)}^t W_{(n)} P_{(n)}\left(\vec{Y}_{(n)} - \vec{M}_{(n)}\right) = \left(\tilde{g}_{j,(n)}^*\right)_{j=0}^p$, that is, of vector $\tilde{G}_{(n)}$ centered with respect to $\vec{M}_{(n)} = (m(x_1), \cdots, m(x_n))^t$.

Proposition 3.2. *Under assumptions A1-A4, we have*

$$\lim_{n\to\infty} nh_n Cov\left(h_n^{-j}\tilde{g}_{i,(n)}^*, h_n^{-i}\tilde{g}_{j,(n)}^*\right) = \nu_{j+i} f(x)(1-\rho)^2\sigma_e^2, \quad 0 \le i,j \le p, \tag{3.3}$$

or, in matrix form

$$\lim_{n\to\infty} nh_n E\left(H_{(n)}^{-1}\tilde{G}_{(n)}^*\tilde{G}_{(n)}^{*t}H_{(n)}^{-1}\right) = \tilde{S} f(x)(1-\rho)^2\sigma_e^2, \tag{3.4}$$

where $\tilde{S}$ *is the matrix* $(p+1)\times(p+1)$ *whose* $(i+1, j+1)$*th element is* $\tilde{s}_{i,j} = \nu_{i+j}$, $i,j = 0,\ldots,p$, *with* $\nu_r = \int u^r K^2(u)\,du$.

Now, using these propositions, the mean squared convergence of $\tilde{\beta}_G(x)$ can be established. For this, from (2.7) and (2.10), we obtain

$$E\left(\tilde{\beta}_G(x)\right) - \vec{\beta}(x) = \frac{m^{(p+1)}(x)}{(p+1)!}\tilde{C}_{(n)}^{-1}\begin{pmatrix} \tilde{c}_{0,p+1}^{(n)} \\ \tilde{c}_{1,p+1}^{(n)} \\ \vdots \\ \tilde{c}_{p,p+1}^{(n)} \end{pmatrix} + o\left(h_n^{p+1}\right)\vec{1}, \tag{3.5}$$

where $\vec{1} = (1, \ldots, 1)^t$.

From this equation and using Proposition 3.1 and A2, the asymptotic bias of $\vec{\beta}(x)$ is obtained.

Corollary 3.1. *Under assumptions in Proposition 3.1, we have*

$$H_{(n)} \left(E \left(\tilde{\beta}_G(x) \right) - \vec{\beta}\,(x) \right) = \frac{m^{(p+1)}(x)}{(p+1)!} h_n^{p+1} S^{-1} \vec{\mu} + o \left(h_n^{p+1} \right) \vec{1}, \tag{3.6}$$

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

With regard to the variance, from

$$Var \left(\sqrt{nh_n} H_{(n)} \tilde{\beta}_G(x) \right) = nh_n H_{(n)} E \left(\tilde{C}_{(n)}^{-1} \tilde{G}_{(n)}^{*} \tilde{G}_{(n)}^{*t} \tilde{C}_{(n)}^{-1t} \right) H_{(n)}, \tag{3.7}$$

and Propositions 3.1 and 3.2, the following corollary is obtained.

Corollary 3.2. *Under assumptions of Proposition 3.1, we have*

$$Var \left(H_{(n)} \tilde{\beta}_G(x) \right) = \frac{\sigma_\varepsilon^2}{nh_n f(x)} \frac{1+\rho}{1-\rho} S^{-1} \tilde{S} S^{-1} + o \left(\frac{1}{nh_n} \right). \tag{3.8}$$

Now, the asymptotic normality of $\tilde{\beta}_G(x)$ is obtained, but first the asymptotic normality of $\tilde{G}_{(n)}^{*}$ is established. Moreover, an additional assumption is necessary

A.5. $\lim\limits_{n\to\infty} nh_n^{2p+3} = C < \infty$.

Proposition 3.3. *Under assumptions A1-A5, the following holds*

$$\sqrt{nh_n} H_{(n)}^{-1} \tilde{G}_{(n)}^{*} \xrightarrow{\mathcal{L}} N_{(p+1)} \left(\vec{0}, \tilde{S} f(x) \left(1-\rho\right)^2 \sigma_e^2 \right), \tag{3.9}$$

where $N_{(p+1)}(\cdot,\cdot)$ *denotes the* $(p+1)$*-variate normal distribution.*

From this result and corresponding Proposition 3.1, the asymptotic normality of estimator $\tilde{\beta}_G(x)$ is obtained.

Theorem 3.1. *Under the assumptions of Proposition 3.3, the following holds:*

$$\sqrt{nh_n}\left[H_{(n)}\left(\tilde{\beta}_G(x)-\vec{\beta}(x)\right)-\frac{m^{(p+1)}(x)}{(p+1)!}h_n^{p+1}S^{-1}\vec{\mu}\right]\xrightarrow{\mathcal{L}}N_{(p+1)}\left(\vec{0},\Sigma\right), \tag{3.10}$$

where $\Sigma=\dfrac{\sigma_\varepsilon^2}{f(x)}\dfrac{1+\rho}{1-\rho}S^{-1}\widetilde{S}S^{-1}=\dfrac{\sigma_e^2}{f(x)}\dfrac{1}{(1-\rho)^2}S^{-1}\widetilde{S}S^{-1}.$

On the other hand, the convergence in probability of the estimator $\hat{\rho}_{(n)}$, given in (2.11), is obtained following similar arguments to those in (Stute, 1995, Corollary 2.5).

Proposition 3.4. *Under assumptions A1, A2 and A4, one gets*

$$\hat{\rho}_{(n)}\to\rho,\quad as\quad n\to\infty,\quad \textit{with probability 1.} \tag{3.11}$$

Now, using Proposition 3.4 and A5, the convergence to zero of the term $\hat{\beta}_F(x)-\tilde{\beta}_G(x)$ is obtained.

Proposition 3.5. *Under assumptions A1-A5, we have*

$$\sqrt{nh_n}H_{(n)}\left(\hat{\beta}_F(x)-\tilde{\beta}_G(x)\right)\to 0,\quad as\quad n\to\infty,\quad \textit{with probability 1.} \tag{3.12}$$

Finally, from (3.10) and (3.12), the asymptotic normality of $\hat{\beta}_F(x)$ is established.

Theorem 3.2. *Under assumptions A1-A5, we obtain*

$$\sqrt{nh_n}\left[H_{(n)}\left(\hat{\beta}_F(x)-\vec{\beta}(x)\right)-\frac{m^{(p+1)}(x)}{(p+1)!}h_n^{p+1}S^{-1}\vec{\mu}\right]\xrightarrow{\mathcal{L}}N_{(p+1)}\left(0,\Sigma\right). \tag{3.13}$$

From these results, it can be observed that $\hat{\beta}_F(x)$ presents the same asymptotic properties as the LP estimator $\hat{\beta}_L(x)$ defined in (2.2). The asymptotic properties of $\hat{\beta}_L(x)$, with respect to model (1.1), were studied by (Francisco-Fernández and Vilar-Fernández, 2001, Theorem 2), and the asymptotic properties of a general nonparametric estimator regression function $\hat{\beta}_0(x) = \hat{m}_n(x)$, under general dependence conditions on the errors, $\varepsilon_{t,n}$, were studied in Roussas et al. (1992), and Tran et al. (1996).

From Theorem 3.2 the asymptotic normality of the individual components $\hat{\beta}_F^{(j)}(x) = \dfrac{\hat{m}_F^{(j)}(x)}{j!}$ can be obtained.

Corollary 3.3. *Under assumptions A1-A5, we have*

$$\sqrt{nh_n^{1+2j}}\left[\left(\hat{m}_F^{(j)}(x) - m^{(j)}(x)\right) - h_n^{p+1-j}\frac{m^{(p+1)}(x)}{(p+1)!}j!B_j\right] \xrightarrow{\mathcal{L}} N\left(0, \sigma_j^2\right), \tag{3.14}$$

where $\sigma_j^2 = \dfrac{\sigma_\varepsilon^2}{f(x)}\dfrac{1+\rho}{1-\rho}(j!)^2 V_j$ *and the terms* B_j *and* V_j *denote, respectively, the* j*th element of* $S^{-1}\vec{\mu}$ *and the* j*th diagonal element of* $S^{-1}\tilde{S}S^{-1}$.

Once the asymptotic properties of this new estimator have been obtained, the following step in our study is to give some guides about the important problem of bandwidth selection. In this point, two plug-in techniques are proposed to obtain the smoothing parameter which can provide local or global bandwidths, depending on the choice of a local or a global measure of the estimation error.

The former of these methods is called asymptotic plug-in and it consists of finding the bandwidths that minimize the asymptotic mean squared error ($AMSE$) or the asymptotic mean integrated squared error ($AMISE$) -depending on the use of local or global bandwidths- and then substituting the unknown quantities that appear in these bandwidths by some estimators. In this particular case, using the asymptotic expressions of the bias and the variance, given in (3.14), the $AMSE$ of $\hat{m}_F^{(j)}(x)$ is given by:

$$AMSE\left(\hat{m}_F^{(j)}(x)\right) = \left(h_n^{p+1-j}\frac{m^{(p+1)}(x)}{(p+1)!}j!B_j\right)^2 + \frac{1}{nh_n^{2j+1}}\frac{\sigma_\varepsilon^2}{f(x)}\frac{1+\rho}{1-\rho}(j!)^2 V_j. \tag{3.15}$$

So, minimizing expression (3.15) in h_n, the asymptotically optimal local bandwidth to estimate the jth derivative of the regression function is obtained. This bandwidth is given by:

$$h_{j,l,as}^{opt}(x) = C_{j,p}(K)\left(\frac{\sigma_\varepsilon^2\left(\frac{1+\rho}{1-\rho}\right)}{n(m^{(p+1)}(x))^2 f(x)}\right)^{1/(2p+3)}, \qquad (3.16)$$

where $C_{j,p}(K)$ is a real number that depends on the kernel K.

On the other hand, using the $AMISE$, we can obtain the asymptotically optimal global bandwidth to estimate the jth derivative of the regression function given by:

$$h_{j,g,as}^{opt}(x) = C_{j,p}(K)\left(\frac{\sigma_\varepsilon^2\left(\frac{1+\rho}{1-\rho}\right)}{n\int(m^{(p+1)}(x))^2 f(x)dx}\right)^{1/(2p+3)}. \qquad (3.17)$$

In (3.16) and (3.17), there are four unknown quantities: σ_ε^2, ρ, $m^{(p+1)}(x)$ and $\int(m^{(p+1)}(x))^2 f(x)dx$. These must be estimated to produce practical smoothing parameters from (3.16) and (3.17). In Francisco-Fernández and Vilar-Fernández (2001), some ideas to that aim are presented.

Another possibility is to design a bandwidth selection procedure based on using the exact expressions of the bias and the variance of the estimator. This technique follows the same idea as the exact plug-in method proposed for the local polynomial estimator in Fan and Gijbels (1996). For the estimator $\tilde{\beta}_G(x)$, given in (2.10), the exact expressions of the bias and the variance are given by:

$$Bias\left(\tilde{\beta}_G(x)\right) = \tilde{C}_{(n)}^{-1} X_{(n)}^t P_{(n)}^t W_{(n)} P_{(n)} \vec{R}_{(n)}$$

and

$$Var\left(\tilde{\beta}_G(x)\right) = \sigma_\varepsilon^2\left(1-\rho^2\right)\tilde{C}_{(n)}^{-1} X_{(n)}^t P_{(n)}^t W_{(n)}^2 P_{(n)} X_{(n)} \tilde{C}_{(n)}^{-1},$$

where $\vec{R}_{(n)} = \vec{M}_{(n)} - X_{(n)}\vec{\beta}(x)$, $P_{(n)}$, σ_ε^2 and ρ are unknown. Using, for example, the approximations of these quantities given in Francisco-Fernández and Vilar-Fernández (2001) and changing $P_{(n)}$ by $\hat{P}_{(n)}$, local or global bandwidths to estimate the jth derivative of the regression function can be obtained. Obviously, the same bandwidths can be used with $\hat{m}_F^{(j)}(x)$, an estimator of the jth regression function derivative, $j = 0, 1, \ldots, p$, obtained from $\hat{\beta}_F(x)$. These kind of smoothing parameters are called local or global exact plug-in bandwidths.

4 Simulation study

In this section, the performance of the proposed nonparametric estimator defined in (2.12) is illustrated. It will be seen that $\hat{\beta}_F$ behaves adequately for regression curve estimation and it is better than other kernel type estimators under the Mean Integrated Square Error (MISE) criterion. For this purpose, a simulation study was carry out to compare the following estimators of the regression function: the Nadaraya-Watson (NW), the Gasser-Müller (GM), the local linear (LL) and the new feasible local linear (FLL) estimators. The simulation studies presented here are representative of many others performed by the authors.

We simulated $B = 200$ samples of size n from a fixed and equally spaced model in the interval $[0, 1]$ with random errors following an $AR(1)$ process with $N(0, \sigma^2)$ distribution. Optimal bandwidths by minimizing the Average Squared Error (ASE) were computed. Using Montecarlo approximations, the integrated squared bias, the integrated variance and the $MISE$ for each of the four estimators were then approximated. In the study, to avoid possible boundary effects, two situations have been considered: estimation of the regression function in interval $[0, 1]$ (global region) and estimation of the function in the central region, $[0.3, 0.7]$. In each case, the obtained optimal bandwidths are different. The kernel function used was the quartic kernel ($K(u) = \frac{15}{16}(1 - u^2)^2$, if $|u| \leq 1$).

As explained in Section 2, to compute $\hat{\beta}_F(x)$ it is necessary to estimate the parameter ρ. This was achieved by using the estimator given in (2.11), with a pilot bandwidth empirically determined. The results are presented in the following tables.

In Table 1, the regression function is $m(x) = \sin(\pi x)$, the sample size is $n = 300$ and the error is $AR(1)$, with $\sigma = 0.1$ and two values for the autocorrelation coefficient, $\rho = 0.9$ (strong dependence) and $\rho = 0.0$ (independence) have been considered.

In Table 2, all the parameters are the same as in Table 1 but the re-

$\rho = \mathbf{0.9}$		**NW**	**GM**	**LL**	**FLL**
Central	Mean h_{opt}	0.1843	0.1842	0.1842	0.3145
	$\int Bias^2$	0.00080	0.00080	0.00080	0.00036
	$\int Variance$	0.00167	0.00167	0.00167	0.00154
	$MISE$	0.00247	0.00247	0.00247	0.00190
Global	Mean h_{opt}	0.1226	0.1695	0.2017	0.3137
	$\int Bias^2$	0.00091	0.00073	0.00062	0.00043
	$\int Variance$	0.00301	0.00237	0.00220	0.00215
	$MISE$	0.00392	0.00310	0.00282	0.00258
$\rho = \mathbf{0.0}$		**NW**	**GM**	**LL**	**FLL**
Central	Mean h_{opt}	0.1086	0.1085	0.1085	0.1115
	$\int Bias^2$	0.00008	0.00008	0.00008	0.00008
	$\int Variance$	0.00019	0.00019	0.00019	0.00019
	$MISE$	0.00026	0.00026	0.00026	0.00026
Global	Mean h_{opt}	0.0645	0.0870	0.1203	0.1243
	$\int Bias^2$	0.00016	0.00009	0.00006	0.00007
	$\int Variance$	0.00036	0.00034	0.00023	0.00024
	$MISE$	0.00052	0.00043	0.00030	0.00031

Table 1: $m(x) = \sin(\pi x)$, $n = 300$, $\sigma = 0.1$, $\rho = 0.9$ *and* $\rho = 0.0$.

$\rho = \mathbf{0.9}$		**NW**	**GM**	**LL**	**FLL**
Central	Mean h_{opt}	0.2070	0.2070	0.2070	0.3649
	$\int Bias^2$	0.00192	0.00192	0.00192	0.00066
	$\int Variance$	0.00191	0.00191	0.00191	0.00154
	$MISE$	0.00383	0.00383	0.00383	0.00220
Global	Mean h_{opt}	0.2016	0.2034	0.2099	0.3642
	$\int Bias^2$	0.00213	0.00135	0.00100	0.00092
	$\int Variance$	0.00223	0.00226	0.00236	0.00220
	$MISE$	0.00436	0.00361	0.00336	0.00312

Table 2: $m(x) = 16x^2(1-x)^2$, $n = 300$, $\sigma = 0.1$, $\rho = 0.9$.

gression function is $m(x) = 16x^2(1-x)^2$ and $\rho = 0.9$.

To observe the importance of the correlation coefficient estimator used in the FLL estimator, in the following two tables, the results for the new generalized local lineal estimator (GLL) obtained from $\tilde{\beta}_G(x)$ are included, where the theoretical correlation coefficient, ρ, is used.

In Table 3 we have considered $m(x) = 1 + 5x$, $n = 100$, $\sigma = 0.5$ and $\rho = 0.9$

$\rho = \mathbf{0.9}$		**NW**	**GM**	**LL**	**FLL**	**GLL**
Central	Mean h_{opt}	0.4970	0.6879	0.7887	0.7951	0.8055
	$\int Bias^2$	0.00434	0.00275	0.00069	0.00049	0.00042
	$\int Variance$	0.05882	0.03801	0.04727	0.04269	0.04010
	$MISE$	0.06316	0.04076	0.04796	0.04317	0.04052
Global	Mean h_{opt}	0.2817	0.4421	0.8171	0.8715	0.8578
	$\int Bias^2$	0.02411	0.01803	0.00084	0.00078	0.00053
	$\int Variance$	0.08898	0.07022	0.07475	0.07041	0.06977
	$MISE$	0.11309	0.08826	0.07559	0.07110	0.07030

Table 3: $m(x) = 1 + 5x$, $n = 100$, $\sigma = 0.5$, $\rho = 0.9$.

Finally, in Table 4, the model used was $m(x) = x^3$, $n = 300$, $\sigma = 0.1$ and $\rho = 0.9$. In general, the FLL estimator performed better than the other estimators studied. The broad study showed that this improvement is greater when ρ is close to one. Furthermore, when ρ is near zero or negative, the new estimator is not worse than the LL estimator, as observed in Table 1.

$\rho = \mathbf{0.9}$		**NW**	**GM**	**LL**	**FLL**	**GLL**
Central	Mean h_{opt}	0.29560	0.29220	0.29320	0.31390	0.31610
	$\int\int Bias^2$	0.00040	0.00043	0.00042	0.00041	0.00042
	$\int Variance$	0.00080	0.00079	0.00080	0.00081	0.00080
	$MISE$	0.00120	0.00122	0.00122	0.00122	0.00122
Global	Mean h_{opt}	0.1656	0.2587	0.3105	0.3278	0.3282
	$\int Bias^2$	0.00076	0.00047	0.00032	0.00030	0.00031
	$\int Variance$	0.00234	0.00171	0.00163	0.00158	0.00157
	$MISE$	0.00310	0.00218	0.00195	0.00188	0.00188

Table 4: $m(x) = x^3$, $n = 300$, $\sigma = 0.1$, $\rho = 0.9$.

A figure is also presented from which similar conclusions can be deduced. Figure 1 shows the plot of $\frac{MISE(Est)}{MISE(NW)} \times 100$ versus ρ, where $MISE(Est)$ is the $MISE$ of FLL, LL and GM estimators in central region and $MISE(NW)$ is the $MISE$ of NW estimator. The model considered in this figure is the same as that in Table 2, that is, $m(x) = 16x^2(1-x)^2$,

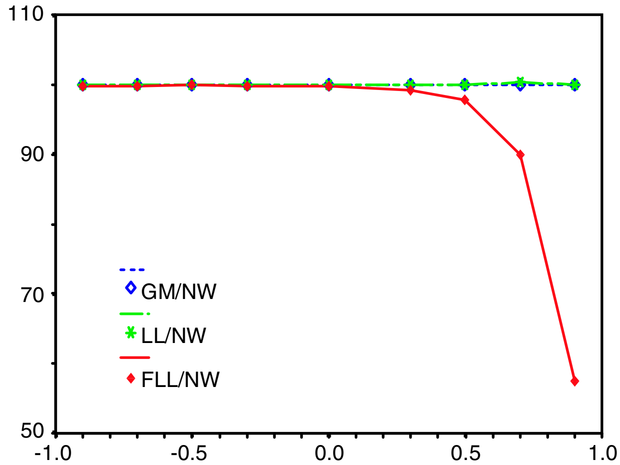

Figure 1: *Plot of* $\frac{MISE(Est)}{MISE(NW)} \times 100$ *versus* ρ, *with* Est = GM, LL *and* FLL. $m(x) = 16x^2(1-x)^2$, $n = 300$, $\sigma = 0.1$.

$n = 300$, $\sigma = 0.1$, ρ goes from -0.9 to 0.9.

Once more, to study the influence of the correlation coefficient estimator in the FLL estimator obtained from $\hat{\beta}_F(x)$, another figure (Figure 2) is presented, where the relative efficiency between the FLL estimator and the ideal GLL estimator as a function of ρ, with the same model as in Figure 1, is shown.

It can be seen, in Figure 1, that the best behavior of the FLL estimator is obtained when ρ is close to one, but, on the other hand, the plot shown in Figure 2 indicates that when ρ decreases, the difference between the MISE of the FLL estimator and the MISE of the GLL estimator becomes larger. So, if we had been more careful in the problem of estimating ρ for all the values of the correlation coefficient -and therefore the FLL estimator would be closer to the ideal GLL estimator-, the improvements of the FLL estimator would have been greater for any value of ρ.

Another interesting point is that when ρ increases, the MISE of the estimators also increases. Thus, the MISE associated with negative ρ models can be slightly lower than in the context of independent data ($\rho = 0$). This is due to the behavior of the variance and it is compatible with the asymptotic expression obtained in Section 3 (see Corollary 3.3).

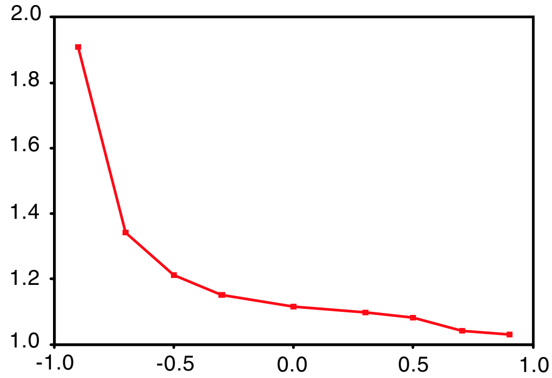

Figure 2: *Plot of* $\frac{MISE(FLL)}{MISE(GLL)}$ *versus* ρ. $m(x) = 16x^2(1-x)^2$, $n = 300$, $\sigma = 0.1$.

In conclusion, the results of Section 3 show that estimators $\hat{\beta}_L$ and $\hat{\beta}_F$ have the same asymptotic properties. But, from the numeric study presented in this section we can deduce that $\hat{\beta}_F$ has a better behavior than $\hat{\beta}_L$ for finite samples when ρ is large.

5 Proofs

In this section, we sketch the proofs of the results presented in Section 2. In what follows, the letter C will be used to indicate generic constants whose values are not important and may vary. We will use the following result of Francisco-Fernández and Vilar-Fernández (2001) (see Propositions 1 and 2 in this paper),

Proposition 5.1. *Under assumptions A1-A4, we have*

$$\lim_{n\to\infty} h_n^{-j} s_j^{(n)} = f(x)\mu_j, \quad 0 \le j \le 2p+1, \tag{5.1}$$

and

$$\lim_{n\to\infty} nh_n Cov\left(h_n^{-i} t^*_{i,(n)}, h_n^{-j} t^*_{j,(n)}\right) = \nu_{j+i} f(x) \frac{\sigma_e^2}{(1-\rho)^2}, \quad 0 \le j, i \le p, \tag{5.2}$$

where $t^*_{i,(n)} = t_{i,(n)} - E\left(t_{i,(n)}\right)$.

Proof of Proposition 3.1. Let $\tilde{c}_{i,j}^{(n)}$, $i,j = 0,...,p$ be the terms of matrix $\tilde{C}_{(n)} = X_{(n)}^t P_{(n)}^t W_{(n)} P_{(n)} X_{(n)}$, then

$$\begin{aligned}
\tilde{c}_{i,j}^{(n)} &= \frac{1}{n}\sum_{r=1}^{n}(x_r - x)^{i+j} K_h(x_r - x) - \frac{\rho^2}{n}(x_1 - x)^{i+j} K_h(x_1 - x) \\
&\quad -\frac{\rho}{n}\sum_{r=2}^{n}(x_r - x)^i (x_{r-1} - x)^j K_h(x_r - x) \\
&\quad -\frac{\rho}{n}\sum_{r=2}^{n}(x_r - x)^j (x_{r-1} - x)^i K_h(x_r - x) \\
&\quad +\frac{\rho^2}{n}\sum_{r=2}^{n}(x_{r-1} - x)^{i+j} K_h(x_r - x) \\
&= \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4 + \Delta_5. \qquad (5.3)
\end{aligned}$$

Thus,

$$\Delta_1 = s_{i,j}^{(n)}, \qquad (5.4)$$

with $s_{i,j}^{(n)}$ defined in (2.4). With regard to the other terms, using assumptions A2 and A4, and (1.2), we obtain

$$\Delta_2 h_n^{-i-j} \leq C\frac{1}{nh_n} = o(1), \qquad (5.5)$$

$$\Delta_3 = \Delta_4 = -\rho s_{i,j}^{(n)} - \frac{\Delta_2}{\rho} - \rho\sum_{k=1}^{j}\binom{j}{k} s_{i+j-k}^{(n)} n^{-k}, \qquad (5.6)$$

$$\Delta_5 = \rho^2 s_{i,j}^{(n)} + \Delta_2 + \rho^2\sum_{k=1}^{i+j}\binom{i+j}{j} s_{i+j-k}^{(n)} n^{-k}. \qquad (5.7)$$

Consequently we have

$$\tilde{c}_{i,j}^{(n)} h_n^{-i-j} = (1-\rho)^2 s_{i,j}^{(n)} + o(1), \qquad (5.8)$$

and using 5.1, equation (5.1), the proof is concluded. □

Proof of Proposition 3.2. Let us denote $\tilde{g}_{j,(n)}^*$, $j = 0,1,...,p$ to the terms of the vector

$$\tilde{G}_{(n)}^* = X_{(n)}^t P_{(n)}^t W_{(n)} P_{(n)}\left(\vec{Y}_{(n)} - \vec{M}_{(n)}\right),$$

therefore

$$\tilde{g}^*_{j,(n)} = \frac{1-\rho^2}{n}(x_1-x)^j K_h(x_1-x)\varepsilon_1 + \frac{1}{n}\sum_{r=2}^{n}\left((x_r-x)^j - \rho(x_{r-1}-x)^j\right) K_h(x_r-x)e_r. \tag{5.9}$$

Taking into account the independence between ε_1 and e_r, $r > 1$, for $i, j = 0, 1, ..., p$, and that variables e_r are independent, we have

$$Cov\left(h_n^{-j}\tilde{g}^*_{j,(n)}, h_n^{-i}\tilde{g}^*_{i,(n)}\right) = G_1 + G_2, \tag{5.10}$$

with

$$\begin{aligned} G_1 &= \frac{\left(1-\rho^2\right)^2}{n^2 h_n^{j+i}}(x_1-x)^{i+j} K_h^2(x_1-x)\sigma_\varepsilon^2 \\ &\leq C\frac{\left(1-\rho^2\right)^2\sigma_\varepsilon^2}{\sigma_\varepsilon^2 n^2 h_n^2} = o\left(n^{-1}h_n^{-1}\right), \end{aligned} \tag{5.11}$$

where the assumptions A2, A4 and (1.2) have been used, and

$$\begin{aligned} G_2 &= \frac{\sigma_e^2}{n^2 h_n^{j+i}}\sum_{r=2}^{n}\left((x_r-x)^j - \rho(x_{r-1}-x)^j\right) \\ &\quad\times\left((x_r-x)^i - \rho(x_{r-1}-x)^i\right) K_h^2(x_r-x). \end{aligned} \tag{5.12}$$

G_2 can be split as follows:

$$\begin{aligned} G_2 &= \frac{\sigma_e^2}{n^2 h_n^{j+i}}\sum_{r=2}^{n}\left((x_r-x)^{i+j} + \rho^2(x_{r-1}-x)^{i+j}\right) K_h^2(x_r-x) \\ &\quad - \frac{\rho\sigma_e^2}{n^2 h_n^{j+i}}\sum_{r=2}^{n}\left((x_r-x)^j(x_{r-1}-x)^i + (x_{r-1}-x)^j(x_r-x)^i\right) \\ &\quad \times K_h^2(x_r-x) \\ &= G_{21} + G_{22}. \end{aligned}$$

Taking into account A2 and A4, and using Riemann approximations, we have

$$\begin{aligned}
&\frac{1}{n^2 h_n^{j+i}} \sum_{r=2}^{n} (x_r - x)^{i+j} K_h^2 (x_r - x) \\
&= \frac{1}{n h_n^2} \int \left(\frac{u - x}{h_n} \right)^{i+j} K^2 \left(\frac{u - x}{h_n} \right) f(u) du + O\left(n^{-2} h_n^{-2}\right) \\
&= \frac{1}{n h_n} f(x) \nu_{i+j} + o\left(n^{-1} h_n^{-1}\right).
\end{aligned}$$

On the other hand, we obtain that

$$\begin{aligned}
&\frac{1}{n^2 h_n^{i+j}} \sum_{r=2}^{n} (x_{r-1} - x)^{i+j} K_h^2 (x_r - x) \\
&= \frac{1}{n^2 h_n^{i+j}} \sum_{r=2}^{n} (x_r - x)^{i+j} K_h^2 (x_r - x) + R,
\end{aligned}$$

where R is a residual part, with

$$\begin{aligned}
R &= \frac{1}{n^2 h_n^{i+j}} \sum_{r=2}^{n} \sum_{s=1}^{i+j} \binom{i+j}{s} (x_{r-1} - x_r)^s (x_r - x)^{i+j-s} K_h^2 (x_r - x) \\
&\leq O\left(\frac{1}{n^2 h_n^2} \right),
\end{aligned}$$

which allows us to conclude that

$$G_{21} = \left(1 + \rho^2\right) \sigma_e^2 f(x) \nu_{i+j} \frac{1}{n h_n} + o\left(\frac{1}{n h_n} \right). \tag{5.13}$$

With respect to G_{22}, using similar arguments, we obtain that

$$\begin{aligned}
&\frac{1}{n^2 h_n^{j+i}} \sum_{r=2}^{n} (x_r - x)^j (x_{r-1} - x)^i K_h^2 (x_r - x) \\
&= \frac{1}{n^2 h_n^{j+i}} \sum_{r=2}^{n} (x_r - x)^{j+i} K_h^2 (x_r - x) + R',
\end{aligned}$$

with $R' \leq O\left(n^{-2} h_n^{-2}\right)$. So,

$$G_{22} = -2\rho\sigma_e^2 f(x) \nu_{i+j} \frac{1}{n h_n} + o\left(\frac{1}{n h_n} \right). \tag{5.14}$$

Now, we conclude that

$$Cov\left(h_n^{-j}\tilde{g}^*_{j,(n)}, h_n^{-i}\tilde{g}^*_{i,(n)}\right) = (1-\rho)^2\,\sigma_e^2 f(x)\,\nu_{i+j}\frac{1}{nh_n} + o\left(\frac{1}{nh_n}\right), \quad (5.15)$$

and (3.4) is a consequence of (5.15). □

Proof of Proposition 3.3. Let Q_n be an arbitrary linear combination of $h_n^{-i}\tilde{g}^*_{i,(n)}$

$$Q_n = \sum_{i=0}^{p} \alpha_i\, h_n^{-i}\tilde{g}^*_{i,(n)}, \qquad \text{with } \alpha_i \in \mathbb{R}. \quad (5.16)$$

By (5.9), $\sqrt{nh_n}Q_n$ can also be written in the form

$$\sqrt{nh_n}Q_n = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \xi_{t,n}, \quad (5.17)$$

where

$$\xi_{1,n} = \left(1-\rho^2\right) h_n^{1/2} K_h(x_1 - x) \sum_{i=0}^{p} \alpha_i (x_1 - x)^i\, h_n^{-i}\varepsilon_1$$

and if $t = 2, \ldots, n$,

$$\xi_{t,n} = h_n^{1/2} K_h(x_t - x)\left(\sum_{i=0}^{p} \alpha_i \left((x_t - x)^i - \rho\,(x_{t-1} - x)^i\right) h_n^{-i}\right) e_t.$$

If the asymptotic normality of $\sqrt{nh_n}Q_n$ is established, then (3.9) follows from the Cramer-Wold Theorem. For this, from (5.9), it follows that $E(Q_n) = 0$ and using Proposition 3.2, we obtain that

$$\lim_{n\to\infty} Var\left(\sqrt{nh_n}Q_n\right) = (1-\rho)^2\,\sigma_e^2 f(x)\sum_{i=0}^{p}\sum_{j=0}^{p} \alpha_i\alpha_j\nu_{i+j} = \sigma_Q^2 < \infty. \quad (5.18)$$

Now, we prove that the standard Lindenberg-Feller condition is satisfied. Here, this condition takes the form

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} E\left(\xi_{t,n}^2 I\left(|\xi_{t,n}| \geq \epsilon\sigma_Q\sqrt{n}\right)\right) = 0,\ \forall\epsilon > 0,\ \epsilon \in \mathbb{R}.$$

For this purpose, we employ a truncation argument. Let M be a fixed truncation point and let us denote

$$e_{t,M} = e_t I\left(|e_t| \leq M\right), \quad t = 2, \ldots, n$$

(the same is done with ε_1). We have that $e_t = e_{t,M} + \tilde{e}_{t,M}$, with $\tilde{e}_{t,M} = I\left\{|e_t| > M\right\}$. Replacing e_t by $e_{t,M}$ (or $\tilde{e}_{t,M}$), we can write $Q_{n,M}$ (or $\tilde{Q}_{n,M}$), as the same linear combination as Q_n. Now, reasoning as in (5.18), we have that

$$\lim_{n\to\infty} Var\left(\sqrt{nh_n}Q_{n,M}\right) = (1-\rho)^2\,\sigma^2_{e,M} f(x) \sum_{i=0}^{p}\sum_{j=0}^{p} \alpha_i\alpha_j\nu_{i+j} = \sigma^2_{Q,M} < \infty,$$

where $\sigma^2_{e,M} = Var\left(e_{t,M}\right)$. By assumptions A2 and A4, we have

$$|\xi_{t,n,M}| \leq \frac{C}{\sqrt{h_n}} M,$$

where $\xi_{t,n,M}$ is obtained replacing e_t by $e_{t,M}$ in $\xi_{t,n}$. Hence

$$\max_{1\leq t\leq n} \frac{1}{\sqrt{n}} |\xi_{t,n,M}| \leq \frac{C}{\sqrt{nh_n}} M \to 0, \quad \text{as } n \to \infty.$$

Therefore, $\left\{\xi^2_{t,n,M} I\left\{|\xi_{t,n,M}| \geq \epsilon\sigma_{Q,M}\sqrt{n}\right\}\right\}$ is an empty set when n is large enough, and the Lindenberg-Feller condition is satisfied. So

$$\sqrt{nh_n}Q_{n,M} \xrightarrow{\mathcal{L}} N\left(0, \sigma^2_{Q,M}\right). \tag{5.19}$$

In order to complete the proof, it suffices to show that

$$\varphi_{Q_n}(t) \longrightarrow \varphi_Z^{\sigma^2_Q}(t), \quad \text{as } n \to \infty, \tag{5.20}$$

where $\varphi_{Q_n}(t)$ and $\varphi_Z^{\sigma^2_Q}(t)$ denote the characteristic functions of $\sqrt{nh_n}\,Q_n$ and of the distribution $N(0, \sigma^2_Q)$, respectively.

We have

$$\begin{aligned}\left|\varphi_{Q_n}(t) - \varphi_Z^{\sigma^2_Q}(t)\right| \;&\leq\; \left|\varphi_{Q_{n,M}}(t)\right| \left|\varphi_{\tilde{Q}_{n,M}}(t) - 1\right| + \left|\varphi_Z^{\sigma^2_{Q,M}}(t) - \varphi_Z^{\sigma^2_Q}(t)\right| \\ &+\; \left|\varphi_{Q_{n,M}}(t) - \varphi_Z^{\sigma^2_{Q,M}}(t)\right| \equiv \Delta_1 + \Delta_2 + \Delta_3.\end{aligned}$$

As $n \to \infty$, the terms Δ_1 and Δ_2 tend to zero by the dominated convergence theorem when $M \uparrow \infty$. And the convergence to zero of Δ_3 follows from (5.19) and the Levy theorem, for every $M > 0$. Now, the proof of Proposition 3.3 is complete. □

Proof of Proposition 3.5 To prove Proposition 3.5, we will use the following auxiliary results.

Lemma 5.1. *Under assumptions A1-A4, we have*

$$H_{(n)}^{-1} X_{(n)}^t \left(\hat{\Omega}_W^{-1} - \Omega_W^{-1} \right) X_{(n)} H_{(n)}^{-1} = o_P(1), \tag{5.21}$$

where $\Omega_W^{-1} = P_{(n)}^t W_{(n)} P_{(n)}$ and $\hat{\Omega}_W^{-1} = \hat{P}_{(n)}^t W_{(n)} \hat{P}_{(n)}$.

Proof. Taking into account the form of the matrix $\hat{\Omega}_W^{-1} - \Omega_W^{-1}$, and applying Propositions 3.1 and 3.4, the conclusion of the lemma is deduced. □

Lemma 5.2. *Under assumptions A1-A4, we have*

$$\sqrt{nh_n} H_{(n)}^{-1} X_{(n)}^t \Omega_W^{-1} \vec{\varepsilon}_{(n)} = O_P(1). \tag{5.22}$$

Proof. From the form of the matrix $P_{(n)}$, it is sufficient to prove that

$$\sqrt{nh_n} H_{(n)}^{-1} X_{(n)}^t W_{(n)} \vec{\varepsilon}_{(n)} = \sqrt{nh_n} H_{(n)}^{-1} \vec{T}_{(n)}^* = O_P(1),$$

where $\vec{T}_{(n)}^* = X_{(n)}^t W_{(n)} \left(\vec{Y}_{(n)} - \vec{M}_{(n)} \right)$. And, this is deduced from Proposition 5.1, statement (5.2). □

Using similar arguments to those in Lemma 5.1, the following lemma is obtained.

Lemma 5.3. *Under assumptions A1-A4, we have*

$$\sqrt{nh_n} H_{(n)}^{-1} X_{(n)}^t \left(\hat{\Omega}_W^{-1} - \Omega_W^{-1} \right) \vec{\varepsilon}_{(n)} = o_P(1). \tag{5.23}$$

Now we can prove Proposition 3.5.

From the definitions of $\tilde{\beta}_G(x)$ and $\hat{\beta}_F(x)$, the regression model (1.1) and (2.6), we have that

$$\sqrt{nh_n}H_{(n)}\left(\hat{\beta}_F(x)-\tilde{\beta}_G(x)\right)=\Gamma_1+\Gamma_2+o\left(1\right), \tag{5.24}$$

where

$$\begin{aligned}\Gamma_1 = \quad & \frac{m^{(p+1)}\left(x\right)}{(p+1)!}\sqrt{nh_n}H_{(n)} \\ & \times\left[\left(X_{(n)}^t\hat{\Omega}_W^{-1}X_{(n)}\right)^{-1}X_{(n)}^t\hat{\Omega}_W^{-1}\begin{pmatrix}(x_1-x)^{p+1}\\ \vdots \\ (x_n-x)^{p+1}\end{pmatrix}\right. \\ & \left.-\left(X_{(n)}^t\Omega_W^{-1}X_{(n)}\right)^{-1}X_{(n)}^t\Omega_W^{-1}\begin{pmatrix}(x_1-x)^{p+1}\\ \vdots \\ (x_n-x)^{p+1}\end{pmatrix}\right],\end{aligned}$$

and

$$\begin{aligned}\Gamma_2 \quad = \quad & \sqrt{nh_n}H_{(n)}\left[\left(X_{(n)}^t\hat{\Omega}_W^{-1}X_{(n)}\right)^{-1}X_{(n)}^t\hat{\Omega}_W^{-1}\vec{\varepsilon}_{(n)}\right. \\ & \left.-\left(X_{(n)}^t\Omega_W^{-1}X_{(n)}\right)^{-1}X_{(n)}^t\Omega_W^{-1}\vec{\varepsilon}_{(n)}\right].\end{aligned}$$

Now, the convergence in probability to zero of Γ_1 and Γ_2 is proved. In the first place, using assumptions A1 and A2, we have

$$\begin{aligned}\Gamma_1 \quad = \quad & C\sqrt{nh_n}H_{(n)} \\ & \left[\left(X_{(n)}^t\hat{\Omega}_W^{-1}X_{(n)}\right)^{-1}X_{(n)}^t\left(\hat{\Omega}_W^{-1}-\Omega_W^{-1}\right)\right. \\ & \left.+\left(\left(X_{(n)}^t\hat{\Omega}_W^{-1}X_{(n)}\right)^{-1}-\left(X_{(n)}^t\Omega_W^{-1}X_{(n)}\right)^{-1}\right)X_{(n)}^t\Omega_W^{-1}\right] \\ & \times\begin{pmatrix}(x_1-x)^{p+1}\\ \vdots \\ (x_n-x)^{p+1}\end{pmatrix}.\end{aligned}$$

Applying Propositions 3.1 and 3.4, we obtain that

$$\Gamma_1=C\sqrt{nh_n}h_n^{p+1}\,o\left(1\right). \tag{5.25}$$

From Assumption A5, the convergence in probability to zero of Γ_1 follows.

With regard to the second term Γ_2, we can write

$$\begin{aligned}\Gamma_2 &= \sqrt{nh_n} H_{(n)} \left[\left(X_{(n)}^t \hat{\Omega}_W^{-1} X_{(n)} \right)^{-1} X_{(n)}^t \left(\hat{\Omega}_W^{-1} - \Omega_W^{-1} \right) \vec{\varepsilon}_{(n)} \right] \\ &\quad + \sqrt{nh_n} H_{(n)} \left[\left(\left(X_{(n)}^t \hat{\Omega}_W^{-1} X_{(n)} \right)^{-1} - \left(X_{(n)}^t \Omega_W^{-1} X_{(n)} \right)^{-1} \right) \right. \\ &\quad \left. X_{(n)}^t \Omega_W^{-1} \vec{\varepsilon}_{(n)} \right] \\ &= \Gamma_{2,1} + \Gamma_{2,2}.\end{aligned}$$

Using Proposition 3.1 and Lemma 5.3, we have

$$\begin{aligned}\Gamma_{2,1} &= \sqrt{nh_n} \left(H_{(n)}^{-1} \hat{C}_{(n)} H_{(n)}^{-1} \right)^{-1} H_{(n)}^{-1} X_{(n)}^t \left(\hat{\Omega}_W^{-1} - \Omega_W^{-1} \right) \vec{\varepsilon}_{(n)} \\ &= C \sqrt{nh_n} H_{(n)}^{-1} X_{(n)}^t \left(\hat{\Omega}_W^{-1} - \Omega_W^{-1} \right) \vec{\varepsilon}_{(n)} = o_P(1). \qquad (5.26)\end{aligned}$$

Finally, only the convergence to zero of $\Gamma_{2,2}$ remains to be proved. Using Lemmas 5.1 and 5.2, and Proposition 3.1, we derive that

$$\begin{aligned}\Gamma_{2,2} &= \left(\left[H_{(n)}^{-1} X_{(n)}^t \left(\hat{\Omega}_W^{-1} - \Omega_W^{-1} \right) X_{(n)} H_{(n)}^{-1} + H_{(n)}^{-1} X_{(n)}^t \Omega_W^{-1} X_{(n)} H_{(n)}^{-1} \right]^{-1} \right. \\ &\quad \left. - \left[H_{(n)}^{-1} X_{(n)}^t \Omega_W^{-1} X_{(n)} H_{(n)}^{-1} \right]^{-1} \right) \sqrt{nh_n} H_{(n)}^{-1} X_{(n)}^t \Omega_W^{-1} \vec{\varepsilon}_{(n)} \\ &= o_P(1). \qquad (5.27)\end{aligned}$$

From (5.24), (5.25), (5.26) and (5.27), the conclusion of Proposition 3.5 is deduced. □

Acknowledgments

The authors wish to thank an Associated Editor and a referee for their helpful comments and suggestions.

References

CLEVELAND, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368):829–836.

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

FAN, J. and GIJBELS, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Association, B*, 57(2):371–394.

FAN, J. and GIJBELS, I. (1996). *Local polynomial modelling and its applications*. Chapman & Hall, London.

FRANCISCO-FERNÁNDEZ, M. and VILAR-FERNÁNDEZ, J. M. (2001). Local polynomial regression estimation with correlated errors. *Communications in Statistics. Theory and Methods*, 30(7):1271–1293.

HÄRDLE, W. and TSYBAKOV, A. (1997). Local polynomial estimators of the volatility function in nonparametric autoregression. *Journal of Econometrics*, 81(1):223–242.

HÄRDLE, W., TSYBAKOV, A., and YANG, L. (1998). Nonparametric vector autoregression. *Journal of Statistical Planning and Inference*, 68(2):221–245.

HASTIE, T. and LOADER, C. (1993). Local regression: automatic kernel carpentry. *Statistical Science*, 8(2):120–143.

MASRY, E. (1996a). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *Journal of Time Series Analysis*, 17(6):571–599.

MASRY, E. (1996b). Multivariate regression estimation—local polynomial fitting for time series. *Stochastic Processes and their Applications*, 65(1):81–101.

MASRY, E. and FAN, J. (1997). Local polynomial estimation of regression functions for mixing processes. *Scandinavian Journal of Statistics. Theory and Applications*, 24(2):165–179.

Roussas, G., Tran, L., and Ionannides, D. (1992). Fixed design regression for time series: asymptotic normality. *Journal of Multivariate Analysis*, 40:262–291.

Ruppert, D., Sheather, S. J., and Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, 90(432):1257–1270.

Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression. *Annals of Statistics*, 22(3):1346–1370.

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

Stute, W. (1995). Bootstrap of a linear model with AR-error structure. *Metrika*, 42:395–410.

Tran, L., Roussas, G., Yakowitz, S., and Truong, V. (1996). Fixed-design regression for linear time series. *Annals of Statistics*, 24(3):975–991.

Tsybakov, A. B. (1986). Robust reconstruction of functions by the local approximation method. *Problems of Information Transmission*, 22:133–146.

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