

# LOCAL POLYNOMIAL REGRESSION ESTIMATION WITH CORRELATED ERRORS

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

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

*Key Words:* nonparametric estimators; local polynomial fitting; autoregressive process.

## ABSTRACT

In this paper, we study the nonparametric estimation of the regression function and its derivatives using weighted local polynomial fitting. Consider the fixed regression model and suppose that the random observation error is coming from a strictly stationary stochastic process. Expressions for the bias and the variance array of the estimators of the regression function and its derivatives are obtained and joint asymptotic normality is established. The influence of the dependence of the data is observed in the expression of the variance. We also propose a variable bandwidth selection procedure. A simulation study and an analysis with real economic data illustrate the proposed selection method.

## 1. INTRODUCTION

Recently, the weighted local polynomial regression estimator has received increasing attention and it has gained acceptance as an attractive method of nonparametric estimation of regression function and its derivatives. This

smoothing method has become a powerful and useful diagnostic tool for data analysis. The estimator is obtained by locally fitting a  $p$ th degree polynomial to the data via weighted least squares and it presents advantages compared with other kernel nonparametric regression estimators. These are its better boundary behavior, its adaptation to estimate regression derivatives and its good minimax properties. Some significant references are, for example, Lejeune (1985), Müller (1988), Tsybakov (1986,1990), Fan (1992, 1993), Hastie and Loader (1993), Fan and Gijbels (1992, 1995), Ruppert and Wand (1994) and Fan et al. (1997). In these papers the independence of the observations is assumed. The statistical properties of local polynomial regression for dependent data have been studied in recent works of Masry and Fan (1997), Masry (1996a,b), Härdle and Tsybakov (1997), Härdle et al. (1998), Opsomer (1997) and Vilar-Fernández and Vilar-Fernández (1998, 2000). In these works the regression model with random design was considered and the assumption of the data satisfying some mixing condition was used. A wide study of this smoothing method can also be found in the recent monograph of Fan and Gijbels (1996).

In this paper, the fixed design and the short-range dependence nonparametric regression model is considered. In what follows it is assumed that univariate data  $Y_{1,n}, Y_{2,n}, \dots, Y_{n,n}$  are observed, and that

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

where  $x_{t,n}$ ,  $1 \leq t \leq n$ , are the design points,  $m(x)$  is a “smooth” regression function defined on  $[0, 1]$ , without any loss of generality, and  $\varepsilon_{t,n}$ ,  $1 \leq t \leq n$ , is a sequence of unobserved random variables with zero mean and finite variance  $\sigma^2$ . We assume, for each  $n$ ,  $\{\varepsilon_{1,n}, \varepsilon_{2,n}, \dots, \varepsilon_{n,n}\}$  have the same joint distribution as  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ , where  $\{\varepsilon_t, t \in Z\}$  is a strictly stationary stochastic process. Also, it is assumed that design  $x_{t,n}$ ,  $1 \leq t \leq n$ , is a regular design generated by a design density  $f$ ; that is, for each  $n$ , the design points are defined by

$$\int_0^{x_{t,n}} f(x)d(x) = \frac{t-1}{n-1}, \quad 1 \leq t \leq n, \quad (2)$$

$f$  being a positive function, defined on  $[0, 1]$  and its first derivative is continuous.

These regression models frequently arise in economic studies, in the analysis of growth curves and usually in situations in which data are collected sequentially over time, for example, in the study of time series with deterministic tendency. Now, the purpose of this paper is to study the properties of the local polynomial kernel estimator of regression function  $m$  and its derivatives.

The organization of the work is as follows: In Section 2, an estimator for a regression function and its derivatives is introduced and its asymptotic properties are studied. Expressions for the bias and variance/covariance matrix are obtained and asymptotic normality is shown. In Section 3, a method of bandwidth selection is proposed and its empirical performance is tested via a small simulation study presented in Section 4. In this section, we also include the analysis of an example of real economic data. Finally, the last section is devoted to the proofs of the obtained results.

## 2. THE ESTIMATOR AND ITS ASYMPTOTIC PROPERTIES

We consider the fixed design regression model given in (1). Our goal is to estimate the unknown regression function  $m(x) = E(Y/x)$  and its derivatives based on an observed sample  $\{(x_t, Y_t)\}_{t=1}^n$ . The nonparametric estimator used is based on a weighted local polynomial fitting.

### 2.1. The estimator.

If we assume that the  $(p+1)$ th derivatives of the regression function at point  $x$  exist and are continuous, local polynomial fitting permits estimating the parameter vector  $\vec{\beta}(x) = (\beta_0(x), \beta_1(x), \dots, \beta_p(x))^t$ , where  $\beta_j(x) = m^{(j)}(x)/(j!)$ , with  $j = 0, 1, \dots, p$ , by minimizing the function

$$\Psi(\vec{\beta}(x)) = \sum_{t=1}^n \left( Y_t - \sum_{j=0}^p \beta_j(x)(x_t - x)^j \right)^2 \omega_{n,t}, \quad (3)$$

where  $\omega_{n,t} = n^{-1}K_n(x_t - x)$  are the weights,  $K_n(u) = h_n^{-1}K(h_n^{-1}u)$ ,  $K$  being a kernel function and  $h_n$  the bandwidth or smoothing parameter that controls the

size of the local neighborhood and so the degree of smoothing. The estimator of  $\vec{\beta}(x)$ , obtained as a solution to the weighted least squares problem given in (3), is called the local polynomial kernel estimator and it is interesting to observe that this class of estimators includes the classical Nadaraya-Watson estimator, which is the minimizer of (3) when  $p = 0$ . Of special interest is also the local linear kernel estimator corresponding to  $p = 1$ .

The minimizing problem is introduced in matrix notation for concise presentation of results. Let us denote

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

and let  $W_{(n)} = \text{diag}(\omega_{n,1}, \dots, \omega_{n,n})$  be the diagonal array of weights. Then by assuming the invertibility  $X_{(n)}^t W_{(n)} X_{(n)}$ , standard weighted least squares theory leads to the solution

$$\hat{\beta}_{(n)}(x) = (X_{(n)}^t W_{(n)} X_{(n)})^{-1} X_{(n)}^t W_{(n)} \vec{Y}_{(n)} = S_{(n)}^{-1} \vec{T}_{(n)}, \quad (4)$$

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

$$s_j^{(n)} = \frac{1}{n} \sum_{t=1}^n (x_t - x)^j K_n(x_t - x), \quad 0 \leq j \leq 2p, \quad (5)$$

and  $\vec{T}_{(n)} = (t_0^{(n)}, t_1^{(n)}, \dots, t_p^{(n)})^t$ , being

$$t_i^{(n)} = \frac{1}{n} \sum_{t=1}^n (x_t - x)^i K_n(x_t - x) Y_t, \quad 0 \leq i \leq p. \quad (6)$$

## 2.2. The Mean Squared Error.

In this subsection asymptotic expressions for the bias and variance/covariance array of the estimate defined in (4) are obtained. The following assumptions will be needed in our analysis:

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

**A.2.** The sequence of bandwidths or smoothing parameters,  $\{h_n\}$ , satisfies that  $h_n > 0$ ,  $h_n \downarrow 0$ ,  $nh_n \uparrow \infty$ .

**A.3.** Denote  $Cov(\epsilon_i, \epsilon_{i+k}) = \sigma^2 c(k)$ ,  $k = 0, \pm 1, \dots$  then  $\sum_{k=1}^{\infty} k |c(k)| < \infty$ .

From assumption A3 we assume that the strength of correlation between error terms is independent of sample size. This assumption, among others, is satisfied by time series that are of the form tendency plus random component, used in Hart (1991), Chu and Marron (1991), and Tran et al. (1996), among others.

The following notations will be used. Let  $\mu_j = \int u^j K(u) du$  and  $\nu_j = \int u^j K^2(u) du$ ,  $j = 0, 1, 2, \dots$  and let us denote  $\vec{\mu} = (\mu_{p+1}, \dots, \mu_{2p+1})^t$ .

To obtain results on the asymptotic behavior of the proposed estimates, we need to establish the convergence for each element array  $S_{(n)}$  and the mean squared convergence of  $\vec{T}_{(n)}^*$ , that is, the variable vector  $\vec{T}_{(n)}$  centered with respect to vector  $\vec{M}_{(n)} = (m(x_1), \dots, m(x_n))^t$ .

**PROPOSITION 1.** Under assumptions A1 and A2, for every  $x \in (h_n, 1 - h_n)$ , we have

$$\lim_{n \rightarrow \infty} h_n^{-j} s_j^{(n)} = f(x) \mu_j, \quad 0 \leq j \leq 2p + 1. \quad (7)$$

This result can be expressed in matrix form as

$$\lim_{n \rightarrow \infty} H_{(n)}^{-1} S_{(n)} H_{(n)}^{-1} = f_X(x) S, \quad (8)$$

where  $H_{(n)} = \text{diag}(1, h_n, h_n^2, \dots, h_n^p)$  and  $S$  is the  $(p+1) \times (p+1)$  array whose  $(i, j)$ th element is  $s_{i,j} = \mu_{i+j-2}$ .

In order to establish the mean squared convergence of  $\hat{\beta}_{(n)}(x)$ , it is necessary to study the asymptotic performance of vector  $\vec{T}_{(n)}^* = (t_{0,(n)}^*, \dots, t_{p,(n)}^*)^t$ , whose  $i$ th component is

$$t_{i,(n)}^* = \frac{1}{n} \sum_{t=1}^n (x_t - x)^i K_n(x_t - x) (Y_t - m(x_t)), \quad 0 \leq i \leq p. \quad (9)$$

**PROPOSITION 2.** Under assumptions A1, A2 and A3, for every  $x \in (h_n, 1 - h_n)$ , we have

$$\lim_{n \rightarrow \infty} nh_n \text{Cov} (h_n^{-j} t_{j,(n)}^*, h_n^{-i} t_{i,(n)}^*) = \nu_{j+i} f(x) c(\varepsilon) \quad \text{for } 0 \leq j, i \leq p, \quad (10)$$

or, in matrix form

$$\lim_{n \rightarrow \infty} nh_n E \left( H_{(n)}^{-1} \vec{T}_{(n)}^* \vec{T}_{(n)}^{*t} H_{(n)}^{-1} \right) = \tilde{S} f(x) c(\varepsilon), \quad (11)$$

where  $\tilde{S}$  is the array whose  $(i, j)$ th element is  $\tilde{s}_{i,j} = \nu_{i+j-2}$  and  $c(\varepsilon) = \sigma^2 (c(0) + 2 \sum_{k=1}^{\infty} c(k))$ .

The proofs of these two previous results may be found in Section 5. Now, using these results we can establish the mean squared convergence of  $\hat{\beta}_{(n)}(x)$ . For it, let

$$\vec{\beta}_{(n)}^*(x) = E \left( \hat{\beta}_{(n)}(x) \right) = S_{(n)}^{-1} X_{(n)}^t W_{(n)} \vec{M}_{(n)}. \quad (12)$$

Performing a  $(p+1)$ th-order Taylor series expansion in a neighborhood of  $x$ , we obtain

$$\vec{M}_{(n)} = X_{(n)} \vec{\beta}(x) + \frac{m^{(p+1)}(x)}{(p+1)!} \begin{pmatrix} (x_1 - x)^{p+1} \\ \vdots \\ (x_n - x)^{p+1} \end{pmatrix} + o \begin{pmatrix} (x_1 - x)^{p+1} \\ \vdots \\ (x_n - x)^{p+1} \end{pmatrix}. \quad (13)$$

Substituting this expression in (12) and using A1,

$$\vec{\beta}_{(n)}^*(x) = \vec{\beta}(x) + S_{(n)}^{-1} \left( \frac{m^{(p+1)}(x)}{(p+1)!} \begin{pmatrix} s_{p+1}^{(n)} \\ \vdots \\ s_{2p+1}^{(n)} \end{pmatrix} + o \begin{pmatrix} h_n^{p+1} \\ \vdots \\ h_n^{2p+1} \end{pmatrix} \right). \quad (14)$$

On the other hand, definitions (9) and (12) yield

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

From both equations, we obtain

$$\hat{\beta}_{(n)}(x) - \vec{\beta}_{(n)}(x) = S_{(n)}^{-1} \vec{T}_{(n)}^* + S_{(n)}^{-1} \left( \frac{m^{(p+1)}(x)}{(p+1)!} \begin{pmatrix} s_{p+1}^{(n)} \\ \vdots \\ s_{2p+1}^{(n)} \end{pmatrix} + o \begin{pmatrix} h_n^{p+1} \\ \vdots \\ h_n^{2p+1} \end{pmatrix} \right). \quad (15)$$

This equation is very important; the first term on the right side of (15) corresponds to the variance of the estimator  $\hat{\beta}_{(n)}(x)$  and the second term corresponds to the bias.

From equation (15) and Propositions 1 and 2, we can deduce the following theorem:

**THEOREM 1.** Under assumptions A1, A2 and A3, for every  $x \in (h_n, 1 - h_n)$ , we have the following asymptotic expression of bias of the estimator  $\hat{\beta}_{(n)}(x)$ ,

$$H_{(n)}E \left( \hat{\beta}_{(n)}(x) - \vec{\beta}(x) \right) = \frac{m^{(p+1)}(x)}{(p+1)!} h_n^{p+1} S^{-1} \vec{\mu} + o \left( h_n^{p+1} (1, \dots, 1)^t \right), \quad (16)$$

and the following asymptotic expression of variance of  $\hat{\beta}_{(n)}(x)$ ,

$$Var \left( H_{(n)} \hat{\beta}_{(n)}(x) \right) = \frac{1}{nh_n} \frac{c(\varepsilon)}{f(x)} S^{-1} \tilde{S} S^{-1} + o \left( \frac{1}{nh_n} \right). \quad (17)$$

Asymptotic expressions for the bias and variance of the regression function estimator and its derivatives are directly derived from Theorem 1.

**COROLLARY 1.** If the assumptions of Theorem 1 are satisfied, then for every  $x \in (h_n, 1 - h_n)$ , we have

$$Bias \left( \hat{m}^{(j)}(x) \right) = h_n^{p+1-j} \frac{m^{(p+1)}(x)}{(p+1)!} j! B_j (1 + o(1)), \quad (18)$$

$$Var \left( \hat{m}^{(j)}(x) \right) = \frac{1}{nh_n^{2j+1}} \frac{c(\varepsilon)}{f(x)} (j!)^2 V_j (1 + o(1)), \quad (19)$$

with  $j = 0, 1, \dots, p$ . The terms  $B_j$  and  $V_j$  denote the  $j$ th element of  $S^{-1} \vec{\mu}$  and the  $j$ th diagonal element of  $S^{-1} \tilde{S} S^{-1}$ , respectively.

The first implication of Corollary 1 is that the dependence of the observations influences the variance of the estimator but not the bias, as expected. However, the rate of convergence in Mean Squared Error of the proposed estimator is the same under independence of the observations as under dependence, although slower in this case due to the larger variance. The results

agree with those obtained by Hart (1991) in his Mean Squared Error study of Gasser-Müller kernel regression function estimator when model (1) is considered.

### 2.3. Asymptotic normality

To establish the asymptotic normality of  $\hat{\beta}_{(n)}(x)$ , it is first appropriate to study the asymptotic performance of vector  $\vec{T}_{(n)}^*$ . For it, the two following additional assumptions are necessary:

**A.4.**  $E|\varepsilon_t|^{2+\delta} < \infty$  for some  $\delta > 0$

**A.5.** The stationary stochastic process  $\{\varepsilon_t\}$  is  $\alpha$ -mixing with mixing coefficients such that  $\sum_{t=1}^{\infty} \alpha(t)^{\delta/(2+\delta)} < \infty$ . And, there exists a sequence of positive integers  $\{s_n\}$ ,  $s_n \rightarrow \infty$  as  $n \rightarrow \infty$  with  $s_n = o\left((nh_n^3)^{1/2}\right)$  and such that  $(nh_n^{-1})^{1/2} \sum_{t=s_n}^{\infty} \alpha(t)^{1-\gamma} < \infty$ , with  $\gamma = 2/(2 + \delta)$ .

**A.6.**  $h_n = O\left(n^{-1/(2p+3)}\right)$ .

Note that the strong mixing condition ( $\alpha$ -mixing), introduced by Rosenblatt (1956), is one of the least restrictive among the numerous dependence conditions and it is satisfied by many processes, for example, the ARMA processes generated by absolutely continuous noise. A thorough study of this condition can be seen in Doukhan (1995).

**PROPOSITION 3.** If assumptions A1-A6 are fulfilled then for every  $x \in (h_n, 1 - h_n)$ , we have

$$\sqrt{nh_n} H_{(n)}^{-1} \vec{T}_{(n)}^* \xrightarrow{\mathcal{L}} N_{(p+1)}\left(\vec{0}, f(x)c(\varepsilon)\tilde{S}\right), \quad (20)$$

where  $N_{(p+1)}(\vec{\mu}, \Sigma)$  denotes a multivariate normal distribution of dimension  $p + 1$ , with mean vector  $\vec{\mu}$  and variance-covariance matrix  $\Sigma$ .

Now, the asymptotic normality of  $\hat{\beta}_{(n)}(x)$  follows from (15) and the convergence results established in Propositions 1 and 3.



**THEOREM 2.** Under assumptions of Proposition 3, we have

$$\sqrt{nh_n} \left[ H_{(n)} \left( \hat{\beta}_{(n)}(x) - \vec{\beta}(x) \right) - \frac{m^{(p+1)}(x)}{(p+1)!} h_n^{p+1} S^{-1} \vec{\mu} \right] \xrightarrow{\mathcal{L}} N_{(p+1)}(0, \Sigma), \quad (21)$$

where  $\Sigma = \frac{c(\varepsilon)}{f(x)} S^{-1} \tilde{S} S^{-1}$ .

The asymptotic normality of the individual components  $\hat{\beta}_j^{(n)}(x) = \frac{\hat{m}^{(j)}(x)}{j!}$  is directly derived from Theorem 2.

**COROLLARY 2.** Under the hypotheses of Theorem 2, we have, for  $j = 0, \dots, p$ ,

$$\sqrt{nh_n^{1+2j}} \left[ \left( \hat{m}^{(j)}(x) - m^{(j)}(x) \right) - h_n^{p+1-j} \frac{m^{(p+1)}(x)}{(p+1)!} j! B_j \right] \xrightarrow{\mathcal{L}} N(0, \sigma_j^2), \quad (22)$$

where  $\sigma_j^2 = \frac{c(\varepsilon)}{f(x)} (j!)^2 V_j$  and the terms  $B_j$  and  $V_j$  are given in Corollary 1.

Masry and Fan (1997) studied the local polynomial kernel estimator in regression models with random design and  $\rho$ -mixing and  $\alpha$ -mixing observations. In their Theorem 5, they obtained a similar result to (22), but in their case the asymptotic variance expression under dependence coincides with the result for independent observations.

### 3. BANDWIDTH SELECTION METHODS

In any problem of nonparametric estimation of curves, the choice of the bandwidth parameter is very important and hence this should be done with extreme care. Several methods of selecting bandwidths have been proposed in kernel estimation of the regression function with dependent errors when the Nadaraya-Watson estimator or the Gasser-Müller estimator were employed. Most of the proposed procedures of bandwidth selection use a *cross-validation*

algorithm or a *plug-in* method. Among those using the former are the modified cross-validation (MCV), that is, simply the “leave( $2l + 1$ )-out” version of cross-validation (see Härdle and Vieu (1992)), the partitioned cross-validation (PCV) (Chu and Marron (1991)) and the time series cross-validation (TSCV) proposed by Hart (1994). Other procedures follow a plug-in method. This method consists of obtaining an estimator of the bandwidth that minimizes the Mean Squared Error (MSE) or Mean Integrated Squared Error (MISE), replacing the unknown quantities by estimators. Along this line are the bandwidth selectors proposed by Herrmann, Gasser and Kneip (1992) and Quintela (1994).

In this section, we propose practical bandwidth selection procedures for local polynomial estimator, given in (4), of the regression function and its derivatives. The proposed procedures are of plug-in type and are based on minimizing the Mean Squared Error, asymptotic or theoretical.

Using the asymptotic expressions of bias and variance, (18) and (19), we can obtain an asymptotically optimal local bandwidth,  $h_{j,l,as}^{opt}(x)$ , for estimator  $\hat{m}^{(j)}(x)$ , minimizing the asymptotic Mean Squared Error,

$$MSE(\hat{m}^{(j)}(x)) = \left( \left( h_n^{p+1-j} \frac{m^{(p+1)}(x)}{(p+1)!} j! B_j \right)^2 + \frac{1}{nh_n^{2j+1}} \frac{c(\varepsilon)}{f(x)} (j!)^2 V_j \right) (1 + o(1)), \quad (23)$$

therefore,

$$h_{j,l,as}^{opt}(x) = C_{j,p}(K) \left( \frac{c(\varepsilon)}{n(m^{(p+1)}(x))^2 f(x)} \right)^{1/(2p+3)}, \quad (24)$$

where  $C_{j,p}(K)$  is a real number that depends on kernel  $K$  (Table 3.2 of Fan and Gijbels (1996) lists some values of  $C_{j,p}(K)$ ). Equation (24) shows the influence of dependence. So, when we wrongly assume independence of the errors, since  $\sigma^2 \neq c(\varepsilon)$ , the asymptotically optimal bandwidth considered will be different from (24). For example, if  $\varepsilon_t$  follows an AR(1) model with  $\rho = 0.8$ , we have  $c(\varepsilon) = 9\sigma^2$  and therefore,  $h(x) \simeq h_{j,l,as}^{opt}(x) / 1.55$ , when  $p = 1$ .

In the obtained expression for the local bandwidth, there are two unknown quantities:  $c(\varepsilon)$ , given by the dependence structure of the errors, and  $m^{(p+1)}(x)$ , the  $(p + 1)$ th derivative of the regression function. In practice, these un-

known quantities are replaced by estimators, and the computed bandwidth is  $\hat{h}_{j,l,as}^{opt}(x)$ . To estimate the unknown quantities before cited, various possibilities arise. In section 4.2 of Fan and Gijbels (1996),  $m^{(p+1)}(x)$  is estimated fitting to the regression function, globally, a polynomial of  $(p+3)$  order,  $\check{m}_{p+3}(x)$ , and next calculating the  $(p+1)$  derivative of this fitted curve,  $\check{m}_{p+3}^{(p+1)}(x)$ . Moreover, we can compute the residuals  $\check{\varepsilon}_t = Y_t - \check{m}_{p+3}(x_t)$ ,  $1 \leq t \leq n$ . On the other hand, this can also be achieved using a local regression polynomial, for which a pilot bandwidth,  $h_{pilot}$ , is necessary. Therefore, with this pilot bandwidth,  $m^{(p+1)}(x)$  is estimated (as explained in Section 2.1) and the nonparametric residuals are calculated as  $\tilde{\varepsilon}_t = Y_t - \tilde{m}_{h_{pilot}}(x_t)$ ,  $1 \leq t \leq n$ . With respect to the other parameter,  $c(\varepsilon) = \sigma^2 (c(0) + 2 \sum_{k=1}^{\infty} c(k))$ , let us consider two situations. If we assume that the error follows an ARMA model or another parametric model, we can obtain the covariances from a small number of parameters. Then, an estimator of  $c(\varepsilon)$  can be obtained by simply estimating these parameters from nonparametric residuals. If, on the other hand, we do not assume a parametric structure of covariances, then an estimator of  $c(\varepsilon)$  is directly obtained from the data. In this way, for  $m$ -dependent residuals, Müller and Stadtmüller (1988) proposed an estimator for  $c(\varepsilon)$  based on first order differences of  $Y_t$ . For residuals that are not necessarily  $m$ -dependent but satisfy some mixing conditions, Herrmann, Gasser and Kneip (1992) suggested an estimator of  $c(\varepsilon)$  based on second order differences of  $Y_t$ .

To obtain a global bandwidth, we use as measure of global error of the estimation, the Mean Integrated Squared Error, given by

$$MISE(h) = \int MSE(\hat{m}^{(j)}(x)) \omega(x) dx, \quad (25)$$

where  $\omega(x) \geq 0$  is a weight function. An asymptotically optimal global bandwidth,  $h_{j,g,as}^{opt}$ , can be obtained by minimizing the asymptotic  $MISE(h)$ , in which case the bandwidth is

$$h_{j,g,as}^{opt} = C_{j,p}(K) \left( \frac{c(\varepsilon)}{n \int (m^{(p+1)}(x))^2 \omega(x) f(x) dx} \right)^{1/(2p+3)}. \quad (26)$$

Again, the global bandwidth depends on two unknown quantities,  $c(\varepsilon)$  and  $\int (m^{(p+1)}(x))^2$ , that we can replace by estimators, obtaining the bandwidth

$\hat{h}_{j,g,as}^{opt}(x)$ . Estimators of these two values can be calculated as we indicate previously.

On the other hand and following the analogous ideas of Fan et al. (1996), the bias and variance of the estimator  $\hat{\beta}_{(n)}(x)$  are derived immediately from its definition in (4).

$$Bias\left(\hat{\beta}_{(n)}(x)\right) = \vec{\beta}_{(n)}^*(x) - \vec{\beta}(x) = S_{(n)}^{-1} X_{(n)}^t W_{(n)} \vec{R}_{(n)}. \quad (27)$$

$$Var\left(\hat{\beta}_{(n)}(x)\right) = S_{(n)}^{-1} X_{(n)}^t W_{(n)} \Gamma_{(n)} W_{(n)} X_{(n)} S_{(n)}^{-1}, \quad (28)$$

where  $\vec{R}_{(n)} = \vec{M}_{(n)} - X_{(n)} \vec{\beta}(x)$  and  $\Gamma_{(n)} = E(\vec{\varepsilon} \vec{\varepsilon}^t)$  is the variance-covariance matrix of the residuals. These two arrays are unknown. The exact Mean Squared Error of  $\hat{\beta}_{(n),j}(x)$  is

$$MSE\left(\hat{\beta}_{(n),j}(x)\right) = b_{(n),j}^2(x) + V_{(n),j}(x), \quad j = 0, 1, \dots, p, \quad (29)$$

where  $b_{(n),j}(x)$  and  $V_{(n),j}(x)$  are the  $(j+1)$ th-element of the bias vector (27) and the  $(j+1)$ th-diagonal element of the variance matrix (28), respectively. Again, the  $MSE\left(\hat{\beta}_{(n),j}(x)\right)$  cannot be computed, but it can be estimated if estimators of the  $Bias\left(\hat{\beta}_{(n)}(x)\right)$  and  $\Gamma_{(n)}$  are obtained. Then, we define the estimated optimal local bandwidth as

$$\hat{h}_{j,l}^{opt} = \arg \min_h M\hat{S}E\left(\hat{\beta}_{(n),j}(x)\right). \quad (30)$$

Using Taylor's expansion of order  $p+a$ , an approximation of  $Bias\left(\hat{\beta}_{(n)}(x)\right)$  can be obtained; Fan and Gijbels (1996) recommend to choose  $a = 2$  because, without a high computational cost, a bandwidth selector with good asymptotic properties is obtained. For simplicity, we use  $a = 1$ , then  $Bias\left(\hat{\beta}_{(n)}(x)\right)$  can be approximated by

$$Bias\left(\hat{\beta}_{(n)}(x)\right) \approx S_{(n)}^{-1} \beta_{p+1} \vec{\mu}. \quad (31)$$

On the right side of (31), quantity  $\beta_{p+1}$  is unknown but can be estimated by using a local polynomial regression of order  $r$  ( $r > p$ ) with a pilot bandwidth  $h^*$ .

This fitting is also used to obtain the estimated residuals  $\varepsilon_i^* = Y_i - Y_i^*$ , where  $Y_i^* = \hat{m}_{h^*}(x_i)$  is the predicted value in  $x_i$  after the local  $r$ th-order polynomial fit. From these residuals an estimator,  $\Gamma^*$ , of the variance-covariance matrix is obtained.

Finally, the estimated optimal global bandwidth is defined as

$$\hat{h}_{j,g}^{opt} = \arg \min_h \int M\hat{S}E \left( \hat{\beta}_{(n),j}(x) \right) dx. \quad (32)$$

In both bandwidth selection procedures, the asymptotic plug-in and direct plug-in, the influence of the dependence of the residuals is observed. So, if high positive correlations are present and they are not considered, the bands selected will be too small.

We now illustrate the behavior of the estimator defined in (4) and some of the bandwidth selectors proposed in this section. We perform a small simulation study and analyze a numerical real data sample. On the one hand, the objective of these studies is to compare the bandwidths obtained by the asymptotic plug-in method and direct method and on the other hand, to observe the effect of obviating the dependence of the data. In our study, global bands given in (26) and (32), of the local linear estimator of the regression function, are considered.

## 4. SIMULATION STUDY AND EXAMPLE

In this section we present some simulation results of the proposed estimator and of the methods of selecting bandwidths presented in Section 3. We also apply this estimator to a set of real-data.

### 4.1. Simulation study.

A first study was carried out in order to compare bandwidths  $\hat{h}_{j,g,as}^{opt}$  and  $\hat{h}_{j,g}^{opt}$  (given in (26) and (32), respectively, using the local linear estimator of the regression function, that is,  $p = 1$  and  $j = 0$ ) computed under independence

and under dependence and also to study the influence of the dependence of data in these selectors. For this purpose, in a first step, we simulated 300 samples of size  $n = 100$ , following the regression model given in (1), where we considered a design of equally spaced points on the unit interval,  $x_t = t/n$ ,  $t = 1, \dots, n$ , with regression function  $m(x) = 5(x - 0.5)^3$  and errors,  $\varepsilon_t$ , following a dependence structure of  $AR(1)$  type,

$$\varepsilon_t = \rho\varepsilon_{t-1} + e_t,$$

where  $\varepsilon_t$  have distribution function  $N(0, \sigma^2)$ , with  $\sigma = 0.3$ . In order to study the influence of dependence of the observations, we considered the following values of correlation coefficient,  $\rho = 0.9, 0.6, 0.3, 0, -0.3, -0.6, -0.9$ . Now, we computed the measurements of discrepancy in the norms  $L_2$  (the mean integrated square error,  $MISE$ ),  $L_1$  (the mean integrated absolute error,  $MIAE$ ) and  $L_\infty$  (the mean uniform absolute error,  $MUAE$ ), between the underlying regression function  $m(x)$  and the estimator  $\hat{m}_h(x)$ ,

$$MISE(h) = E \int (\hat{m}_h(t) - m(t))^2 dt, \quad (33)$$

$$MIAE(h) = E \int |\hat{m}_h(t) - m(t)| dt, \quad (34)$$

$$MUAE(h) = E \left( \sup_t |\hat{m}_h(t) - m(t)| \right). \quad (35)$$

This is done for every bandwidth in a grid of equally spaced values of  $h$ . The integrals appearing in the three norms are approximated by means of Riemman sums. By minimizing functions (33), (34) and (35) numerically in  $h$ , some numerical approximations to the values  $h_{MISE}$ ,  $h_{MIAE}$  and  $h_{MUAE}$  are found.

The second step consists in drawing another 300 random samples of sample size  $n = 100$  and computing the four plug-in bandwidths for every sample. Using Montecarlo approximations, once more, the expected value, the standard deviation for every selector and the mean squared of the error committed,  $M(\hat{h})$ , with respect to its minimum,  $\min_{h>0} M(h)$ , given by

$$\Delta M = E \left( M(\hat{h}) - \min_{h>0} M(h) \right)^2$$

where  $M = MISE$ ,  $MIAE$  or  $MUAE$  can be approximated.

In the computation of the global bandwidths  $\hat{h}_{j,g,as}^{opt}$  and  $\hat{h}_{j,g}^{opt}$ , we need to estimate parameter  $c(\varepsilon) = \sigma^2$ , if we suppose that the errors are independent, and  $c(\varepsilon) = \left(\frac{1+\rho}{1-\rho}\right)\sigma^2$ , under the hypothesis that the errors follow an  $AR(1)$  dependent model. For this purpose, we use the following consistent estimators

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 \quad \text{and} \quad \hat{\rho} = \frac{\sum_{i=2}^n \hat{\varepsilon}_i \hat{\varepsilon}_{i-1}}{\sum_{i=1}^n \hat{\varepsilon}_i^2},$$

where

$$\hat{\varepsilon}_i = Y_i - \hat{m}_{h_{TSCV}}(x_i), \quad i = 1, 2, \dots, n.$$

These nonparametric residuals were obtained using a pilot bandwidth,  $h_{TSCV}$ , computed by Time Series Cross-Validation method, proposed by Hart (1994). On the other hand, an empirically chosen pilot bandwidth,  $h_{pilot} = 0.95$ , was used to carry out a local polynomial fitting of third order from which  $m''(x)$  (needed to calculate  $\hat{h}_{0,g,as}^{opt}$ ), and  $\beta_2$  (used to compute  $\hat{h}_{0,g}^{opt}$ ) are estimated.

Table I shows the approximated optimal bandwidths with respect to each of the three criteria and the Montecarlo approximation of the mean of every plug-in selector,  $\hat{h}_{0,g,as}^{opt}$  and  $\hat{h}_{0,g}^{opt}$ , under independence and dependence.

TABLE I							
bandwidth $\times 10^{-1}$	-0.9	-0.6	-0.3	$\rho = 0$	0.3	0.6	0.9
$h_{MISE}$	1.666	2.090	2.454	2.878	3.424	4.333	5.787
$h_{MIAE}$	1.363	1.969	2.272	2.636	3.121	4.030	6.333
$h_{MUAE}$	2.151	2.575	2.878	3.424	3.848	4.463	4.333
$\hat{h}_{0,g,as}^{opt}$ (under dep.)	1.477	1.949	2.268	2.571	2.900	3.253	3.254
$\hat{h}_{0,g}^{opt}$ (under dep.)	1.487	1.915	2.264	2.619	2.985	3.292	3.107
$\hat{h}_{0,g,as}^{opt}$ (under indep.)	2.596	2.598	2.596	2.615	2.633	2.600	2.220
$\hat{h}_{0,g}^{opt}$ (under indep.)	2.654	2.659	2.654	2.673	2.668	2.578	2.127

Table 1: Optimal bandwidths, and  $\hat{h}_{0,g,as}^{opt}$  and  $\hat{h}_{0,g}^{opt}$  under independence and dependence.

Table II includes the efficiency measures,  $\Delta MISE$ , as a function of  $\rho$  and Table III the results of the simulation study with  $\rho = 0.9$ .

<b>TABLE II</b>				
$\Delta MISE$	$\rho = -0.9$	$\rho = -0.6$	$\rho = -0.3$	$\rho = 0$
$\hat{h}_{0,g,as}^{opt}$ (dependence)	$1.00E - 9$	$1.00E - 9$	$9.00E - 9$	$5.60E - 8$
$\hat{h}_{0,g}^{opt}$ (dependence)	$1.00E - 9$	$3.00E - 9$	$1.70E - 8$	$8.80E - 8$
$\hat{h}_{0,g,as}^{opt}$ (independence)	$1.49E - 7$	$3.80E - 8$	$1.12E - 8$	$4.80E - 8$
$\hat{h}_{0,g}^{opt}$ (independence)	$1.94E - 7$	$8.00E - 8$	$3.60E - 8$	$8.00E - 8$
$\Delta MISE$	$\rho = 0.9$	$\rho = 0.6$	$\rho = 0.3$	
$\hat{h}_{0,g,as}^{opt}$ (dependence)	$1.70E - 5$	$2.00E - 6$	$3.11E - 7$	
$\hat{h}_{0,g}^{opt}$ (dependence)	$2.50E - 5$	$3.00E - 6$	$3.78E - 7$	
$\hat{h}_{0,g,as}^{opt}$ (independence)	$8.00E - 5$	$1.00E - 5$	$7.35E - 7$	
$\hat{h}_{0,g}^{opt}$ (independence)	$9.80E - 5$	$1.20E - 5$	$9.13E - 7$	

Table 2:  $\Delta MISE$  as a function of  $\rho$ .

<b>TABLE III</b>				
$\rho = 0.9$	Bandwidth	$MISE$	$MIAE$	$MUAE$
$h_{MISE}$	$5.78E - 1$	$4.04E - 2$	$1.57E - 1$	$3.92E - 1$
$h_{MIAE}$	$6.33E - 1$	$4.05E - 2$	$1.57E - 1$	$3.96E - 1$
$h_{MUAE}$	$4.33E - 1$	$4.10E - 2$	$1.59E - 1$	$3.87E - 1$
$\rho = 0.9$	Mean (St. dev.)	$\Delta MISE$	$\Delta MIAE$	$\Delta MUAE$
$\hat{h}_{0,g,as}^{opt}$ (dep.)	0.325 (0.0897)	$1.70E - 5$	$9.20E - 5$	$3.01E - 4$
$\hat{h}_{0,g}^{opt}$ (dep.)	0.310 (0.0926)	$2.50E - 5$	$1.29E - 4$	$4.75E - 4$
$\hat{h}_{0,g,as}^{opt}$ (indep.)	0.222 (0.0611)	$8.00E - 5$	$3.72E - 4$	$1.70E - 3$
$\hat{h}_{0,g}^{opt}$ (indep.)	0.212 (0.0645)	$9.80E - 5$	$4.41E - 4$	$2.13E - 3$

Table 3: Results of the simulation study with  $\rho = 0.9$ .



For a better interpretation of these results, we also include two figures. In Figure 1, the studied bandwidths as a function of  $\rho$  are represented. In Figure 2, the values of  $\log(\Delta MISE)$  for every selector, as a function of  $\rho$ , are represented.

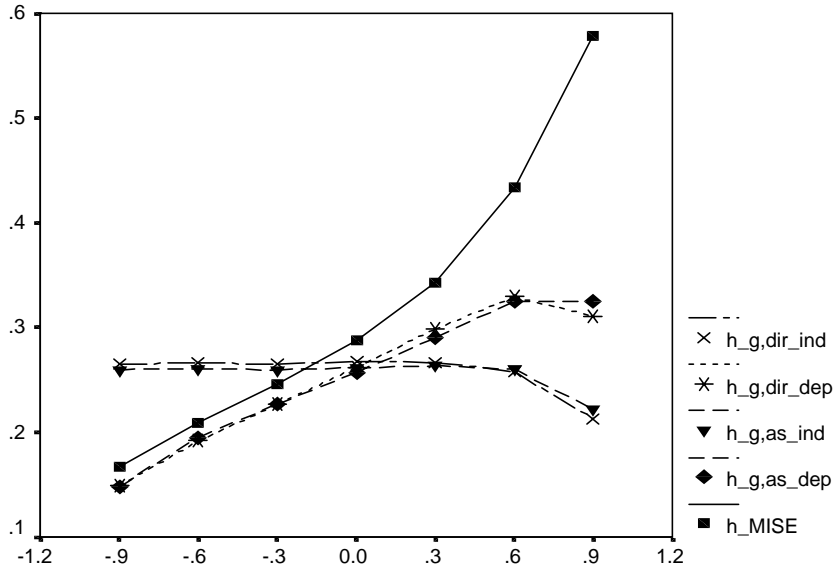


Figure 1: Bandwidths as a function of  $\rho$

In the simulation results, we can observe that the amount of positive dependence makes that the optimal bandwidths,  $h_{MISE}$ ,  $h_{MIAE}$  and  $h_{MUAE}$ , increase remarkably. The plug-in selectors,  $\hat{h}_{0,g,as}^{opt}$  and  $\hat{h}_{0,g}^{opt}$ , under dependence have the same behavior and so they take into account the influence of dependence. For  $\rho = 0.9$ , the results are worse. This may be due to a need for, owing to the strong dependence, using larger sample sizes to obtain good results. On the other hand, the plug-in bandwidths, computed under the hypothesis of independence, show a worse behavior, in the sense that they are insensitive to the dependence of the observations (Figure 1) and do not follow the optimal bandwidths. In the simulation study here presented and in others we

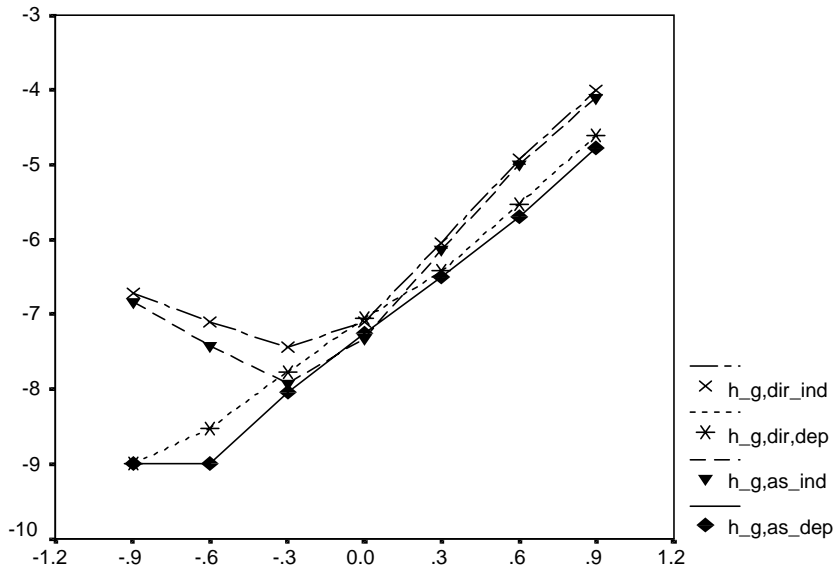


Figure 2:  $\log(\Delta MISE)$ , for every selector, as a function of  $\rho$ .

have performed, we can observe that bandwidths  $\hat{h}_{j,g,as}^{opt}$  and  $\hat{h}_{j,g}^{opt}$ , under dependence, present a similar and good performance. So, it is important to take into account the dependence of the observations when a bandwidth is chosen and to use this information in the computation of the bandwidth, as with the proposed selectors,  $\hat{h}_{j,g,as}^{opt}$  and  $\hat{h}_{j,g}^{opt}$ .

#### 4.2. A numerical example.

We now illustrate the behavior of weighted local polynomial regression with the data of accumulative (last twelve months) retail price index of Spain. The studied series is of 224 months, from January 1979 to August 1997. Each observation indicates the accumulated R.P.I. in the last twelve months and they were obtained from the database TEMPUS of Instituto Nacional de Estadística de España (National Institute of Statistics of Spain).

The regression model can be fitted to these data with  $x_t = t/n$  and smoothing parameters are obtained using the same methods as in the above simulation study.

TABLE IV	
$\hat{h}_{0,g,as}^{opt}$ (under dependence)	0.3236
$\hat{h}_{0,g}^{opt}$ (under dependence)	0.3745
$\hat{h}_{0,g,as}^{opt}$ (under independence)	0.1669
$\hat{h}_{0,g}^{opt}$ (under independence)	0.2148

Table 4: Plug-in bandwidths obtained with data of monthly accumulated R.P.I. from January 1979 to August 1997 in Spain.

Figure 3 shows the scatterplots of the R.P.I. and two estimators of the regression function, the first obtained under dependence with bandwidth  $\hat{h}_{0,g}^{opt} = 0.3745$  (bold line) and the second, under independence, with bandwidth  $\hat{h}_{0,g}^{opt} = 0.2148$  (thin line).

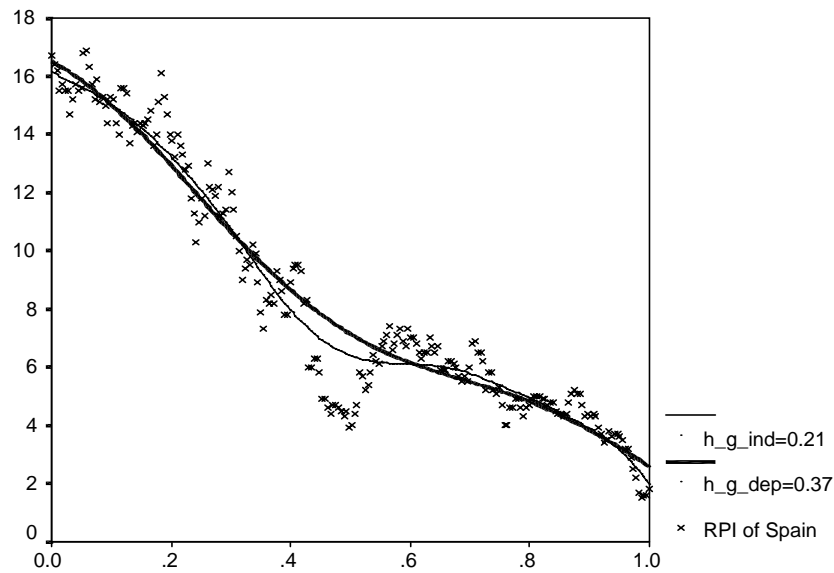


Figure 3: Graphs of R.P.I. and  $\hat{m}_h(x)$  with  $h = 0.3745$  and  $h = 0.2148$ .

## 5. PROOFS

In this section, we sketch proofs of the results presented in Section 2.

### PROOF OF PROPOSITION 1.

Let  $k_n$  be the number of elements of  $J_x = \{i / x_i \in (x - h, x + h)\}$ . From the design (2) of the regression model it follows that  $k_n = O(nh_n)$ . Using assumption A1, we obtain that

$$s_j^{(n)} = \int_0^1 (u - x)^j K_n(u - x) f(u) du + O\left(\frac{h_n^{j-1}}{n}\right), \quad 0 \leq j \leq 2p. \quad (36)$$

Now, changing variable  $y = \frac{u - x}{h}$  and using Taylor's expansion of  $f(x + yh_n)$ , we have

$$s_j^{(n)} = h_n^j \mu_j f(x) + o(h_n^j) + O\left(\frac{h_n^{j-1}}{n}\right), \quad 0 \leq j \leq 2p,$$

from which, using assumption A2, it follows (7) and (8).

### PROOF OF PROPOSITION 2.

If the Mean Value Theorem with integral remainder is considered, then we find

$$K\left(\frac{x_t - x}{h_n}\right) = K\left(\frac{x_i - x}{h_n}\right) + \frac{x_t - x_i}{h_n} \left(\int_0^1 K'\left(\frac{x_i - x + z(x_t - x_i)}{h_n}\right) dz\right),$$

and

$$(x_t - x)^r = (x_i - x)^r + (x_t - x_i) \left(\int_0^1 r(x_i - x + z(x_t - x_i))^{r-1} dz\right).$$

Using these two equations, we can deduce for  $j, r = 0, 1, \dots, p$ ,

$$Cov\left(h_n^{-j} t_{j,(n)}^*, h_n^{-r} t_{r,(n)}^*\right) \leq$$

$$\begin{aligned}
& \frac{1}{h_n^{j+r}} \frac{\sigma^2}{n^2 h_n^2} \sum_{i=1}^n (x_i - x)^{j+r} K^2 \left( \frac{x_i - x}{h_n} \right) \sum_{t=1}^n c(|i - t|) + o \left( \frac{1}{nh_n} \right) \\
&= \Delta_1 + o \left( \frac{1}{nh_n} \right). \tag{37}
\end{aligned}$$

Using arguments similar to those used in the proof of Proposition 1, it follows that,

$$\Delta_1 = \frac{f(x)c(\varepsilon)}{nh_n} \int_{-1}^1 y^{j+r} K^2(y) dy + o \left( \frac{1}{nh_n} \right). \tag{38}$$

So, from (37) and (38), we deduce (10) and (11).

### PROOF OF PROPOSITION 3.

A similar approach to that employed in Masry and Fan (1997) or Vilar-Fernández and Vilar-Fernández (1998) is used to prove asymptotic normality.

Let  $Q_n$  be an arbitrary linear combination of  $h_n^{-j} t_{j,(n)}^*$ ,

$$Q_n = \sum_{j=0}^p a_j h_n^{-j} t_{j,(n)}^* \quad \text{with } a_j \in \mathbb{R}. \tag{39}$$

If the asymptotic normality of  $\sqrt{nh_n} Q_n$  is established, then (20) is obtained from the Cramer-Wold theorem. To achieve this, we have used the well known “small-blocks and large-blocks” method.

From (9), it follows that  $E(Q_n) = 0$  and using Proposition 2 we obtain that

$$\lim_{n \rightarrow \infty} nh_n \text{Var}(Q_n) = f(x)c(\varepsilon) \int C^2(u) du = \sigma_Q^2(x). \tag{40}$$

We will show that  $\sqrt{nh_n} Q_n \rightarrow N(0, \sigma_Q(x))$ .

Denote

$$\sqrt{nh_n} Q_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i = \frac{1}{\sqrt{n}} S_n, \tag{41}$$

where

$$Z_i = C_h(x_i - x) (Y_i - m(x_i)) \sqrt{h_n},$$

with

$$C(u) = \sum_{j=0}^p a_j u^j K(u) \quad \text{and} \quad C_h(u) = \frac{1}{h_n} C(u/h_n).$$

The sum,  $S_n$ , is written in terms of small and large blocks as follows: for each  $n$ , let  $k_n = \left\lfloor \frac{n}{b_n + s_n} \right\rfloor$ , where  $b_n$  and  $s_n$  depend on  $n$  such that  $b_n \uparrow \infty$ ,  $s_n \uparrow \infty$ ,  $b_n/n \downarrow 0$ ,  $s_n/b_n \downarrow 0$ ,  $b_n$  is the size of the big block and  $s_n$  is the size of the small block. Then, let us split  $S_n$  into  $2k_n + 1$  summands in the following way:

$$S_n = \sum_{j=0}^{k_n-1} \Phi_j + \sum_{j=0}^{k_n-1} \Psi_j + S_{r,n} = S_{b,n} + S_{s,n} + S_{r,n}, \quad (42)$$

being

$$\Phi_j = \sum_{i=1}^{b_n} Z_{e_j+i}, \quad \Psi_j = \sum_{i=b_n+1}^{b_n+s_n} Z_{e_j+i}, \quad S_{r,n} = \sum_{i=e_{k_n}+1}^n Z_i, \quad (43)$$

with  $e_j = j(b_n + s_n)$ , for  $j = 0, \dots, k_n - 1$ . Thus, each  $\Phi_j$  represents a large block summing  $b_n$  variables, each  $\Psi_j$  is a small block summing  $s_n$  variables and, finally,  $S_{r,n}$  is a residual block.

We now choose the block sizes. By assumption A6, there exists a sequence,  $r_n \uparrow \infty$ , such as

$$r_n s_n = o\left((nh_n^3)^{1/2}\right) \quad \text{and} \quad r_n \left(\frac{n}{h_n}\right)^{1/2} \sum_{i=s_n}^{\infty} [\alpha(i)]^{1-\gamma} \rightarrow 0.$$

Define the large block size by  $b_n = \left\lfloor (nh_n)^{1/2} / r_n \right\rfloor$ , where  $[a]$  denotes the integer part of  $a$ . Then, it can easily be deduced that, as  $n \uparrow \infty$ ,

$$\frac{s_n}{b_n} \rightarrow 0, \quad \frac{b_n}{n} \rightarrow 0, \quad \frac{b_n}{(nh_n)^{1/2}} \rightarrow 0, \quad (44)$$

$$\frac{1}{h_n} \sum_{i=s_n}^{\infty} [\alpha(i)]^{1-\gamma} \rightarrow 0, \quad \text{and} \quad \frac{n}{b_n} \alpha(s_n) \rightarrow 0. \quad (45)$$

We will show that as  $n \uparrow \infty$

$$\frac{1}{n}E(S_{s,n}^2) \rightarrow 0 \quad \text{and} \quad \frac{1}{n}E(S_{r,n}^2) \rightarrow 0, \quad (46)$$

$$\left| E(e^{iuS_{b,n}}) - \prod_{j=0}^{k_n-1} E(e^{iu\Phi_j}) \right| \rightarrow 0, \quad (47)$$

$$\frac{1}{n} \sum_{j=0}^{k_n-1} E(\Phi_j)^2 \rightarrow \sigma_Q^2(x), \quad (48)$$

$$\frac{1}{n} \sum_{j=0}^{k_n-1} E(\Phi_j^2 I\{|\Phi_j| \geq \varepsilon \sigma_Q(x) \sqrt{n}\}) \rightarrow 0, \quad \forall \varepsilon > 0, \quad \varepsilon \in \mathbb{R}. \quad (49)$$

(46) implies that the sum of the small blocks,  $S_{s,n}$ , and the residual block,  $S_{r,n}$ , are asymptotically negligible. (47) implies that the summands  $\Phi_j$  in the large blocks,  $S_{b,n}$ , are asymptotically independent, and (48) and (49) are the standard Lindeberg-Feller condition for asymptotic normality of  $S_{b,n}$  under independence.

So, from (46)–(49) the asymptotic normality of (41) is deduced.

The proofs of (46)–(49) are similar to those in Masry and Fan (1997). Using (44), (45), Davydov inequality and assumptions A1, A4 and A5, (46) is established. Using Volkonskii and Rozanov (1959) lemma and (45) again, we obtain (47).

From (42) and (46), we can deduce

$$\frac{1}{n}E(S_{b,n})^2 = \frac{1}{n} \sum_{j=0}^{k_n-1} E(\Phi_j)^2 + \frac{2}{n} \sum_{i=0}^{k_n-1} \sum_{\substack{j=0 \\ i>j}}^{k_n-1} E(\Phi_i \Phi_j) \rightarrow \sigma_Q^2(x), \quad \text{as } n \uparrow \infty. \quad (50)$$

We have that using the arguments employed before for showing (46), we obtain

$$\frac{2}{n} \sum_{i=0}^{k_n-1} \sum_{\substack{j=0 \\ i>j}}^{k_n-1} E(\Phi_i \Phi_j) \rightarrow 0, \quad \text{as } n \uparrow \infty. \quad (51)$$

Now, (48) follows from (50) and (51).

Finally, to show (49), like Masry and Fan (1997), we employ a truncation argument. Let  $M$  be a positive real number and denote  $\varepsilon_{t,M} = \varepsilon_t I\{|\varepsilon_t| \leq M\}$  and  $\tilde{\varepsilon}_{t,M} = \varepsilon_t I\{|\varepsilon_t| > M\}$ . Then,  $\varepsilon_t = \varepsilon_{t,M} + \tilde{\varepsilon}_{t,M}$ . Let  $Q_{n,M}$  (and  $\tilde{Q}_{n,M}$ ) be the same linear combination as  $Q_n$  but replacing  $\varepsilon_t$  by  $\varepsilon_{t,M}$  (or  $\tilde{\varepsilon}_{t,M}$ ). Similar notation is used in other terms. Now, the errors,  $\varepsilon_{t,M}$ , are bounded, and condition (49) is satisfied.

By assumption A1 and using (44), we have

$$\frac{1}{\sqrt{n}} |\Phi_{j,M}| \leq \frac{1}{\sqrt{n}} \sum_{i=1}^{b_n} |Z_{e_j+i,M}| \leq C \frac{b_n}{\sqrt{nh_n}} \longrightarrow 0, \text{ as } n \uparrow \infty.$$

Therefore,  $\{|\Phi_{j,b,M}| \geq \varepsilon \sigma_{Q,M}(x) \sqrt{n}\}$  is an empty set when  $n$  is large enough, and therefore (49) holds and we conclude that  $\sqrt{nh_n} Q_{n,M}$  converges in distribution to  $N(0, \sigma_{Q,M}^2(x))$ .

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

$$\varphi_{Q_n}(t) \longrightarrow \varphi_Z^{\sigma_Q^2}(t), \text{ as } n \uparrow \infty, \quad (52)$$

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

We have

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

As  $n \uparrow \infty$ ,  $\sigma_{Q_{n,M}}^2$  tends to zero by the dominated convergence theorem when  $M \uparrow \infty$ . Therefore,  $\Delta_1$  goes to zero. Again, using the dominated convergence theorem, we obtain the convergence to zero of the second term. Finally, the convergence to zero of  $\Delta_3$  follows from (52) and the Levy theorem, for every  $M > 0$ . Now, the proof of Proposition 3 is complete.

## ACKNOWLEDGEMENTS

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



This work was partially supported by the grants XUGA10501B97 and PB98-0182-c02-01.

## BIBLIOGRAPHY

- Chu, C.K. and Marron, J.S. (1991). “Comparison of two bandwidth selectors with dependent errors,” *The Annals of Statistics*, **19**, 4, 1906–1918.
- Cleveland, W.S. and Devlin, S.J. (1988). “Locally weighted regression: an approach to regression analysis by local fitting,” *J. Am. Stat. Assoc.*, **83**, 597–610.
- Doukhan, P. (1995). *Mixing. Properties and examples*, Notes in Statistics, **85**, Berlin, Springer.
- Fan, J. (1992). “Design-adaptive nonparametric regression,” *J. Am. Stat. Assoc.*, **87**, 998–1004.
- Fan, J. (1993). “Local linear regression smoothers and their minimax efficiency,” *Ann. Stat.*, **21**, 196–216.
- Fan, J. and Gijbels, I. (1992). “Variable bandwidth and local linear regression smoothers,” *Ann. Stat.*, **20**, 2008–2036.
- Fan, J. and Gijbels, I. (1995). “Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation,” *J. Roy. Stat. Soc. B*, **57**, 2, 371–394.
- Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications*, Chapman and Hall. London.
- Fan, J., Gijbels, I., Hu, T. C. and Huang, L. S. (1996). “A study of variable bandwidth selection for local polynomial regression,” *Statistica Sinica*, **6**, 113–127.
- Fan, J., Gasser, T., Gijbels, I., Brockmann, M., and Engel, J. (1997). “Local polynomial regression: optimal kernels and asymptotic minimax efficiency,” *Ann. Inst. Stat. Math.*, **49**, 1, 79–99.

- Hardle, W. and Tsybakov, A. (1997). “Local polynomial estimators of the volatility function in nonparametric autoregression,” *J. of Econometrics*, **81**, 223–242.
- Hardle, W., Tsybakov, A., Yang L. (1998). “Nonparametric vector autoregression,” *J. of Statistical Planning and Inference*, **68**, 221–245.
- Hardle, W. and Vieu, P. (1992). “Kernel regression smoothing of time series,” *J. Time Ser. Anal.*, **13**, 209–232.
- Hart, J. (1991). “Kernel regression estimation with time series errors,” *J.R. Statist. Soc. B*, **53**, 173–187.
- Hart, J. (1994). “Automated kernel smoothing of dependent data by using time series cross-validation,” *J.R. Statist. Soc. B*, **56**, 3, 529–542.
- Hastie, T. and Loader, C. (1993). “Local regression: automatic kernel carpentry,” *Stat. Science*, **8**, 2, 120–143.
- Herrmann, E., Gasser, T. and Kneip, A. (1992). “Choice of bandwidth for kernel regression when residuals are correlated,” *Biometrika*, **79**, 4, 783–795.
- Lejeune, M. (1985). “Estimation non-paramétrique par noyaux: régression polynomiale mobile,” *Revue de Statist. Appl.*, **33**, 43–68.
- Masry, E. (1996a). “Multivariate regression estimation: local polynomial fitting for time series,” *Stoch. Proc. Appl.*, **65**, 81–101.
- Masry, E. (1996b). “Multivariate local polynomial regression for time series: uniform strong consistency and rates,” *J. Time Ser. Anal.*, **17**, 571–599.
- Masry, E. and Fan, J. (1997). “Local polynomial estimation of regression function for mixing processes,” *Scand. J. of Stat.*, **24**, 165–179.
- Müller, H.G. (1988). *Nonparametric regression analysis of longitudinal data*, Notes in Statistics, **46**, Berlin, Springer.
- Müller, H.G and Stadtmüller (1988). “Detecting dependencies in smooth regression models,” *Biometrika*, **75**, 4, 639–650.

- Opsomer, J.D. (1997). “Nonparametric regression in the presence of correlated errors,” *Modelling Longitudinal and Spatially Correlated Data: Methods, Applications and Future Directions*, (T.G. Gregoire, D.R. Brillinger, P.J. Diggle, E. Russek-Cohen, W.G. Warren and R.D. Wolfinger, Eds.), Springer, New York, 339-348.
- Quintela, A. (1994). “A plug-in technique in nonparametric regression with dependence,” *Comm. In Statist., Theory and Methods*, **23**, 9, 2581–2605.
- Rosenblatt, M.(1956) “A central limit theorem and strong mixing conditions,” *Proc. Nat. Acad. Sci.*, 4, 43–47.
- Ruppert, D. and Wand, P. (1994). “Multivariate locally weighted least squares regression,” *Ann. Stat.*, **22**, 1346–1370.
- Tsybakov, A. B. (1986). “Robust reconstruction of functions by the local-approximation method,” *Problems of Information Transmission*, **22**, 133–146.
- Vilar-Fernández, J.A. and Vilar-Fernández, J.M. (1998). “Recursive estimation of regression functions by local polynomial fitting,” *Ann. Ins. Stat. Math.*, **50**, 4, 729–754.
- Vilar-Fernández, J.M. and Vilar-Fernández, J.A. (2000). “Recursive local polynomial regression under dependence conditions,” *to appear in TEST*.
- Volkonskii, V.A. and Rozanov, Yu. A. (1959). “Some limit theorems for random functions,” *Theory Prob. Appl.*, **4**, 178–197.

## FIGURE CAPTIONS

FIG. 1. Bandwidths as a function of  $\rho$

FIG. 2.  $\log(\Delta MISE)$ , for every selector, as a function of  $\rho$ .

FIG. 3. Graphs of R.P.I. and  $\hat{m}_h(x)$  with  $h = 0.3745$  and  $h = 0.2148$ .