



# Weighted Local Nonparametric Regression with Dependent Errors: Study of Real Private Residential Fixed Investment in the USA

MARIO FRANCISCO-FERNANDEZ\* and JUAN M. VILAR-FERNANDEZ

*Department of Mathematics, University of A Coruña, 15071 A Coruña, Spain*

**Abstract.** This paper presents an overview of the existing literature on the nonparametric local polynomial (LPR) estimator of the regression function and its derivatives when the observations are dependent. When the errors of the regression model are correlated and follow an ARMA process, Vilar-Fernández and Francisco-Fernández (2002) proposed a modification of the LPR estimator, called the generalized local polynomial (GLPR) estimator, based on, first, transforming the regression model to get uncorrelated errors and then applying the LPR estimator to the new model. Some of the most significant asymptotic properties of these two weighted local estimators, including some guidelines on how to select the bandwidth parameter, are reviewed. Finally, these techniques are used to study the real private residential fixed investment in the USA.

**AMS 1991 Subject Classification:** 62G07, 62H12, 62M09.

**Key words:** local polynomial estimator, dependent data, smoothing parameter.

## 1. Introduction

The main goal of regression analysis is to estimate the mean function  $E(Y|X) = m(X)$  for a given set of observations  $\{(X_i, Y_i)\}_{i=1}^n$ , where the responses  $Y_i$  are scalar and the predictors (or design variables)  $X_i$  are either univariate or multivariate. To achieve this goal, a regression model

$$Y_t = m(X_t) + \varepsilon_t, \quad 1 \leq t \leq n, \quad (1)$$

where  $\varepsilon_t$  are random errors, is set up. If the  $X$ -variable is selected (sometimes equally spaced) by the experimenter, the model is called a fixed regression model. On the other hand, if the  $X$ -variable is random, it is referred to as a random regression model.

Nonparametric regression is attractive since it does not require a parametric form for the mean function. Because of recent theoretical developments and widespread use of fast and inexpensive computers, nonparametric regression has become a rapidly growing and exciting field of statistics. Researchers have realized

---

\*Author for correspondence: e-mail: mariofr@udc.es

that for many real data sets, parametric regression is not sufficiently flexible to adequately fit curves or surfaces. Recent monographs on nonparametric regression (Eubank, 1988; Müller, 1988; Härdle, 1990; Fan and Gijbels, 1996) have shown that for a variety of interesting examples, applications of nonparametric regression have yielded analyses essentially unobtainable by other techniques.

Basically, the nonparametric estimators of the regression function (sometimes called smoothers) may be classified as: kernel methods, spline methods and series expansion methods. Perhaps the most popular in the statistical community are the nonparametric kernel estimators. One of these kernel techniques is the local polynomial (LPR) estimator. The LPR estimator has gained wide acceptance as an attractive method for estimating the regression function and its derivatives. Some of the advantages of this nonparametric estimation method, compared with other kernel estimators like Nadaraya–Watson estimator or Gasser–Müller estimator, are better boundary behavior, adaptation to estimate regression derivatives, easy computation, good minimax properties, etc. Some significant references on this method, when the observations are independent, are, for example, Stone (1977), Cleveland (1979), Tsybakov (1986), Fan (1992, 1993), Ruppert and Wand (1994), Fan and Gijbels (1995) and Ruppert et al. (1995). See Fan and Gijbels (1996) for additional references.

Sometimes, the observations in the sample data cannot be assumed independent, for example, if they are gathered sequentially in time. In this case, the statistical properties of the LPR estimator presented in the papers mentioned above can change in this new setting of dependence.

In this article, some of the most relevant literature and methods regarding the LPR estimator under dependence are presented. Moreover, the application of these techniques to a real economic data set is shown.

Classical issues in nonlinear time series, like estimating time trend or constructing predictive intervals, compelled some authors to study the LPR estimator in a dependence context. Some related works are: Masry (1996a, b), Masry and Fan (1997), Härdle and Tsybakov (1997), Härdle et al. (1998) and Vilar-Fernández and Vilar-Fernández (1998). In these papers, a regression model considering a random data sample,  $\{(X_t, Y_t)\}_{t=1}^n$ , satisfying some mixing conditions, was used. Basically, under suitable conditions on the mixing coefficients, the results obtained show that the asymptotic properties of the LPR estimator under mixing conditions and under independence are the same. On the other hand, in Francisco-Fernández and Vilar-Fernández (2001), a regular fixed design regression model with short-range correlated errors is considered. These regression models frequently arise in economic studies, in the analysis of growth curves and usually in the study of time series with deterministic trend. In this case, while the asymptotic bias of the regression estimator is exactly the same as that obtained under independence, the asymptotic variance of the estimator changes. Now, the sum of the covariances of the errors in this term appears instead of simply the variance of the errors, as in the case of independent observations. Some of these asymptotic expressions will be

presented in the next Section. The properties of the LPR estimator under long-range dependent errors were recently studied in Anh et al. (1999), Masry and Mielniczuk (1999) and Beran and Feng (2002). Opsomer (1997) and Opsomer et al. (2001) provide other references on the topic of nonparametric regression in the presence of correlation.

As with any nonparametric regression procedure, with the LPR estimator one has to choose a tuning parameter, called the bandwidth or smoothing parameter, which determines the size of the neighborhood of the point to estimate. Adequate selection of this parameter is crucial for correct estimation of the regression function and its derivatives. If a very small bandwidth is used for the estimation, an extremely wiggly curve will be produced and if the bandwidth is too big, a very smooth curve will result. Some automated methods have been proposed in the literature to select this important parameter from the sample data when the observations are independent. Among them, perhaps the best known are the cross-validation techniques, plug-in methods and bootstrap methods. The direct application of these methods to dependent data usually produces bad results (see, e.g., Hart and Wehrly (1986) or Francisco-Fernández and Vilar-Fernández (2001)). In this case, several techniques have been proposed. These methods are basically a revision of those mentioned before, adapted to take the dependence of the observations into account.

The organization of this work is as follows: Section 2 includes some asymptotic properties of the LPR estimator in a fixed regression model with dependent errors. Moreover, another nonparametric estimator of the regression function (the generalized local polynomial, GLPR, estimator) based on merging ideas of the nonparametric LPR estimator with the parametric generalized least squares estimator will be defined. The asymptotic properties of this estimator will be briefly presented and compared with the properties of the classical LPR estimator. Other topics like bandwidth selection will be explained in this Section. Finally, in Section 3, these estimators will be used to analyze some real economic datasets and to extract some interesting conclusions.

## 2. The LPR Estimator Under Dependence

As stated in the previous Section, our goal is to estimate the unknown regression function and its derivatives based on an observed sample  $\{(X_t, Y_t)\}_{t=1}^n$ .

### 2.1. THE LPR ESTIMATOR

Assuming 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  $\boldsymbol{\beta}(x) = (\beta_0(x), \beta_1(x), \dots, \beta_p(x))^t$ , where  $\beta_j(x) = m^{(j)}(x)/(j!)$ , with  $j = 0, 1, \dots, p$ , by minimizing the function

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

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, that is,  $K$  is a symmetric and positive function with  $\int K = 1$ , and  $h_n$  the bandwidth or smoothing parameter. The estimator of  $\boldsymbol{\beta}(x)$ , obtained as a solution to the weighted least squares problem given in (2), is called the local polynomial kernel estimator and it is interesting to observe that this type of estimators includes the classical Nadaraya–Watson estimator, which is the minimizer of (2) when  $p = 0$ . Of special interest is also the local linear kernel estimator corresponding to  $p = 1$ .

In matrix notation, the minimizing problem (2) can be written as:

$$\min_{\boldsymbol{\beta}(x)} (\mathbf{Y}_{(n)} - X_{p,(n)} \boldsymbol{\beta}(x))^t W_{(n)} (\mathbf{Y}_{(n)} - X_{p,(n)} \boldsymbol{\beta}(x)), \quad (3)$$

where

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

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

$$\hat{\boldsymbol{\beta}}_{(n)}(x) = (X_{p,(n)}^t W_{(n)} X_{p,(n)})^{-1} X_{p,(n)}^t W_{(n)} \mathbf{Y}_{(n)}. \quad (4)$$

Then, the estimator of the  $j$ th derivative of the regression function,  $\hat{m}^{(j)}(x)$ , is given by  $\hat{\beta}_j(x)(j!)$ ,  $\hat{\beta}_j(x)$  being the  $(j + 1)$ th component of the vector in (4),  $j = 0, 1, \dots, p$ .

Obviously, these estimators can be used whether the observations are independent or not. This paper concerns the second case.

## 2.2. DEPENDENCE STRUCTURES IN A REGRESSION MODEL

The structure of dependence of the observations in a regression model is designed in two ways:

- *Random design.* A stationary sequence  $\{(X_i, Y_i)\}_{i=1}^n$ , which may be stochastically dependent, is observed and the conditional mean function  $m(x) = E(Y|X = x)$  is estimated based on this random sample.

In this context, the problem of predicting in time series can be included: let  $\{Z_i\}_{i=1}^n$  be a sample of a time series. By defining  $Z_i$  as  $X_i$  and  $Z_{i+l}$  as  $Y_i$ , the problem of predicting  $Z_{n+l}$  from  $\{Z_i\}_{i=1}^n$  can be considered as a nonparametric regression problem for dependent data, namely, to estimate  $m(z) = E(Z_{i+l}|Z_i = z)$  based on sample  $\{(X_i, Y_i)\}_{i=1}^{n-l} = \{(Z_i, Z_{i+l})\}_{i=1}^{n-l}$ .

- *Fixed design.* The observation errors  $\{\varepsilon_{i,n}\}$  in a fixed regression model

$$Y_{i,n} = m(x_{i,n}) + \varepsilon_{i,n}, \quad 1 \leq i \leq n \quad (5)$$

form a sequence of correlated random variables. In many cases,  $x_{i,n} = i/n$ ,  $1 \leq i \leq n$ , and the aim is to estimate the trend of a time series.

Some references studying the LPR estimator in each one of the previous situations were given in the Introduction. For the first setting, the assumption of observations satisfying a dependence mixing condition is usually considered. Among these mixing conditions, the  $\alpha$ -mixing condition is the most used and one of the least restrictive.  $\alpha$ -mixing condition is defined as: let  $\mathbb{N}$  denote the set of positive integers, and for any  $a$  and  $b$  in  $\mathbb{N} \cup \{\infty\}$  ( $a \leq b$ ), define  $\mathcal{F}_a^b$  as the  $\sigma$ -algebra of events generated by the random variables  $\{(X_i, Y_i) | a \leq i \leq b\}$ . The stationary process  $\{(X_i, Y_i), i \geq 1\}$  is called  $\alpha$ -mixing (or strong mixing) if for a sequence  $\{\alpha_k\}$  tending to zero,

$$\sup_{A \in \mathcal{F}_1^n, B \in \mathcal{F}_{n+k}^\infty} |P(A \cap B) - P(A)P(B)| \leq \alpha_k,$$

for all  $n \geq 1$ ,  $k \geq 1$ . This condition is satisfied by many processes, for example, by ARMA processes which are generated by continuous white noise. A wide and complete study on this and other mixing conditions can be seen in Bradley (1986) and Doukhan (1995).

With respect to the second context, given in (5), classically, it is assumed that the design points  $x_{i,n}$  are within a finite interval  $[a, b]$  (for simplicity,  $[0, 1]$ ) and the errors satisfying that

$$E(\varepsilon_{i,n}) = 0, \quad \text{Var}(\varepsilon_{i,n}) = \sigma^2, \quad \text{Corr}(\varepsilon_{i,n}, \varepsilon_{j,n}) = \rho_n(x_{i,n} - x_{j,n}), \quad (6)$$

with  $\sigma^2$  and  $\rho_n$  unknown. The dependence of the correlation  $\rho_n$  on the sample size is indicated by the subscript. Researchers who have studied the properties of kernel-based estimators of the regression function have mainly focused on the time series case, in which the design points are fixed and usually equally spaced, that is,

$$Y_i = m\left(\frac{i}{n}\right) + \varepsilon_{i,n}, \quad i = 1, 2, \dots, n,$$

where  $\{\varepsilon_{1,n}, \dots, \varepsilon_{n,n}\}$  is a sample from a covariance stationary time series with zero mean, so,

$$E(\varepsilon_{i,n}) = 0, \quad i = 1, 2, \dots, n$$

and

$$\text{Corr}(\varepsilon_{i,n}, \varepsilon_{i+k,n}) = \rho_n(|k|), \quad |k| = 0, 1, \dots \quad (7)$$

The consistency properties of the LPR estimator will depend on the behavior of the correlation function as  $n$  increases. To understand this behavior, let us consider a sequence of positive integers  $\{k_n(u)\}$ , such that  $\lim_{n \rightarrow \infty} k_n(u)/n = u \in [0, 1]$ . Then, for large  $n$ ,  $\rho_n(k_n(u))$  is the correlation between observations  $Y$ 's whose design points,  $i/n$  and  $(i + k_n(u))/n$ , are separated by a distance  $u$  ( $\lim_{n \rightarrow \infty} ((i + k_n(u))/n) - i/n = \lim_{n \rightarrow \infty} ((k_n(u))/n) = u$ ). Two cases of interest exist:

CASE 1

$$\lim_{n \rightarrow \infty} \rho_n(k_n(u)) = \rho(u), \quad (8)$$

for each  $u \in [0, 1]$ , where  $\rho$  is a real-valued function.

CASE 2

$$\lim_{n \rightarrow \infty} \rho_n(k_n(u)) = I_{\{0\}}(u), \quad (9)$$

for each  $u \in [0, 1]$ .

In Case 1, the correlation between data values depends on the distance of the corresponding design points. The observed data are, in the limit, a sample path of a continuous time process on  $[0, 1]$ . This process was discussed by Hart and Wehrly (1986). They have shown that if only a single realization of the process has been observed, there is no consistent linear estimator for the mean function as the design points are sampled more and more densely on the unit interval (see also Fraiman and Meloche (1994)).

In the second case, data values at design points a fixed distance apart are asymptotically uncorrelated. This situation includes the important case  $\rho_n(k) = \rho(k)$  for all  $n$  and  $k$ . This happens if a sample of a time series is taken and the design variable (time) is rescaled into the unit interval. This kind of dependence is assumed in this paper.

### 2.3. SOME ASYMPTOTIC PROPERTIES OF THE LPR ESTIMATOR

In what follows, a fixed regression model like (5) is assumed, where  $m(x)$  is a 'smooth' regression function defined on  $[0, 1]$ , the errors  $\varepsilon_{t,n}$ ,  $1 \leq t \leq n$ , are a sequence of unobserved random variables with zero mean and finite variance  $\sigma_\varepsilon^2$ , where, for each  $n$ ,  $\{\varepsilon_{1,n}, \varepsilon_{2,n}, \dots, \varepsilon_{n,n}\}$  have the same joint distribution as  $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n\}$ , with  $\{\varepsilon_t, t \in \mathbb{Z}\}$  being a strictly stationary stochastic process. Also, it is assumed that the 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

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

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

Considering this regression model and under usual conditions on the kernel function  $K$ , the sequence of bandwidths  $\{h_n\}$ , and assuming that the covariance between the errors is given by  $\text{Cov}(\epsilon_i, \epsilon_{i+k}) = \sigma_\epsilon^2 c(k)$ , with  $\sum_{k=1}^{\infty} k|c(k)| < \infty$ , Francisco-Fernández and Vilar-Fernández (2001) obtained the asymptotic expressions for the bias and the variance–covariance matrix of LPR estimator (4). These expressions are:

$$H_{(n)}\text{E}\left(\hat{\beta}_{(n)}(x) - \beta(x)\right) = \frac{m^{(p+1)}(x)}{(p+1)!} h_n^{p+1} S^{-1} \boldsymbol{\mu} + o\left(h_n^{p+1}(1, \dots, 1)^t\right) \quad (10)$$

and

$$\text{Var}\left(H_{(n)}\hat{\beta}_{(n)}(x)\right) = \frac{1}{nh_n} \frac{c(\epsilon)}{f(x)} S^{-1} \tilde{S} S^{-1} + o\left(\frac{1}{nh_n}\right), \quad (11)$$

where  $H_{(n)} = \text{diag}(1, h_n, h_n^2, \dots, h_n^p)$ .  $S$  is the  $(p+1) \times (p+1)$  matrix whose  $(i+1, j+1)$ th element is  $s_{i,j} = \mu_{i+j}$ ,  $i, j = 0, \dots, p$ , with  $\mu_r = \int u^r K(u) du$ . The vector  $\boldsymbol{\mu} = (\mu_{p+1}, \dots, \mu_{2p+1})^t$ .  $\tilde{S}$  is the  $(p+1) \times (p+1)$  matrix whose  $(i+1, j+1)$ th element is  $\tilde{s}_{i,j} = v_{i+j}$ ,  $i, j = 0, \dots, p$ , with  $v_r = \int u^r K^2(u) du$  and  $c(\epsilon) = \sigma_\epsilon^2(c(0) + 2\sum_{k=1}^{\infty} c(k))$ .

Asymptotic expressions for the bias and the variance of the regression function estimator and its derivatives are directly derived from (10) and (11).

$$\text{Bias}\left(\hat{m}^{(j)}(x)\right) = h_n^{p+1-j} \frac{m^{(p+1)}(x)}{(p+1)!} j! B_j (1 + o(1)), \quad (12)$$

$$\text{Var}\left(\hat{m}^{(j)}(x)\right) = \frac{1}{nh_n^{2j+1}} \frac{c(\epsilon)}{f(x)} (j!)^2 V_j (1 + o(1)), \quad (13)$$

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

Moreover, in the paper of Francisco-Fernández and Vilar-Fernández (2001), the asymptotic normality of  $\hat{\beta}_{(n)}(x)$  and of  $\hat{m}^{(j)}(x)$  are obtained. For this, additional assumptions on the errors of the regression model were made: the  $(2 + \delta)$ -moment of the errors exists for some  $\delta > 0$ , and these are  $\alpha$ -mixing with the mixing coefficients satisfying some summability conditions.

Masry and Fan (1997) studied the LPR estimator with random design and  $\rho$ -mixing and  $\alpha$ -mixing observations. In their Theorem 5, they obtained similar results as those stated before, but in their case the asymptotic variance expression under dependence coincides with that for independent observations. That is, (11) changes, now  $\sigma_\epsilon^2$  appearing instead of  $c(\epsilon)$ . So, with a random design, the asymptotic mean squared error (AMSE) of the LPR estimator is the same under dependence as under independence, but in a fixed design with positive correlated errors the asymptotic variance of the LPR estimator is larger than with independence. The reason for this fact is that in random design, the random variables  $K_n(X_i - x)$

and  $K_n(X_j - x)$  are nearly uncorrelated, as  $h_n \rightarrow 0$ , but in a fixed design, the spatial distance between the  $x$ 's variables coincides with distance in time and the data points  $x_i$  closer to  $x$  are strongly dependent between them.

#### 2.4. BANDWIDTH SELECTION FOR THE LPR ESTIMATOR

Once the asymptotic properties of LPR estimator under dependence are obtained, the next step is to provide guidelines on the important problem of bandwidth selection. Considering the work of Francisco-Fernández and Vilar-Fernández (2002), an asymptotic plug-in method is proposed to obtain the smoothing parameter which can provide local or global bandwidths, depending on the choice of a local or global measure of the estimation error. A local bandwidth is required if the aim is to estimate the regression function at a point  $x_0$ ,  $m(x_0)$ . This happens in local problems, such as the problem of predicting. On the other hand, it will be interesting to obtain a global bandwidth when the estimation of the whole regression function is the main interest of the problem. In the above paper, seven of the most important techniques (including cross-validation methods, plug-in methods and bootstrap methods) used to select the smoothing parameter for the LPR estimator in a fixed regression model with dependent errors are described and compared through a simulation study. Although it is difficult to find a single smoothing parameter showing the best performance in all situations, the asymptotic plug-in bandwidth revealed a remarkable behavior. Moreover, this smoothing parameter is easy and fast to compute.

The asymptotic plug-in method consists in finding the bandwidth that minimizes the 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 with some estimators. In this particular case, using the asymptotic expressions of the bias and variance, (12) and (13), the AMSE of  $\hat{m}^{(j)}(x)$  is given by

$$\text{AMSE}(\hat{m}^{(j)}(x)) = \left( h_n^{p+1-j} \frac{m^{(p+1)}(x)}{(p+1)!} j! B_j \right)^2 + \frac{1}{nh_n^{2j+1}} \frac{c(\varepsilon)}{f(x)} (j!)^2 V_j. \quad (14)$$

So, minimizing expression (14) in  $h_n$ , the asymptotically optimal local bandwidth to estimate the  $j$ th derivative of the regression function is obtained. This bandwidth is given by

$$h_{j,l,as}^{\text{opt}}(x) = C_{j,p}(K) \left( \frac{c(\varepsilon)}{n(m^{(p+1)}(x))^2 f(x)} \right)^{1/(2p+3)}, \quad (15)$$

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



On the other hand, if the AMISE is used, the asymptotically optimal global bandwidth to estimate the  $j$ th derivative of the regression function is obtained, given by

$$h_{j,g,as}^{\text{opt}} = C_{j,p}(K) \left( \frac{c(\varepsilon)}{n \int (m^{(p+1)}(x))^2 f(x) dx} \right)^{1/(2p+3)}. \quad (16)$$

In (15) and (16), there are two unknown quantities,  $c(\varepsilon)$  and  $m^{(p+1)}(x)$ . These must be estimated in order to produce practical smoothing parameters from (15) and (16). Once  $c(\varepsilon)$  and  $m^{(p+1)}(x)$  are replaced by estimators, the computed bandwidths are denoted by  $\hat{h}_{j,l,as}^{\text{opt}}(x)$  and  $\hat{h}_{j,g,as}^{\text{opt}}$ , respectively.

To estimate  $c(\varepsilon)$  and  $m^{(p+1)}(x)$  various possibilities arise. With respect to the sum of the covariances,  $c(\varepsilon)$ , a parametric dependence structure, depending on a small number of parameters (e.g. an ARMA model), can be assumed and then, an estimator of  $c(\varepsilon)$  can be obtained by simply estimating these parameters, using, for example, nonparametric residuals. To do this, a pilot bandwidth,  $h_{\text{pilot}}$ , is needed to evaluate the LPR estimator at the design points and obtain the residuals given by  $\hat{\varepsilon}_t = Y_t - \hat{m}_{h_{\text{pilot}}}(x_t)$ ,  $t = 1, 2, \dots, n$ . Another possibility is to estimate  $c(\varepsilon)$  directly from the data using difference-based methods. Along these lines, Müller and Stadtmüller (1988) proposed an estimator for  $c(\varepsilon)$  based on first order differences of  $Y_t$ , for  $m$ -dependent errors. For errors that are not necessarily  $m$ -dependent but satisfy some mixing conditions, Hermann et al. (1992) suggested an estimator of  $c(\varepsilon)$  based on second order differences of  $Y_t$ . Recently, Hall and Van Keilegom (2003) showed that difference-based methods can be used to construct simple and explicit estimators of  $c(\varepsilon)$  and autoregressive parameters in nonparametric regression with time-series errors.

As far as  $m^{(p+1)}(x)$  is concerned, basically, a parametric or nonparametric fit can be used. In the former,  $m^{(p+1)}(x)$  is estimated by globally fitting to the regression function a polynomial of  $(p+3)$  order,  $\tilde{m}_{p+3}(x)$ , and next by calculating the  $(p+1)$  derivative of this fitted curve,  $\tilde{m}_{p+3}^{(p+1)}(x)$ . The nonparametric approximation involves estimating  $m^{(p+1)}(x)$  using the LPR estimator as explained in Section 2.1, for which a pilot bandwidth is necessary.

In Francisco-Fernández et al. (2002), the asymptotic global bandwidth obtained from (16) was studied for the local linear estimator. In this paper, the authors propose to estimate  $c(\varepsilon)$  from nonparametric residuals, assuming a parametric ARMA structure for the errors. For  $\int (m''(x))^2 f(x) dx$  they use  $1/n \sum_{i=1}^n (\hat{m}_g''(x_i))^2$  as estimator, where  $\hat{m}_g''(x)$  is the LPR estimator of the second derivative of the regression function obtained from (4) with a pilot bandwidth  $g$ . In this paper, the rate of convergence of the plug-in bandwidth considered to the optimal bandwidth that minimizes the mean integrated squared error is obtained. This rate of convergence is of the same order as that obtained under independence (see Ruppert et al., 1995). Moreover, a simulation study is shown where  $\hat{h}_{0,g,as}^{\text{opt}}$  is compared with two other bandwidth estimators: the same bandwidth as  $\hat{h}_{0,g,as}^{\text{opt}}$  but considering that the errors

are independent, that is, assuming that  $c(\varepsilon) = \sigma_\varepsilon^2$ , and the bandwidth proposed in Opsomer (1997) where the estimate of  $c(\varepsilon)$  is constructed by estimating the spectral density at 0 nonparametrically. In the three cases the pilot bandwidth needed to estimate  $c(\varepsilon)$  was computed by the time series cross-validation (TSCV) method, proposed in Hart (1994). The general conclusion is that it is important to take the dependence of the observations into account when a bandwidth is chosen and to use this information in the computation of the bandwidth. Besides, parametric estimation of the correlation function seems to provide a more reliable approach for estimating the optimal bandwidth, at least in that context.

Two reasons exist for selecting the TSCV bandwidth as a pilot bandwidth to estimate the regression function: first, this parameter showed good performance in the simulation study presented in Francisco-Fernández and Vilar-Fernández (2002) in a fixed regression model with autoregressive errors, and second, this technique is completely automatic, that is, it does not need a preliminary parameter for its computation. Therefore, the TSCV bandwidth is a good method for selecting the smoothing parameter when the observations are dependent. This method is a modification of the classical cross-validation method, but here considering the dependence of the observations. The cross-validation method, that finds the bandwidth  $h$  that minimizes

$$\text{CV}(h) = n^{-1} \sum_{i=1}^n (Y_i - \tilde{m}_{h,i}(x_i))^2 w(x_i), \quad (17)$$

with the estimator  $\tilde{m}_{h,i}$  used with the sample without the  $(x_i, Y_i)$  data and  $w$  a weight function introduced to allow reduction of boundary effects, produces reasonable bandwidths when the data are uncorrelated. This occurs because, in this case,

$$\text{E}(Y_j | Y_1, \dots, Y_{j-1}, Y_{j+1}, \dots, Y_n) = m(x_j), \quad (18)$$

that is, a good predictor of  $Y_j$  and a good estimator of  $\text{E}(Y_j)$  are one and the same. But if the error terms in the model are correlated, then

$$\text{E}(Y_j | Y_1, \dots, Y_{j-1}, Y_{j+1}, \dots, Y_n) = m(x_j) + g(\varepsilon_1, \dots, \varepsilon_{j-1}, \varepsilon_{j+1}, \dots, \varepsilon_n), \quad (19)$$

for some function  $g$  and  $\varepsilon_j = Y_j - m(x_j)$ ,  $j = 1, 2, \dots, n$ . So now, the best mean squared error predictor depends on  $Y_1, \dots, Y_{j-1}, Y_{j+1}, \dots, Y_n$ , and it is not necessarily a good estimate of the mean.

To solve this problem the TSCV method uses expression (19) and selects a model that yields good predictions of future observations. The idea is similar to that in the cross-validation method, but the TSCV method uses only  $Y_1, \dots, Y_{j-1}$  to predict  $Y_j$ .

To explain how the TSCV method works, let us suppose, for simplicity, that the errors follow an AR(1) process, that is,

$$\varepsilon_j = \rho \varepsilon_{j-1} + e_j, \quad j = 2, \dots, n,$$

where  $\rho$  is unknown and  $e_2, \dots, e_n$  are iid random variables. If  $\rho$  and the trend,  $m(x)$ , were known, the best mean squared error predictor of  $Y_j$  based on past data would be

$$\hat{Y}_j = m(x_j) + \rho (Y_{j-1} - m(x_{j-1})).$$

However, if the trend is unknown, it can be estimated by using a kernel smoother based on the data through time  $j - 1$ . Define a predictor

$$\hat{Y}_j(a, h) = \tilde{m}_{jh}(x_j) + a(Y_{j-1} - \tilde{m}_{jh}(x_{j-1})),$$

where  $\tilde{m}_{jh}(x_k)$  is a kernel estimate of  $m(x_k)$ ,  $k = j - 1, j$ , that uses only  $Y_1, \dots, Y_{j-1}$  and has smoothing parameter  $h$ . The TSCV method consists in choosing  $h$  and  $a$  to minimize

$$P(a, h) = \frac{1}{n-1} \sum_{j=2}^n \left( \hat{Y}_j(a, h) - Y_j \right)^2. \quad (20)$$

For any given  $h$ , the minimizer  $\hat{\rho}_h$  of  $P(a, h)$  with respect to  $a$  can be obtained explicitly. This leads to a criterion function  $P(h) = P(\hat{\rho}_h, h)$  whose minimum with respect to  $h$  may be approximated numerically.

The method can be generalized by assuming that the errors process,  $\{\varepsilon_j = Y_j - m(x_j)\}$ , follows a given parametric model  $M_\theta$ . Then, a predictor of  $Y_j$  of the form

$$\hat{Y}_j(h, \theta) = \tilde{m}_{jh}(x_j) + g_\theta(\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_{j-1}) \quad (21)$$

can be considered, where, as previously,  $\tilde{m}_{jh}(x_k)$  is a kernel estimate of  $m(x_k)$  that uses only the data  $Y_1, \dots, Y_{j-1}$ ,  $\hat{\varepsilon}_i = Y_i - \tilde{m}_{jh}(x_i)$ ,  $i = 1, \dots, j - 1$  and  $g_\theta(\varepsilon_1, \dots, \varepsilon_{j-1})$  is the optimal mean squared error predictor of  $\varepsilon_j$ , given  $\varepsilon_1, \dots, \varepsilon_{j-1}$ . Therefore, as with AR(1) errors, in the TSCV method,  $h$  and  $\theta$  are chosen to minimize

$$P(h, \theta) = \sum_{j=2}^n \left( \hat{Y}_j(h, \theta) - Y_j \right)^2. \quad (22)$$

Since TSCV uses only the data to the left of a time point  $j$ , Hart (1994) suggests that  $\tilde{m}_{jh}(x_k)$  be a kernel smoother using a boundary-type kernel, say  $L$ . However, if the aim is to efficiently estimate the trend,  $m(x_j)$ , using data on either side of time point  $j$  with a kernel  $K$ , it will be necessary to multiply the bandwidth minimizing (22) by a constant  $C_{K,L}$  depending on the kernels  $K$  and  $L$  (see Hart (1994) for this constant).

Other methods used to select the smoothing parameter for LPR estimator in the regression model considered in Section 2.3 are: the modified cross-validation method, the partitioned cross-validation method, the cross-validation for prediction, the exact plug-in and bootstrap methods. See Francisco-Fernández and Vilar-Fernández (2002) for a description and comparison using a simulation study of these techniques. It appears that the plug-in methods, the TSCV method and the bootstrap method give good results if dependence is considered.

### 3. A Modification of the LPR Estimator

In this Section, a modification of the LPR estimator of the regression function and its derivatives is introduced. Moreover, some asymptotic properties of these estimators are presented and compared with those obtained for the LPR estimator in the previous Section.

#### 3.1. MOTIVATION AND DEFINITION OF THE NEW ESTIMATOR

Another way to estimate the regression function, the parametric perspective, is to assume for this function a parametric model, depending on some parameters, and estimate the regression function by simply estimating the unknown parameters. For example, if a multiple linear regression model for the regression function is assumed, that is,

$$m(\mathbf{x}) = m(x_1, x_2, \dots, x_p) = \boldsymbol{\beta}\mathbf{x} = \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p, \quad (23)$$

the vector  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)$  has to be estimated to obtain a parametric estimator of  $m$ . If a sample data,  $\{(\mathbf{x}_i, Y_i)\}_{i=1}^n$ , being  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})^t$ ,  $i = 1, 2, \dots, n$ , has been observed and assuming that the regression function is as that given in (23), the observations satisfy

$$Y_i = \boldsymbol{\beta}\mathbf{x}_i + \varepsilon_i, \quad i = 1, 2, \dots, n. \quad (24)$$

The least squares (LS) method consists in finding the vector  $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)$  that minimizes the function

$$\Phi(\boldsymbol{\beta}) = \sum_{t=1}^n \left( Y_t - \sum_{j=1}^p \beta_j x_{tj} \right)^2. \quad (25)$$

Problem (25) can be written in matrix form,

$$\min_{\boldsymbol{\beta}} \Phi(\boldsymbol{\beta}) = \min_{\boldsymbol{\beta}} (\mathbf{Y}_{(n)} - X\boldsymbol{\beta})^t (\mathbf{Y}_{(n)} - X\boldsymbol{\beta}), \quad (26)$$

where

$$\mathbf{Y}_{(n)} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \quad \text{and} \quad X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}.$$

Simple algebra leads to the solution of (26), given by

$$\hat{\boldsymbol{\beta}}_{\text{LS}} = (X^t X)^{-1} X^t \mathbf{Y}_{(n)}. \quad (27)$$

If the errors vector,  $\boldsymbol{\varepsilon}_{(n)} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$ , satisfies  $E(\boldsymbol{\varepsilon}_{(n)}) = \mathbf{0}$  and  $E(\boldsymbol{\varepsilon}_{(n)} \boldsymbol{\varepsilon}_{(n)}^t) = \sigma_\varepsilon^2 I_n$  ( $I_n$  being the identity matrix of order  $n$ ), the estimator given in (27) is unbiased, that is,

$$E(\hat{\boldsymbol{\beta}}_{\text{LS}}) = \boldsymbol{\beta}$$

and its variance–covariance matrix is given by

$$\text{Var}(\hat{\boldsymbol{\beta}}_{\text{LS}}) = \sigma_\varepsilon^2 (X^t X)^{-1}.$$

Moreover, according to the Gauss-Markov theorem,  $\hat{\boldsymbol{\beta}}_{\text{LS}}$  is the BLUE (best linear unbiased estimator) estimator of  $\boldsymbol{\beta}$ .

On the other hand, if  $E(\boldsymbol{\varepsilon}_{(n)}) = \mathbf{0}$  and  $E(\boldsymbol{\varepsilon}_{(n)} \boldsymbol{\varepsilon}_{(n)}^t) = \sigma_\varepsilon^2 \Psi$ , where  $\Psi \neq I_n$  is a known symmetric and definite positive matrix (the correlation matrix of  $\boldsymbol{\varepsilon}_{(n)}$ ), the LS estimator is unbiased and its variance–covariance matrix is

$$\text{Var}(\hat{\boldsymbol{\beta}}_{\text{LS}}) = \sigma_\varepsilon^2 (X^t X)^{-1} X^t \Psi X (X^t X)^{-1}.$$

Since  $\Psi \neq I_n$ ,  $\text{Var}(\hat{\boldsymbol{\beta}}_{\text{LS}})$  is different from that obtained when the errors are independent,  $\sigma_\varepsilon^2 (X^t X)^{-1}$ .

Alternatively, when the errors are dependent, instead of directly using the LS estimator, first the model can be transformed to get uncorrelated errors and then obtain the LS estimator from the new transformed data sample. This estimator is called the generalized least squares (GLS) estimator.

So, the first step in the computation of the GLS estimator is to find a matrix  $P_{(n)}$ , such that

$$P_{(n)} \Psi P_{(n)}^t = I_n. \quad (28)$$

Since  $\Psi$  is symmetric and positive definite, a matrix  $P_{(n)}$  which has the property (28) always exists. From (28), it follows that

$$\Psi = P_{(n)}^{-1} (P_{(n)}^t)^{-1} \quad \text{and} \quad \Psi^{-1} = P_{(n)}^t P_{(n)}.$$

Using  $P_{(n)}$  to transform the model yields

$$P_{(n)} \mathbf{Y}_{(n)} = P_{(n)} X \boldsymbol{\beta} + P_{(n)} \boldsymbol{\varepsilon}_{(n)}$$

or

$$\mathbf{Y}_{(n)}^* = X^* \boldsymbol{\beta} + \boldsymbol{\varepsilon}_{(n)}^*,$$

where  $\mathbf{Y}_{(n)}^* = P_{(n)}\mathbf{Y}_{(n)}$ ,  $X^* = P_{(n)}X$  and  $\boldsymbol{\varepsilon}_{(n)}^* = P_{(n)}\boldsymbol{\varepsilon}_{(n)}$ . It is easy to prove that

$$E(\boldsymbol{\varepsilon}_{(n)}^*) = 0 \quad \text{and} \quad E(\boldsymbol{\varepsilon}_{(n)}^* \boldsymbol{\varepsilon}_{(n)}^{*t}) = \sigma_\varepsilon^2 I_n.$$

Now, the LS method is applied to this model, that is, the vector,  $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)$ , that minimizes the expression

$$\begin{aligned} \Phi_G(\boldsymbol{\beta}) &= \min_{\boldsymbol{\beta}} (\mathbf{Y}_{(n)}^* - X^* \boldsymbol{\beta})^t (\mathbf{Y}_{(n)}^* - X^* \boldsymbol{\beta}) \\ &= \min_{\boldsymbol{\beta}} (\mathbf{Y}_{(n)} - X \boldsymbol{\beta})^t \underbrace{P_{(n)}^t P_{(n)}}_{\Psi^{-1}} (\mathbf{Y}_{(n)} - X \boldsymbol{\beta}), \end{aligned} \quad (29)$$

is computed and the GLS estimator of  $\boldsymbol{\beta}$  is given by

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = (X^{*t} X^*)^{-1} X^{*t} \mathbf{Y}_{(n)}^* = (X^t P_{(n)}^t P_{(n)} X)^{-1} X^t P_{(n)}^t P_{(n)} \mathbf{Y}_{(n)}. \quad (30)$$

Now,

$$E(\hat{\boldsymbol{\beta}}_{\text{GLS}}) = \boldsymbol{\beta}$$

and

$$\text{Var}(\hat{\boldsymbol{\beta}}_{\text{GLS}}) = \sigma_\varepsilon^2 (X^t \Psi^{-1} X)^{-1}.$$

This estimator is the best linear unbiased estimator in this case. In practice, the matrix  $\Psi^{-1} = P_{(n)}^t P_{(n)}$  is unknown and, in expression (30),  $P_{(n)}$  is replaced with an estimator of it,  $\hat{P}_{(n)}$ , which leads to the estimated generalized least squares (EGLS) estimator, given by

$$\hat{\boldsymbol{\beta}}_{\text{EGLS}} = (X^t \hat{P}_{(n)}^t \hat{P}_{(n)} X)^{-1} X^t \hat{P}_{(n)}^t \hat{P}_{(n)} \mathbf{Y}_{(n)}.$$

A thorough study of these parametric estimators of the regression function appears in Judge et al. (1988).

So, it is clear that in a parametric multiple linear regression model with dependent errors, it is better to use the GLS estimator than the LS estimator. On the other hand, the relationship between the LPR estimator and the LS estimator is obvious just by observing the minimizing problems (3) and (26). After all, the LPR estimator is obtained from a least squares problem, but local and weighted. Therefore, when the errors are dependent, one can take advantage of the benefits of the generalized method by previously transforming the observations to get the errors to be uncorrelated and then use the LPR estimator.

Mathematically, these ideas can be written as follows: Consider the fixed regression model with dependent errors given in (5). If the variance–covariance matrix of the errors is  $E(\boldsymbol{\varepsilon}_{(n)} \boldsymbol{\varepsilon}_{(n)}^t) = \sigma_\varepsilon^2 \Psi$ , with  $\Psi \neq I_n$ , a matrix  $P_{(n)}$  such that

$\Psi^{-1} = P_{(n)}^t P_{(n)}$  exists. Performing a Taylor series expansion in a neighborhood of  $x$ , one obtains

$$m(x_t) = \sum_{j=0}^p \frac{m^{(j)}(x)}{j!} (x_t - x)^j + \frac{m^{(p+1)}(x)}{(p+1)!} (x_t - x)^{p+1} + o(x_t - x)^{p+1},$$

$$t = 1, \dots, n, \quad (31)$$

or, in matrix form,

$$\mathbf{M}_{(n)} = X_{(n)} \boldsymbol{\beta}(x) + \frac{m^{(p+1)}(x)}{(p+1)!} \begin{pmatrix} (x_1 - x)^{p+1} \\ \vdots \\ (x_n - x)^{p+1} \end{pmatrix} + \begin{pmatrix} o(x_1 - x)^{p+1} \\ \vdots \\ o(x_n - x)^{p+1} \end{pmatrix}, \quad (32)$$

where  $\mathbf{M}_{(n)} = (m(x_1), m(x_2), \dots, m(x_n))^t$ . So, model (5) can be approximated by

$$\mathbf{Y}_{(n)} \approx X_{(n)} \boldsymbol{\beta}(x) + \boldsymbol{\varepsilon}_{(n)}. \quad (33)$$

The errors of the following model are uncorrelated:

$$P_{(n)} \mathbf{Y}_{(n)} = P_{(n)} X_{(n)} \boldsymbol{\beta}(x) + P_{(n)} \boldsymbol{\varepsilon}_{(n)}. \quad (34)$$

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

$$\tilde{\beta}_{G,n}(x) = (X_{(n)}^t P_{(n)}^t W_{(n)} P_{(n)} X_{(n)})^{-1} X_{(n)}^t P_{(n)}^t W_{(n)} P_{(n)} \mathbf{Y}_{(n)}. \quad (35)$$

This estimator is called the GLPR estimator.

In practice, the estimator given in (35) cannot be computed because matrix  $P_{(n)}$  is unknown. Once  $P_{(n)}$  is estimated by a consistent estimator of it,  $\hat{P}_{(n)}$ , and plugged into (35), the estimated generalized local polynomial (EGLPR) estimator of  $\boldsymbol{\beta}(x)$  is given by

$$\hat{\beta}_{GE,n}(x) = (X_{(n)}^t \hat{P}_{(n)}^t W_{(n)} \hat{P}_{(n)} X_{(n)})^{-1} X_{(n)}^t \hat{P}_{(n)}^t W_{(n)} \hat{P}_{(n)} \mathbf{Y}_{(n)}, \quad (36)$$

where it is supposed that  $(X_{(n)}^t \hat{P}_{(n)}^t W_{(n)} \hat{P}_{(n)} X_{(n)})^{-1}$  exists.

### 3.2. ASYMPTOTIC ANALYSIS

In Vilar-Fernández and Francisco-Fernández (2002), the GLPR estimator and the EGLPR estimator are defined and some asymptotic properties (asymptotic bias, asymptotic variance and asymptotic normality), when the errors have an ARMA dependence structure, are shown.

In the particular case that the errors of the regression model  $\varepsilon_t$  follow an AR(1) type correlation structure, that is,

$$\varepsilon_t = \rho \varepsilon_{t-1} + e_t, \quad t \in \mathbb{Z},$$

with  $|\rho| < 1$  and  $\{e_t\}_{t \in \mathbb{Z}}$  a noise process with zero mean and finite variance,  $\sigma_e^2$ , the variance–covariance matrix of this process is  $E(\mathbf{e}_{(n)}\mathbf{e}_{(n)}^t) = \sigma_e^2 \Omega_{(n)}$ , where  $\Omega_{(n)}$  is a nonsingular matrix and positive definite, given by

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

So, in this case the transformation matrix  $P_{(n)}$  needed to obtain the GLPR estimator is

$$P_{(n)} = \begin{pmatrix} \sqrt{1 - \rho^2} & 0 & 0 & \dots & 0 \\ -\rho & 1 & 0 & \dots & 0 \\ 0 & -\rho & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & -\rho & 1 \end{pmatrix}. \quad (37)$$

A natural estimator of  $P_{(n)}$  to compute the EGLPR estimator is obtained by replacing  $\rho$  with its estimator  $\hat{\rho}$ ,

$$\hat{\rho} = \frac{\sum_{t=1}^{n-1} \hat{\varepsilon}_t \hat{\varepsilon}_{t+1}}{\sum_{t=1}^n \hat{\varepsilon}_t^2}, \quad (38)$$

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 LPR estimator.

Assuming usual conditions on the kernel function, the regression function and on the bandwidth, in the mentioned paper by Vilar-Fernández and Francisco-Fernández (2002), considering that the errors are AR(1), it is established that

$$\sqrt{nh_n} \left[ H_{(n)} \left( \tilde{\beta}_{G,n}(x) - \boldsymbol{\beta}(x) \right) - \frac{m^{(p+1)}(x)}{(p+1)!} h_n^{p+1} S^{-1} \boldsymbol{\mu} \right] \xrightarrow{\mathcal{L}} N_{(p+1)}(\mathbf{0}, \Sigma) \quad (39)$$

and

$$\sqrt{nh_n} \left[ H_{(n)} \left( \hat{\beta}_{GE,n}(x) - \boldsymbol{\beta}(x) \right) - \frac{m^{(p+1)}(x)}{(p+1)!} h_n^{p+1} S^{-1} \boldsymbol{\mu} \right] \xrightarrow{\mathcal{L}} N_{(p+1)}(\mathbf{0}, \Sigma), \quad (40)$$

where the variance–covariance matrix,  $\Sigma$ , is

$$\Sigma = \frac{\sigma_e^2}{f(x)} \frac{1 + \rho}{1 - \rho} S^{-1} \tilde{S} S^{-1} = \frac{\sigma_e^2}{f(x)} \frac{1}{(1 - \rho)^2} S^{-1} \tilde{S} S^{-1}.$$



From (39) and (40), it is easy to obtain the asymptotic normality of the estimators of the derivatives of the regression function as in (12) and (13).

If the expressions of the bias and variance given in (39) and (40) are compared with the same quantities for the LPR estimator in (10) and (11), it is seen that they are exactly the same. Hence, these three estimators (LPR, GLPR and EGLPR) are asymptotically equivalent. To illustrate the performance of the EGLPR estimator with finite samples, in the cited paper, a simulation study is shown where this estimator is compared with other kernel nonparametric estimators of the regression function: the Nadaraya–Watson estimator, the Gasser–Müller estimator and the local linear estimator. The general conclusion is that the EGLPR estimator behaves adequately for regression curve estimation and it is better than the other estimators compared, under the mean integrated squared error criterion, especially when the dependence is strong.

On this point, something should be said about the problem of bandwidth selection for the EGLPR estimator. In the case of AR(1) errors and taking into account that the asymptotic expressions for the bias and the variance of this new estimator are the same as those obtained for the LPR estimator, asymptotic local and global plug-in bandwidths could be designed to estimate the regression function and its derivatives with the same expressions as those given in (15) and (16). These would be

$$h_{j,l,as}^{\text{opt}}(x) = C_{j,p}(K) \left( \frac{\sigma_\varepsilon^2 ((1+\rho)/(1-\rho))}{n(m^{(p+1)}(x))^2 f(x)} \right)^{1/(2p+3)} \quad (41)$$

and

$$h_{j,g,as}^{\text{opt}} = C_{j,p}(K) \left( \frac{\sigma_\varepsilon^2 ((1+\rho)/(1-\rho))}{n \int (m^{(p+1)}(x))^2 f(x) dx} \right)^{1/(2p+3)}, \quad (42)$$

respectively. The unknown quantities appearing in (41) and (42),  $\sigma_\varepsilon^2$ ,  $\rho$ ,  $m^{(p+1)}(x)$  and  $\int (m^{(p+1)}(x))^2 f(x) dx$ , could be estimated as explained in Section 2.4.

#### 4. Study of the Real Private Residential Fixed Investment in the USA

In this Section, the study of some real economic data, using the nonparametric estimators of the regression function previously presented, is shown.

The sample data considered are 222 quarterly observations of the real private residential fixed investment in the USA, from 1947 to 2002. Each observation indicates the seasonally adjusted annual rate, measured in billions of chained 1996 dollars (in the year 2002, only two observations are available). The source of these data is the U.S. Department of Commerce, Bureau of Economic Analysis, obtained from the web page: <http://www.research.stlouisfed.org/fred/data/gdp.html>.

A fixed regression model can be fitted to these data, considering an equally spaced design on  $[0, 1]$ , that is,

$$Y_t = m(t/222) + \varepsilon_t, \quad t = 1, 2, \dots, 222.$$

The aim is to estimate the regression function using the LPR estimator or the EGLPR estimator with the global smoothing parameters proposed before. Only the linear case, that is,  $p = 1$ , will be considered. The reason for this choice is that in the estimation of the regression function odd order fits are preferable; while higher order polynomial approximations give a smaller order of the bias, the asymptotic variance is the same when moving from any even order approximation to its consecutive odd order approximation. However, the asymptotic variance increases when moving from an odd order approximation to its consecutive even order approximation. So, a local linear fit is preferable to a local constant fit, a local cubic fit is preferable to a local quadratic fit and so on. On the other hand, there is no preference between local linear and local cubic fits. To estimate the regression function, the order  $p$  is usually 1 or 3, nevertheless, a local linear fit suffices to produce good results and, moreover, it is computationally simpler (see Fan and Gijbels (1996) for a mathematical explanation). Depending on the assumption considered on the errors (whether the noise is white or not) and taking the background explained in the previous sections into account, if the errors are considered independent, the LPR estimator with the plug-in bandwidth obtained from (16), with  $c(\varepsilon) = \sigma_\varepsilon^2$  (this bandwidth is denoted by  $\hat{h}_{\text{ind}}$ ), should be used; or, if the errors are assumed to be correlated, the EGLPR estimator with the bandwidth (42) (for simplicity, this bandwidth is denoted by  $\hat{h}_{\text{dep}}$ ) is used. It should be noted that when the shape of both the mean and correlation functions are not specified at all, the separation of trend and noise remains a subjective matter, so that, both approximations could give a correct fit. The choice of estimator and bandwidth selection method therefore depends on the specific type of data under consideration. In this case, the kind of observations, gathered sequentially in time, suggests that, perhaps, the second approach is more appropriate. In any case, a more thorough study of the problem (including an analysis of the residuals) will give more light on the subject.

First, let us assume that the data are independent. For this case, Figure 1 shows the scatterplots of the data and the estimator of the regression function computed with the LPR estimator,  $\hat{m}_{\text{LPR}}$ . The plug-in bandwidth obtained from (16) is  $\hat{h}_{\text{ind}} = 0.0494$ . For its computation two unknown quantities,  $c(\varepsilon) = \sigma_\varepsilon^2$  and  $\int (m''(x))^2 f(x) dx$ , are estimated. The estimators used here are the same as those in Francisco-Fernández et al. (2002). So,  $\int (m''(x))^2 f(x) dx$  is estimated by  $1/n \sum_{i=1}^n (\hat{m}_g''(x_i))^2$ , where  $\hat{m}_g''(x)$  is the LPR estimator (using a cubic fit) of the second derivative obtained from (4) with a pilot bandwidth  $g = 0.25$ , empirically chosen and  $\sigma_\varepsilon^2$  is estimated by

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2, \quad (43)$$

where

$$\hat{\varepsilon}_i = Y_i - \hat{m}_{h_{\text{pilot}}}(x_i), \quad i = 1, 2, \dots, n. \quad (44)$$

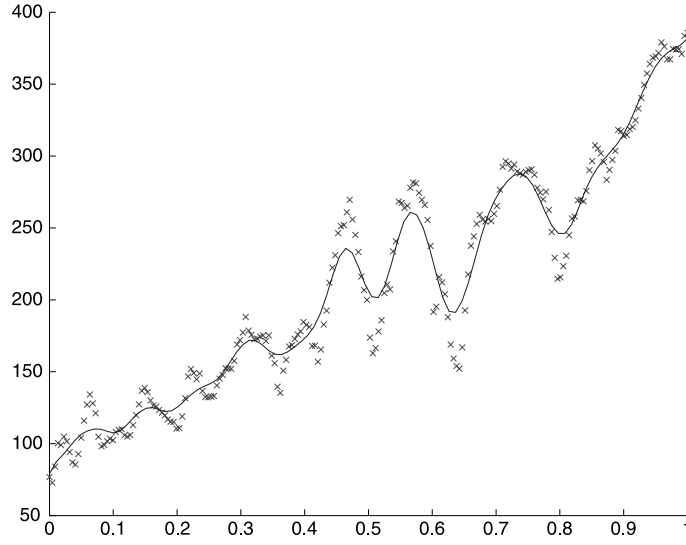


Figure 1. Sample data and  $\hat{m}_{\text{LPR}}$  computed with  $\hat{h} = 0.0494$ .

These nonparametric residuals are obtained using the LPR estimator with a pilot bandwidth computed using the cross-validation method ( $h_{\text{pilot}} = h_{\text{cv}}$ ), consisting in finding the bandwidth  $h$  that minimizes the score function (17). The results obtained are  $h_{\text{cv}} = 0.0061$  and  $\hat{\sigma}_{\varepsilon}^2 = 1.702$ .

On the other hand, assuming that the errors are correlated, based on the explanations given in Section 3, it is clear that the modification of the LPR estimator could give better results than those of the LPR estimator. So, under this assumption of dependence, the EGLPR estimator with the bandwidth obtained from (42) is used. To compute the EGLPR estimator the transformation matrix  $P_{(n)}$  must be estimated. A first approach could be to suppose that the errors follow an AR(1) process. It is important to note that, in many cases, although the errors do not follow an AR(1) model exactly, this assumption can eliminate a big part of the influence of the dependence of the observations in the nonparametric fit. In this case  $P_{(n)}$  is given by (37) and a natural estimator of  $P_{(n)}$  is obtained by replacing  $\rho$  with its estimator  $\hat{\rho}$ ,

$$\hat{\rho} = \frac{\sum_{t=1}^{n-1} \hat{\varepsilon}_t \hat{\varepsilon}_{t+1}}{\sum_{t=1}^n \hat{\varepsilon}_t^2}, \quad (45)$$

where  $\hat{\varepsilon}_t$ ,  $1 \leq t \leq n$ , are the residuals given by (44) with a pilot bandwidth computed by the TSCV method (since the data are considered dependent). This estimator of  $\rho$  and the estimator of  $\sigma_{\varepsilon}^2$  given in (43), but using a pilot bandwidth computed by the TSCV method, are used to estimate  $\sigma_{\varepsilon}^2((1 + \rho)/(1 - \rho))$  in the bandwidth (42). With the data considered in this study, the results are:  $h_{\text{TSCV}} = 0.1672$ ,  $\hat{\rho} = 0.9334$ ,  $\hat{\sigma}_{\varepsilon}^2 = 652.94$  and  $\hat{h}_{\text{dep}} = 0.3187$ . For the denominator of (42), the same estimator as in the independent case is used. Figure 2 includes

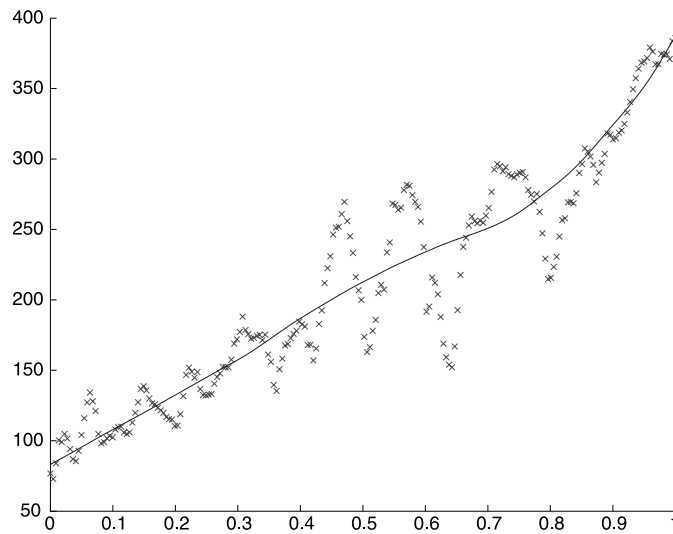


Figure 2. Sample data and  $\hat{m}_{\text{EGLPR}}$  computed with  $\hat{h} = 0.3187$ .

the sample data and the estimator of the regression function using the EGLPR estimator, assuming that the errors are AR(1),  $\hat{m}_{\text{EGLPR}}$ , following what has just been described.

If Figure 2, where the errors are assumed to have an AR(1) structure, is compared with Figure 1, where the data are assumed to be independent, in the independent case the estimator considers that the trend in the data is due mainly to the mean function, and therefore the estimated curve follows the sample data, producing a variable fit. On the other hand, when the errors are assumed to be AR(1), the estimator considers that the sample variability is due, in part, to the correlated errors and tries to eliminate this effect to uncover the true regression function.

The study of the estimated autocorrelation function (ACF) and the estimated partial autocorrelation function (PACF) of the residuals of this fit show that the assumption that the errors follow an AR(1) model is not correct and then, the estimation of the transformation matrix  $P_{(n)}$  under this assumption could not be good. In fact, the more reliable dependent structure for the errors seems to be an AR(2) model. Therefore, although the results obtained when the errors are assumed to follow an AR(1) model clearly improve those obtained under the assumption of independence, perhaps, better results could be obtained if a parametric structure for the errors is not assumed and the transformation matrix  $P_{(n)}$  is estimated directly from the data. Along these lines, Figure 3 shows the same plot as that in Figure 2, but in this case the EGLPR estimator is computed without assuming any parametric structure for the errors (this estimator is denoted by  $\hat{m}_{\text{EGLPR2}}$ ). In this case, first the

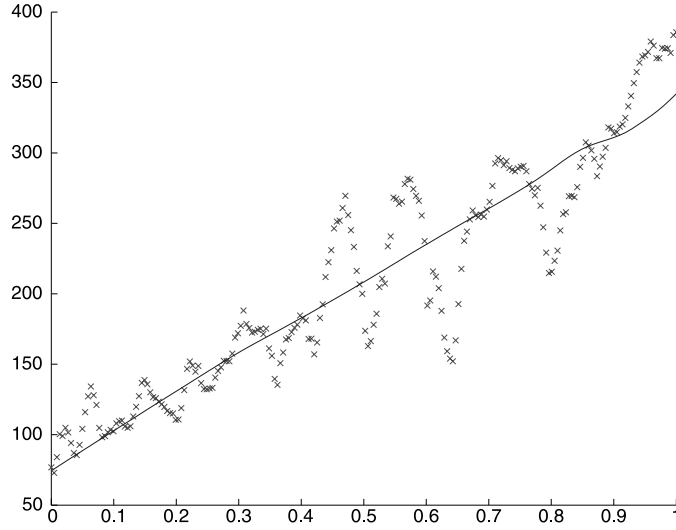


Figure 3. Sample data and  $\hat{m}_{\text{EGLPR2}}$  computed with  $\hat{h} = 0.3187$ .

correlation matrix of the errors,  $\hat{\Omega}_{(n)}$ , is estimated using estimators obtained by the method of the moments for their components:

$$\hat{\rho}(s) = \frac{1/n \sum_{i=1}^{n-s} \hat{\varepsilon}_i \hat{\varepsilon}_{i+s}}{1/n \sum_{i=1}^n \hat{\varepsilon}_i^2}, \quad s = 0, 1, 2, \dots, k, \quad (46)$$

as estimator of

$$\rho(s) = \frac{E(\varepsilon_k \varepsilon_{k+s})}{\sigma_\varepsilon^2}, \quad s = 0, 1, 2, \dots, k \quad (47)$$

and then, the matrix  $\hat{P}_{(n)}$ , such that,  $\hat{\Omega}_{(n)}^{-1} = \hat{P}_{(n)}' \hat{P}_{(n)}$ , is obtained numerically. The pilot bandwidth used to obtain the residuals in (46) is calculated by the TSCV method.

In the study of the residuals of this model, the ACF and the PACF of the residuals are computed and shown in Figures 4 and 5, respectively.

From Figures 4 and 5, it can be deduced that the errors fit an AR(2) model, more specifically, they follow the model:

$$Z_t - 1.4173Z_{t-1} + 0.4990Z_{t-2} = a_t.$$

Based on the estimator obtained in Figure 3 and the analysis of the residuals, it seems reasonable to assume that the series under study has a linear trend and a residual component with AR(2) structure. The hypothesis of that 'the trend is linear' can be tested using nonparametric methods. This problem has been studied by González-Manteiga and Vilar-Fernández (1995) and Vilar-Fernández and

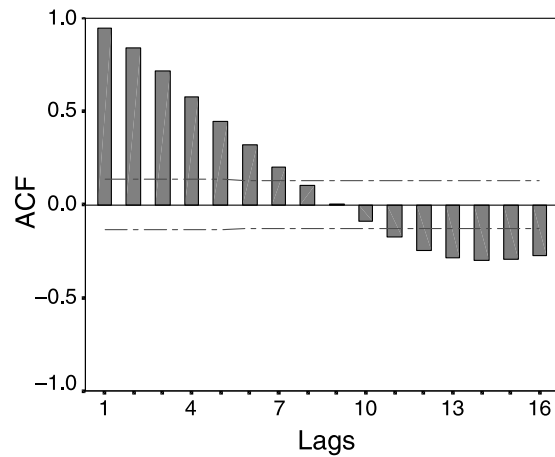


Figure 4. ACF of the residuals obtained using  $\hat{m}_{\text{EGLPR2}}$ .

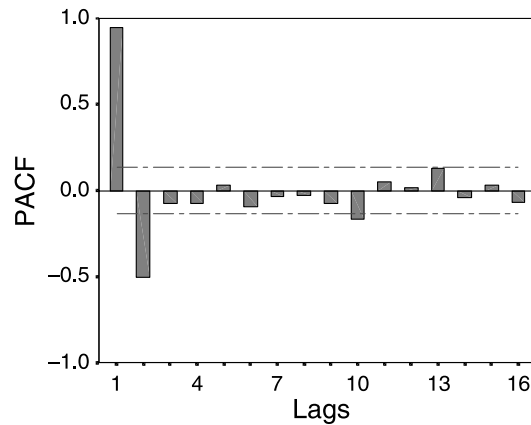


Figure 5. PACF of the residuals obtained using  $\hat{m}_{\text{EGLPR2}}$ .

González-Manteiga (1996). They propose to use as statistic a Cramer–von-Mises type functional distance between a nonparametric estimator of the regression function,  $\hat{m}_n$ , and the member of the class defined in the null hypothesis that is closest to  $\hat{m}_n$  in terms of this distance.

In Figure 6, the three nonparametric estimators,  $\hat{m}_{\text{LPR}}$ ,  $\hat{m}_{\text{EGLPR}}$  and  $\hat{m}_{\text{EGLPR2}}$  are represented.

Figure 6 shows that  $\hat{m}_{\text{EGLPR2}}$  is slightly less wiggly than the estimator obtained assuming that the errors are AR(1),  $\hat{m}_{\text{EGLPR}}$ . Therefore, at least with this time series, if a parametric shape for the errors is not assumed and the whole correlation matrix of the errors is estimated from the data, only small differences are observed with respect to the results obtained under the assumption of errors being AR(1). On the other hand, with this simple assumption, a big improvement is observed with

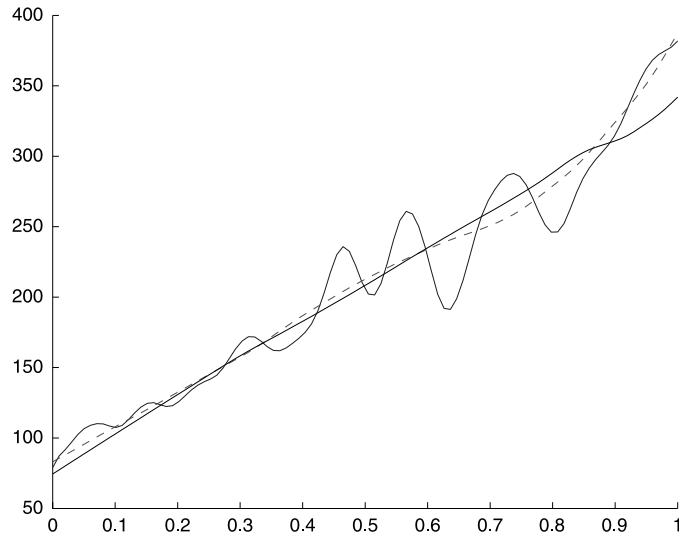


Figure 6.  $\hat{m}_{LPR}$  (wiggly solid line),  $\hat{m}_{EGLPR}$  (dashed line) and  $\hat{m}_{EGLPR2}$  (smooth solid line).

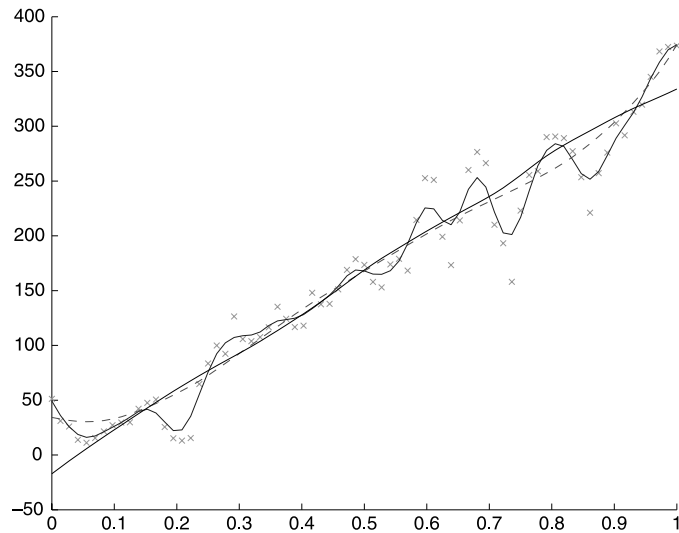


Figure 7. Sample data,  $\hat{m}_{LPR}$  (wiggly solid line),  $\hat{m}_{EGLPR}$  (dashed line) and  $\hat{m}_{EGLPR2}$  (smooth solid line).

respect to the hypothesis of independence and only one parameter, the correlation coefficient,  $\rho$ , has to be estimated to compute the EGLPR estimator.

Finally, the study of the real private residential fixed investment in the USA is completed considering the series observed annually, but for a longer period of time. In this case, the sample data consists of 73 annual observations of this index, from 1929 to 2001. The observations were obtained from the same source and each data point indicates the billions of chained 1996 dollars in each year.

Along the same lines as Figures 1–3, Figure 7 includes the sample data and three estimators of the trend, the LPR estimator assuming independent observations,  $\hat{m}_{\text{LPR}}$ , the EGLPR supposing that the errors follow an AR(1) process,  $\hat{m}_{\text{EGLPR}}$ , and the EGLPR estimator without assuming any parametric structure for the errors,  $\hat{m}_{\text{EGLPR}2}$ .

The shape of the three estimators computed, as well as the general conclusions, basically, are the same as those before cited for the quarterly data and they could be repeated here.

### Acknowledgements

The authors wish to thank a referee for the helpful comments and suggestions. This work was partially supported by the grants PB98-0182-C02-01, PGIDT01PXI10505PR, MCyT Grant BFM2002-00265 (European FEDER support included) and PGIDT03PXIC10505PN.

### References

- Anh, V., Wolff, R., Gao, J. and Tieng, Q.: Local linear kernel regression with long-range dependence errors, *Aust. New Zealand J. Stat.* **41**(4) (1999), 463–479.
- Beran, J. and Feng, Y.: Local polynomial fitting with long-memory, short memory and antipersistent errors, *Ann. Inst. Stat. Math.* **54**(2) (2002), 291–311.
- Bradley, R. C.: Basic properties of strong mixing conditions, In: E. Elberlein and M. S. Taquq (eds), *Dependence in Probability and Statistics*, Birkhäuser, 1986, pp. 165–192, Oberwolfack, 1985.
- Cleveland, W. S.: Robust locally weighted regression and smoothing scatterplots, *J. Am. Stat. Assoc.* **74** (1979), 829–836.
- Doukhan, P.: *Mixing. Properties and Examples*, Lectures Notes in Statistics 85, Springer-Verlag, Berlin, 1995.
- Eubank, R. L.: *Spline Smoothing and Nonparametric Regression*, Marcel Dekker, New York, 1988.
- Fan, J.: Design-adaptive nonparametric regression, *J. Am. Stat. Assoc.* **87** (1992), 998–1004.
- Fan, J.: Local linear regression smoothers and their minimax efficiency, *Ann. Stat.* **21** (1993), 196–216.
- Fan, J. and Gijbels, I.: Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation, *J. Royal Stat. Soc. Series B* **57**(2) (1995), 371–394.
- Fan, J. and Gijbels, I.: *Local Polynomial Modelling and Its Applications*, Chapman and Hall, London, 1996.
- Fraiman, R. and Meloche, J.: Smoothing dependent observations, *Stat. Prob. Lett.* **2** (1994), 203–214.
- Francisco-Fernández, M. and Vilar-Fernández, J. M.: Local polynomial regression estimation with correlated errors, *Commun. Stat. Theor. Methods* **30**(7) (2001), 1271–1293.
- Francisco-Fernández, M. and Vilar-Fernández, J. M.: Bandwidth selection for the local polynomial estimator under dependence: a simulation study, Technical Report (2002).
- Francisco-Fernández, M., Opsomer, J. D. and Vilar-Fernández, J. M.: A plug-in bandwidth selector for local polynomial regression estimator with correlated errors, Preprint No. 02-19, Department of Statistics, Iowa State University (2002).
- González-Manteiga, W. and Vilar-Fernández, J. M.: Testing linear regression models using non-parametric regression estimators when errors are non-independent, *Comput. Stat. Data Anal.* **20** (1995), 521–541.



- Hall, P. and Van Keilegom, I.: Using difference-based methods for inference in nonparametric regression with time-series errors, *J. Roy. Statist. Soc. Ser B* **65** (2003), 443–456.
- Härdle, W.: *Applied Nonparametric Regression*, Cambridge University Press, Cambridge, 1990.
- Härdle, W. and Tsybakov, A.: Local polynomial estimators of the volatility function in nonparametric autoregression, *J. Econometric.* **81** (1997), 223–242.
- Härdle, W., Tsybakov, A. and Yang, L.: Nonparametric vector autoregression, *J. Stat. Plan. Infer.* **68** (1998), 221–245.
- Hart, J.: Automated kernel smoothing of dependent data by using time series cross-validation, *J. Roy. Stat. Soc. Series B* **56**(3) (1994), 529–542.
- Hart, J. and Wehrly, T. E.: Kernel regression using repeated measurements data, *J. Am. Stat. Assoc.* **81** (1986), 1080–1088.
- Herrmann, E., Gasser, T. and Kneip, A.: Choice of bandwidth for kernel regression when residuals are correlated, *Biometrika* **79**(4) (1992), 783–795.
- Judge, G., Griffiths, W., Carter, R., Lütkepohl, H. and Lee, T. C.: *The Theory and Practice of Econometrics*, Wiley, New York, 1988.
- Masry, E.: Multivariate regression estimation: local polynomial fitting for time series, *Stoch. Process. Appl.* **65** (1996a), 81–101.
- Masry, E.: Multivariate local polynomial regression for time series: uniform strong consistency and rates, *J. Time Series Anal.* **17** (1996b), 571–599.
- Masry, E. and Fan, J.: Local polynomial estimation of regression function for mixing processes, *Scand. J. Stat.* **24** (1997), 165–179.
- Masry, E. and Mielniczuk, J.: Local linear regression for time series with long-range dependence, *Stoch. Process. Appl.* **82** (1999), 173–193.
- Müller, H. G.: *Nonparametric Regression Analysis of Longitudinal Data*, Lectures Notes in Statistics 46, Springer-Verlag, Berlin, 1988.
- Müller, H. G. and Stadtmüller, U.: Detecting dependencies in smooth regression models, *Biometrika* **75**(4) (1988), 639–650.
- Opsomer, J. D.: Nonparametric regression in the presence of correlated errors, In: T. G. Gregoire, D. R. Brillinger, P. J. Diggle, E. Russek-Cohen, W. G. Warren and R. D. Wolfinger (eds), *Modeling Longitudinal and Spatially Correlated Data: Methods, Applications and Future Directions*, Springer-Verlag, Berlin, 1997, pp. 339–348.
- Opsomer, J. D., Wang, Y. and Yang, Y.: Nonparametric regression with correlated errors, *Stat. Sci.* **16** (2001), 134–153.
- Ruppert, D. and Wand, M. P.: Multivariate locally weighted least squares regression, *Ann. Stat.* **22** (1994), 1346–1370.
- Ruppert, D., Sheather, S. J. and Wand, M. P.: An effective bandwidth selector for local least squares regression, *J. Am. Stat. Assoc.* **90** (1995), 1257–1270.
- Stone, C. J.: Consistent nonparametric regression, *Ann. Stat.* **5** (1977), 595–620.
- Tsybakov, A.: Robust reconstruction of functions by the local-approximation method, *Probl. Informat. Transmission* **22** (1986), 133–146.
- Vilar-Fernández, J. M. and Francisco-Fernández, M.: Local polynomial regression smoothers with AR-error structure, *TEST* **11**(2) (2002), 439–464.
- Vilar-Fernández, J. M. and González-Manteiga, G.: Bootstrap test of goodness of fit to a linear model when errors are correlated, *Commun. Stat. Theory Methods* **25**(12) (1996), 2925–2953.
- Vilar-Fernández, J. M. and Vilar-Fernández, J. A.: Recursive estimation of regression functions by local polynomial fitting, *Ann. Inst. Stat. Math.* **50**(4) (1998), 729–754.