Bandwidth selection for the local polynomial estimator under dependence: A simulation study

Article in Computational Statistics · January 2005

DOI: 10.1007/BF02741314

2 authors:

Mario Francisco-Fernández
University of A Coruña

Juan M Vilar
University of A Coruña

Bandwidth Selection for the Local Polynomial Estimator under Dependence: a Simulation Study

Mario Francisco-Fernández[1] and Juan M. Vilar-Fernández[1]

[1] Departamento de Matemáticas, Facultad de Informática, Universidad de A Coruña, Campus de Elviña, s/n, 15071, A Coruña, SPAIN

Summary

Seven of the most popular methods for bandwidth selection in regression estimation are compared by means of a thorough simulation study, when the local polynomial estimator is used and the observations are dependent. The study is completed with two plug-in bandwidths for the generalized local polynomial estimator proposed by Vilar-Fernández & Francisco-Fernández (2002).

Keywords: Local Polynomial Fitting, Dependent Data, Bandwidth Selection

1 Introduction

Consider the univariate fixed regression model given by

$$Y_t = m(x_t) + \varepsilon_t, \quad 1 \leq t \leq n, \tag{1}$$

where $m(x)$ is a smooth regression function, x_t are the points of the design and ε_t, the errors of the model, is a sequence of unobserved random variables with zero mean and finite variance σ^2.

We can estimate $m(x)$ from the sample data $\{(x_t, Y_t)\}_{t=1}^n$ using the nonparametric local polynomial regression (LPR) estimator. Some of the advantages of this nonparametric estimation method compared with other nonparametric regression estimators like Nadaraya-Watson or Gasser-Müller are better boundary behavior, straightforwardness to obtain good estimators of regression derivatives, easy computation, good minimax properties, etc. Some references on this method, when the observations are independent, are: Fan (1992), Ruppert & Wand (1994) and Ruppert et al. (1995). See Fan & Gijbels (1996) for additional references.

The LPR estimator is obtained by locally fitting a pth degree polynomial to the data by weighted least squares. The expression for the LPR estimator of $m(x)$ is given by the first component of the following vector:

$$\hat{\beta}_{(n)}(x) = \left(X_{p,(n)}^t W_{(n)} X_{p,(n)} \right)^{-1} X_{p,(n)}^t W_{(n)} \vec{Y}_{(n)}, \tag{2}$$

where

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

and $W_{(n)} = \operatorname{diag}(\omega_{n,1}, \ldots, \omega_{n,n})$ is a diagonal array of weights, where, for $t = 1, \ldots, n$, $\omega_{n,t} = n^{-1} h_n^{-1} K\left(h_n^{-1}(x_t - x)\right)$, K being a kernel function and h_n the bandwidth or smoothing parameter, which controls the amount of local averaging performed to obtain the regression estimate. In practice, it would be useful to have available an automated method based on sample observations for the selection of the smoothing parameter.

Sometimes, for example in the important case that the observations are gathered sequentially in time, the sample data cannot be assumed to be independent and some of the properties obtained for the LPR estimator in previous papers change in this new setting. Some papers studying the LPR estimator in a context of dependence are: Masry (1996), Masry & Fan (1997), Härdle & Tsybakov (1997), Härdle et al. (1998), Vilar-Fernández & Vilar-Fernández (1998) and Francisco-Fernández & Vilar-Fernández (2001). Opsomer et al. (2001) provide other references on this topic.

In the works cited the dependence of the observations and the regression model is treated in two ways: either the sample data $\{(X_t, Y_t)\}_{t=1}^{n}$ are a bivariate dependent random sample (mainly mixing) or a fixed regression model with the dependence coming through the errors of the model is present. In the first case, the dependence influences the variance of the LPR estimator in terms of second order, and then, under weak conditions on the mixing coefficients, the asymptotic behavior of the LPR estimator is the same as under independence (see Masry & Fan (1997)). In the case of fixed design, the influence of the dependence is stronger, because this influence is on the leading term of the variance (see Francisco-Fernández & Vilar-Fernández (2001)). So, under dependence, classical automated bandwidth selection methods for independent nonparametric regression break down (see, for example, Hart & Wehrly (1986)).

When the errors of the regression model are dependent and follow an $ARMA$ process, Vilar-Fernández & Francisco-Fernández (2002) propose a modification of the LPR estimator to estimate $m(x)$ 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. The GLPR estimator of $m(x)$ is the first component of the vector

$$\hat{\beta}_G(x) = \left(X_{p,(n)}^t P_{(n)}^t W_{(n)} P_{(n)} X_{p,(n)} \right)^{-1} X_{p,(n)}^t P_{(n)}^t W_{(n)} P_{(n)} \vec{Y}_{(n)}, \qquad (3)$$

where $P_{(n)}$ is a transforming matrix, such that $P_{(n)}^t P_{(n)} = \Omega_{(n)}^{-1}$, where $\Omega_{(n)}$ is the correlation matrix of the errors. In a practical situation, $P_{(n)}$ has to be estimated from the data. Again, a bandwidth must also be chosen.

The purpose of this paper is to review and compare, through a simulation study, the most important selectors for the smoothing parameter in the LPR estimator (2) and in the GLPR estimator (3), in a fixed regression model with dependent errors.

Properties of interest of all the procedures to be analyzed, such as consistency or rates of convergence to the optimal bandwidth, have been proven in the papers that will be cited later. On the other hand, when these methods of bandwidth selection are used, there will be the added problem of choosing one or more new parameters which influence the final smoothing parameter selected by each method. In any case, the selection of these new parameters is expected to be less important.

The organization of the paper is as follows: in Section 2 the bandwidth selection methods to be compared are briefly presented. The comparative simulation study is carried out in Section 3. Finally, in Section 4 the results are analyzed and some conclusions are drawn.

2 Data-driven bandwidth selectors

In what follows, let us consider the fixed regression model (1), with stationary errors and supposing that $\sum_{k=1}^{\infty} c(k) < \infty$, where $c(k) = Cov(\varepsilon_i, \varepsilon_{i+k})$. This assumption is satisfied by $ARMA$ type errors, among others.

2.1 Bandwidths for the LPR estimator

Seven selectors are compared: four selectors of the cross-validation type (modified cross-validation, partitioned cross-validation, time series cross-validation and cross-validation for prediction), two plug-in bandwidths (asymptotic plug-in and exact plug-in) and a smoothing parameter choice by bootstrap.

2.1.1 Modified cross-validation ($\hat{h}_{CVM}$)

This technique is a modification of the classical cross-validation method and has been studied in, among others, Hart & Vieu (1990), Chu & Marron (1991) and Härdle & Vieu (1992). It consists in choosing the bandwidth that minimizes the cross-validation function

$$CV(h) = \frac{1}{n} \sum_{t=1}^{n} \left(\hat{m}_{(t)}(x_t) - Y_t \right)^2 w(x_t), \tag{4}$$

$w(\cdot)$ being a weight function included to avoid the boundary effect and $\hat{m}_{(t)}(x_t)$ the estimation of $m(x_t)$ with the LPR estimator, using the sample data $\{(x_i, Y_i)\}_{i=1}^{n}$without the consecutive observations $(x_u, Y_u), \ldots, (x_v, Y_v)$, with $u = \max\{1, t-l_n\}$ and $v = \min\{n, t+l_n\}$. So, up to $2l_n+1$ observations are removed from the original sample data, where l_n is a parameter to choose, such that, Y_t and Y_{t+l_n+1} are practically independent. In practice, l_n can be chosen empirically, as the smallest value such that $\gamma(|l_n|) \simeq 0$, where $\gamma(k)$ is the k order autocovariance. If $l_n = 0$, we have the classical cross-validation rule when independence is assumed.

2.1.2 Partitioned cross-validation ($\hat{h}_{PCV}$)

Introduced by Marron (1987) for bandwidth selection in density estimation, it has been compared with the modified cross-validation in the fixed design regression in Chu & Marron (1991).

In this method the n sample observations are divided into g ($g \geq 1$) subgroups, and the least-squares cross-validation function (4), with $l_n = 0$, $CV_k(h)$ is calculated for each subgroup ($k = 1, 2, \ldots, g$). $CV_k(h)$ is given by

(4) but using the subsample $\left\{\left(x_{(i-1)\times g+k}, Y_{(i-1)\times g+k}\right)\right\}_{i=1}^{[n/g]}$, $k = 1, 2, \ldots, g$, where $[a]$ denotes the integer part of a. After this, the average of these functions is minimized:

$$CV^*(h) = \frac{1}{g}\sum_{k=1}^{g} CV_k(h). \tag{5}$$

Let $\hat{h}_{CV}^*$ be the value of h that minimizes (5). Chu & Marron (1991) proved that $\hat{h}_{CV}^*$ is of the order $(n/g)^{-1/5}$ and since the optimal bandwidth is of the order $n^{-1/5}$, the partitioned cross-validation bandwidth is defined as $\hat{h}_{PCV} = g^{-1/5}\hat{h}_{CV}^*$.

In this technique, the parameter g that plays a similar role to that of the integer l_n in the modified cross-validation must be chosen. Obviously, if $g = 1$, the partitioned cross-validation would be the ordinary cross-validation.

Some asymptotic properties (central limit theorems and rates of convergence) are obtained in Chu & Marron (1991) for $\hat{h}_{CVM}$ and $\hat{h}_{PCV}$. They used the Nadaraya–Watson estimator and assumed that the errors had an $ARMA$ dependence structure. In conclusion, the rate of convergence of $\hat{h}_{PCV}$ is faster than that of $\hat{h}_{CVM}$, and in both cases it is equal to that obtained when assuming independence ($n^{-1/10}$ and $g^{-2/5}n^{-1/10}$, respectively).

2.1.3 Time series cross-validation ($\hat{h}_{TSCV}$)

This method was proposed in Hart (1994). In that paper the author explains why the ordinary cross-validation method yields a poor bandwidth when the data are correlated. When the data are independent,

$$E\left(Y_j | Y_1, \ldots, Y_{j-1}, Y_{j+1}, \ldots, Y_n\right) = m(x_j), \tag{6}$$

that is, a good predictor of Y_j and a good estimator of $E(Y_j)$ are one and the same, but if the error terms in the model are correlated, then

$$E\left(Y_j | Y_1, \ldots, Y_{j-1}, Y_{j+1}, \ldots, Y_n\right) = m(x_j) + g(\varepsilon_1 \ldots \varepsilon_{j-1}\varepsilon_{j+1} \ldots \varepsilon_n) \tag{7}$$

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

To solve this problem, the TSCV method uses expression (7) and selects a model that yields good predictions of future observations. The idea is similar to that of the cross-validation method, but bases a prediction of Y_t only on the previous data $Y_1, \ldots, Y_{j-1}$, rather than on the data $Y_1, \ldots, Y_{j-1}, Y_{j+1}, \ldots, Y_n$.

Suppose that in our trend estimation problem we assume a parametric model M_θ for the process $\{\varepsilon_j = Y_j - m(x_j)\}$. Now using (7), consider a predictor of Y_j of the form

$$\hat{Y}_j(h,\theta) = \tilde{m}_{jh}(x_j) + g_\theta(\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_{j-1}), \tag{8}$$

where $\tilde{m}_{jh}(x_k)$ is a kernel estimate of $m(x_k)$ that uses the data $Y_1, \ldots, Y_{j-1}$, $\hat{\varepsilon}_i = Y_i - \tilde{m}_{jh}(x_i)$, $i = 1, \ldots, j-1$ and $g_\theta(\varepsilon_1, \ldots, \varepsilon_{j-1})$ is the optimal mean squared error predictor of ε_j, given $\varepsilon_1, \ldots, \varepsilon_{j-1}$ and assuming that the probability model is M_θ. The TSCV method consists in choosing h and θ to minimize

$$P(h,\theta) = \sum_{j=2}^{n} \left(\hat{Y}_j(h,\theta) - Y_j \right)^2 . \tag{9}$$

If, for example, the errors have an autoregressive dependence structure, it is possible, first, to estimate, for each h, the value $\hat{\theta}(h)$ that minimizes $P(h,\theta)$ with respect to θ. So, the problem now is to find the bandwidth h that minimizes $P(h) = P\left(h, \hat{\theta}(h)\right)$.

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 one is interested in efficiently estimating the trend, $m(x_j)$, using data on either side of time point j with a kernel K, we will have to multiply the bandwidth, that minimizes (9), by a constant $C_{K,L}$, dependent on the kernels K and L (see Hart (1994) for an explicit expression of this constant).

2.1.4 Cross-validation for prediction ($\hat{h}_{CVP}$)

This technique was proposed in Yao & Tong (1998) and uses a cross-validation algorithm but with some modifications to deal with the bad performance of this method when the observations are dependent. In this paper, the authors apply the new smoothing parameter to the LPR estimator, assuming that the data have a $\rho-$mixing kind of dependence. The steps to follow to obtain the sought bandwidth are: First, the bandwidth $\hat{h}_b$ minimizing the following cross-validation function

$$ECV_b(h) = \frac{1}{n-b} \sum_{t=b+1}^{n} \left(Y_t - \hat{m}_{b,h}(x_t)\right)^2 w(x_t), \tag{10}$$

must be found, where $\hat{m}_{b,h}(x_t)$ is the kernel estimator of the regression function computed from the subsample $\{(x_i, Y_i)\}_{i=1}^{b}$. The parameter b represents the number of observations in the sample data used to predict the $n-b$ remaining observations. That is, the function $ECV_b(h)$ gives the error of this prediction. Finally, the bandwidth is given by $\hat{h}_{CVP} = \hat{h}_b \left(\frac{b}{n}\right)^{1/5}$.

Simulation studies by Yao and Tong show that the new bandwidth has, on average, good behavior, but its variability is high. Moreover, if the integrated squared error is used as a comparison criterion, the new bandwidth is worse than the modified cross-validation selector, but its performance is better when the predictions are used. Finally, better behavior for this selector is observed when the data are highly negatively correlated.

2.1.5 Asymptotic plug-in ($\hat{h}_{PLIA}$)

The plug-in methods consist in obtaining an estimator of the bandwidth that minimizes the mean squared error (local bandwidth) or the mean integrated squared error (global bandwidth), replacing the unknown quantities by estimators.

For the LPR estimator under suitable conditions, it follows from the expressions of the bias and the variance given in Corollary 1 of Francisco-Fernández & Vilar-Fernández (2001) that for p odd, the asymptotically optimal bandwidth to estimate the regression function is given by

$$h_{PLIA} = C_{0,p}\left(K\right)\left(\frac{c(\varepsilon)}{n\int\left(m^{(p+1)}(x)\right)^2 f(x)dx}\right)^{1/(2p+3)}, \qquad (11)$$

where $C_{0,p}\left(K\right)$ is a real number that depends only on kernel K and $c(\varepsilon)$ is the sum of the covariances of the errors ($c(\varepsilon) = \sigma_\varepsilon^2\left(c(0) + 2\sum_{k=1}^{\infty} c\left(k\right)\right)$).

In (11) there are two unknown quantities: $c(\varepsilon)$ and $\int\left(m^{(p+1)}(x)\right)^2 f(x)dx$. These must be estimated to obtain a practical smoothing parameter from this bandwidth. In Francisco-Fernández & Vilar-Fernández (2001), some suggestions on how to do this are presented. Once these unknown quantities are estimated and these estimators plugged into (11), the asymptotically optimal plug-in global bandwidth to estimate $m(x)$ with a p-degree polynomial is denoted by $\hat{h}_{PLIA}$.

In the simulation study presented here, the estimators of the unknown quantities $c(\varepsilon)$ and $\int\left(m^{(p+1)}(x)\right)^2 f(x)dx$ are the same as those used in Francisco-Fernández et al. (2004), where the rate of convergence of this bandwidth to the optimal one is shown. In that paper, the authors propose to estimate $c(\varepsilon)$ assuming a parametric $ARMA$ structure for the errors, depending on a small number of parameters. Then, an estimator of $c(\varepsilon)$ can be obtained by simply estimating these parameters from nonparametric residuals (using a pilot bandwidth computed, in practice, by the TSCV method). Recently, Hall & Van Keilegom (2003) have proposed a differencing-based approach to estimating $c(\varepsilon)$. For the other parameter, given by $\int\left(m''(x)\right)^2 f(x)dx$ when a local linear fitting, $p = 1$, is used, a local polynomial fitting, using a pilot bandwidth, will be used. See Francisco-Fernández & Vilar-Fernández (2001) for other ways to estimate these unknown parameters.

2.1.6 Exact plug-in ($\hat{h}_{PLIE}$)

This selector was proposed in Francisco-Fernández & Vilar-Fernández (2001) in the fixed regression model with correlated errors, following analogous ideas to those given in Fan et al. (1996) for the case of independent observations. As in that case, the exact expressions for the bias and the variance of the LPR estimator are derived immediately from its definition in (2):

$$Bias\left(\hat{\beta}_{(n)}(x)\right) = S_{(n)}^{-1} X_{p,(n)}^{t} W_{(n)} \vec{R}_{(n)}. \tag{12}$$

$$Var\left(\hat{\beta}_{(n)}(x)\right) = S_{(n)}^{-1} X_{p,(n)}^{t} W_{(n)} \Gamma_{(n)} W_{(n)} X_{p,(n)}^{t} S_{(n)}^{-1}, \tag{13}$$

where $S_{(n)} = X_{p,(n)}^{t} W_{(n)} X_{p,(n)}$, $\vec{R}_{(n)} = \vec{M}_{(n)} - X_{(n)} \vec{\beta}(x)$ (with $\vec{\beta}(x) = \left(m(x), m'(x), \ldots, m^{(p)}(x)/p!\right)^{t}$) and $\Gamma_{(n)} = E\left(\vec{\varepsilon}\vec{\varepsilon}^{t}\right)$ is the variance-covariance matrix of the residuals. The arrays $\vec{R}_{(n)}$ and $\Gamma_{(n)}$ are unknown. An estimator of the mean integrated squared error of the regression function estimator can be obtained by substituting these unknown quantities by estimators. Now, minimizing this estimator of the mean integrated squared error, the exact plug-in bandwidth ($\hat{h}_{PLIE}$) is obtained. Basically, expression (12) is estimated using a proper LPR estimator with a pilot bandwidth and $\Gamma_{(n)}$ in (13) is estimated in the same way as $c(\varepsilon)$ in the previous selector, assuming a parametric dependence structure for the errors.

2.1.7 Bootstrap method ($\hat{h}_{BOOT}$)

Classical bootstrap techniques to select the smoothing parameter designed for independent observations can produce wrong bandwidths under dependence, so these methods have been modified to take into account the dependence in the resampling step and to obtain more reliable bandwidths. Some of these methods are the block bootstrap or the stationary bootstrap. On the other hand, if an explicit dependence structure for the errors (say an AR(p), for example) can be assumed, this information can be used to adapt the ordinary bootstrap to this model. In Cao (1999) and Härdle et al. (2003) an overview of the existing literature about bootstrapping in time series is given.

The bootstrap method used in the simulation study presented here is based on the same lines as those studied in Hall et al. (1995). In that paper, the authors analyze methods based on the block bootstrap and modified cross-validation, for bandwidth selection in nonparametric regression when errors have an almost arbitrarily long-range dependence. They show that, provided block length or leave-out number are chosen appropriately, both techniques produce first-order optimal bandwidths. Nevertheless, the block bootstrap has far better empirical properties, particularly under long-range dependence.

The global bootstrap bandwidth used here is obtained as follows: Let $\hat{m}_1$ and $\hat{m}_2$ denote two estimators of m constructed according to (2), but employing respective values h_1 and h_2 of the bandwidth h. In what follows $\hat{m}_1$ will be used to compute residuals and $\hat{m}_2$ to generate bootstrap data.

For each $i = 1, 2, \ldots, n$, $\hat{\varepsilon}_{1i} = Y_i - \hat{m}_1(x_i)$, $\bar{\hat{\varepsilon}}_{1\cdot} = \frac{1}{n}\sum \hat{\varepsilon}_{1i}$ and $\hat{\varepsilon}_i = \hat{\varepsilon}_{1i} - \bar{\hat{\varepsilon}}_{1\cdot}$ These are the centered residuals. Let ε_i^* be the bootstrap errors (we have assumed a parametric explicit structure to define them, instead of block bootstrap as in Hall et al. (1995)). In terms of these errors, let $Y_i^* = \hat{m}_2(x_i) + \varepsilon_i^*$, $i = 1, 2, \ldots, n$. Now, a bootstrap sample, $(x_1, Y_1^*), (x_2, Y_2^*), \ldots, (x_n, Y_n^*)$, is obtained, and from it, a bootstrap estimator, $\hat{m}^*(\cdot)$, using the LPR estimator. Repeating this B times, an estimator for the mean integrated squared error is obtained, given by:

$$\widehat{MISE}(h) = \frac{1}{n}\sum E\left(\hat{m}^*(x) - \hat{m}_2(x)\right)^2. \tag{14}$$

The value of h that minimizes (14) is the bandwidth, $\hat{h}_{BOOT}$, used here. h_1 and h_2 are selected as in Hall et al. (1995).

To obtain the bootstrap errors assuming, for example, an AR(1) structure for the errors of the model, the following was done: once the residuals $\{\hat{\varepsilon}_{11}, \hat{\varepsilon}_{12}, \ldots, \hat{\varepsilon}_{1n}\}$ have been obtained, the correlation coefficient, ρ, is estimated from them, using, for example, the estimator obtained from the method of the moments. Estimators of the white noise process are computed by: $\hat{e}_{1t} = \hat{\varepsilon}_{1t} - \hat{\rho}_{(n)}\hat{\varepsilon}_{1t-1}$, $t = 2, \ldots, n$. Next, they are centered, $\tilde{e}_t = \hat{e}_{1t} - \bar{\hat{e}}_{1\cdot}$, $t = 2, \ldots, n$, where $\bar{\hat{e}}_{1\cdot} = \frac{1}{n-1}\sum_{i=2}^{n}\hat{e}_{1i}$. A sample, $e_{-M}^*, \ldots, e_0^*, \ldots, e_n^*$, from $\{\tilde{e}_2, \ldots, \tilde{e}_n\}$ of sample size $M + n$ (M being a large enough integer, for example, $M = 50$), is selected. Finally, the bootstrap residuals are obtained by $\varepsilon_t^* = \hat{\rho}_{(n)}\varepsilon_{t-1}^* + e_t^*$, $t = 1, 2, \ldots, n$, where $\varepsilon_0^* = \sum_{j=0}^{M}\hat{\rho}_{(n)}^j e_{-j}^*$.

2.2 Bandwidths for the GLPR estimator

Two plug-in bandwidths (asymptotic plug-in and exact plug-in) for this estimator are considered.

2.2.1 Asymptotic plug-in ($\hat{h}_{PLIAG}$)

In Vilar-Fernández & Francisco-Fernández (2002) the asymptotic properties of the estimator (3) are presented and studied. The results obtained in this paper show that if we have a fixed regression model with dependent errors following an autoregressive model, the new estimator and the classical LPR estimator have the same asymptotic expression for the mean integrated squared error. Taking this into account, an asymptotic plug-in bandwidth using the formula (11) can be selected. The same discussion on how to select the unknown quantities in (11) applies here. In practice, we have used the same assumptions for the errors and the same estimators as those used in the asymptotic plug-in bandwidth for the LPR estimator.

2.2.2 Exact plug-in ($\hat{h}_{PLIEG}$)

Following the lines exposed for the exact plug-in bandwidth for the LPR estimator, $\hat{h}_{PLIE}$, Vilar-Fernández & Francisco-Fernández (2002) have proposed an exact plug-in smoothing parameter for the new nonparametric estimator of the regression function and its derivatives obtained from (3). Basically, the method consists in using the exact expressions of the bias and the variance of (3) and obtaining an estimator of the mean integrated squared error substituting the unknown quantities appearing in the exact bias and variance expressions. Then, the bandwidth $\hat{h}_{PLIEG}$ is computed by simply minimizing this estimator of the mean integrated squared error.

The same kind of estimators for the unknown elements of the bias and the variance as in the smoothing parameter $\hat{h}_{PLIE}$ are used in this case: a suitable nonparametric fit using the LPR estimator for the bias and estimators for the parameters of the errors, assuming a parametric dependence structure for these errors. These estimators, obtained with the method of the moments, are constructed from the residuals computed from a nonparametric fit. Two pilot bandwidths are necessary for this process. In practice, as in $\hat{h}_{PLIA}$, a pilot bandwidth computed by the TSCV method was used to obtain the nonparametric residuals.

3 Simulation study

A simulation study was carried out to compare the bandwidths presented in the previous section. Different regression models have been considered, but in all of them an equally spaced design on the unit interval was taken. In the first part of the study AR(1) errors having a normal ($N(0, \sigma_\varepsilon^2)$) distribution were generated. Different values of correlation coefficient to study the influence of the dependence of the observations were considered.

In a first step, 300 samples of size n, following the model (1), were simulated. Using these samples, we computed the measurement of discrepancy

$$MISE(h) = E \int \left(\hat{m}_h(t) - m(t)\right)^2 dt. \qquad (15)$$

By minimizing (15) in h, some numerical approximation to the optimal values h_{MISE} was found. The second step consists in drawing another 300 random samples of sample size n of the regression model and computing the bandwidths to be compared for every sample. Using Monte Carlo approximations, the expected value, the standard deviation for every selector and the efficiency measure (16) can be approximated.

$$\Delta MISE = E\left(MISE\left(\hat{h}\right) - \min_{h>0} MISE(h)\right)^2. \tag{16}$$

As observed in Section 2, to calculate most of the selectors compared here, it is necessary to set an additional parameter. We have chosen several representative values of this parameter for each of the bandwidth selectors and the best result was taken for each method.

- For $\hat{h}_{CVM}$: let $2l_n + 1$ be the number of data observations removed from the sample to obtain the estimator of the regression function used in the cross-validation function. $l_n = 0$, 2, 5, 7 and 10 were considered.
- For $\hat{h}_{PCV}$: let g be the number of subgroups considered from the sample. $g = 1$, 3, 5, 7 and 10 were considered.
- For $\hat{h}_{CVP}$: let b be the number of data observations used to predict the remaining $n - b$ observations. $b = 30$, 60, 70 and 80 were considered.
- For $\hat{h}_{PLIA}$, $\hat{h}_{PLIE}$, $\hat{h}_{PLIAG}$ and $\hat{h}_{PLIEG}$: let h_{pilot} be the pilot bandwidth used to estimate the second derivative of the regression function. $h_{pilot} = 0.3$, 0.4, 0.5 and 0.6 were considered.
- For $\hat{h}_{BOOT}$: let B be the number of bootstrap replicas assuming that the errors are AR(1). $B = 150$ was considered.

Moreover, the weight function used in $\hat{h}_{CVM}$, $\hat{h}_{PCV}$ and $\hat{h}_{CVP}$ was taken as $w(t) \equiv 1$. For brevity, we present here only some of the results obtained, the rest being available from the authors upon request.

Model (1) $m(x) = \sin(\pi x)$, $n = 100$, $\sigma_\varepsilon = 0.3$ and $\rho = -0.9, -0.6, -0.3, 0, 0.3, 0.6, 0.9$.

In Table 1 the optimal bandwidth and the means of the bandwidths studied for the LPR estimator with their standard deviations as a function of ρ are presented.

We have grouped the bandwidths in two groups. In Figure 1, h_{MISE}, as well as $\hat{h}_{PLIA}$, $\hat{h}_{PLIE}$ and $\hat{h}_{BOOT}$, as a function of ρ, are shown.

Table 1: h_{MISE}, $\hat{h}_{CVM}$,$\hat{h}_{PCV}$, $\hat{h}_{TSCV}$, $\hat{h}_{CVP}$, $\hat{h}_{PLIA}$, $\hat{h}_{PLID}$ and $\hat{h}_{BOOT}$ as a function of ρ in Model (1)

h $\times 10^{-1}$	-0.9	-0.6	-0.3	$\rho = 0$	0.3	0.6	0.9
h_{MISE}	1.444	1.848	2.131	2.374	2.697	2.939	3.586
$\hat{h}_{CVM}$	1.663	2.534	2.645	2.449	3.171	3.348	3.619
St. dev.	0.0335	0.0399	0.0380	0.0688	0.0785	0.0968	0.1017
$\hat{h}_{PCV}$	2.306	2.833	2.645	2.870	2.954	2.896	2.294
St. dev.	0.0714	0.0239	0.0380	0.0430	0.0360	0.0415	0.0732
$\hat{h}_{TSCV}$	1.126	1.633	1.930	2.181	2.448	2.616	2.455
St. dev.	0.0189	0.0196	0.0234	0.0286	0.3191	0.0307	0.0568
$\hat{h}_{CVP}$	3.531	3.693	3.905	4.233	4.393	4.263	4.587
St. dev.	0.0620	0.0776	0.0928	0.1005	0.1086	0.1185	0.0910
$\hat{h}_{PLIA}$	1.255	1.650	1.896	2.111	2.319	2.521	2.403
St. dev.	0.0021	0.0049	0.0079	0.0113	0.0151	0.0233	0.0312
$\hat{h}_{PLIE}$	1.475	1.807	2.024	2.201	2.360	2.457	2.165
St. dev.	0.0046	0.0083	0.0116	0.0156	0.0199	0.0270	0.0331
$\hat{h}_{BOOT}$	1.452	1.854	2.099	2.101	2.159	2.360	2.357
St. dev.	0.0098	0.0124	0.0256	0.0398	0.0455	0.0523	0.0671

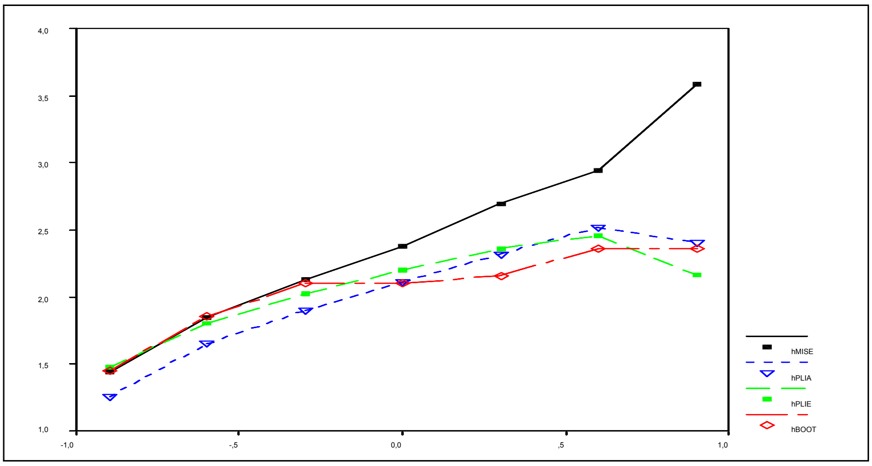

Figure 1: h_{MISE}, $\hat{h}_{PLIA}$, $\hat{h}_{PLIE}$ and $\hat{h}_{BOOT}$ as a function of ρ in Model (1)

Figure 2 includes the evolution, as a function of ρ, of the bandwidths obtained using a cross-validation technique ($\hat{h}_{CVM}$, $\hat{h}_{PCV}$, $\hat{h}_{TSCV}$ and $\hat{h}_{CVP}$) and the optimal bandwidth h_{MISE}.

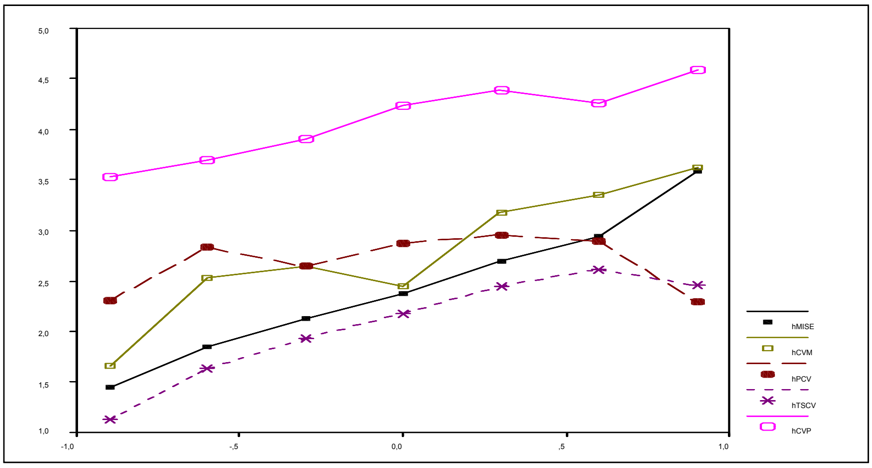

Figure 2: h_{MISE}, $\hat{h}_{CVM}$, $\hat{h}_{PCV}$, $\hat{h}_{CVP}$ and $\hat{h}_{TSCV}$ as a function of ρ in Model (1)

In Figure 3, $\ln(\Delta MISE)$ for $\hat{h}_{PLIA}$, $\hat{h}_{PLIE}$ and $\hat{h}_{BOOT}$, as a function of ρ, is presented. Figure 4 contains the same information as Figure 3 for $\hat{h}_{CVM}$, $\hat{h}_{PCV}$, $\hat{h}_{TSCV}$ and $\hat{h}_{CVP}$.

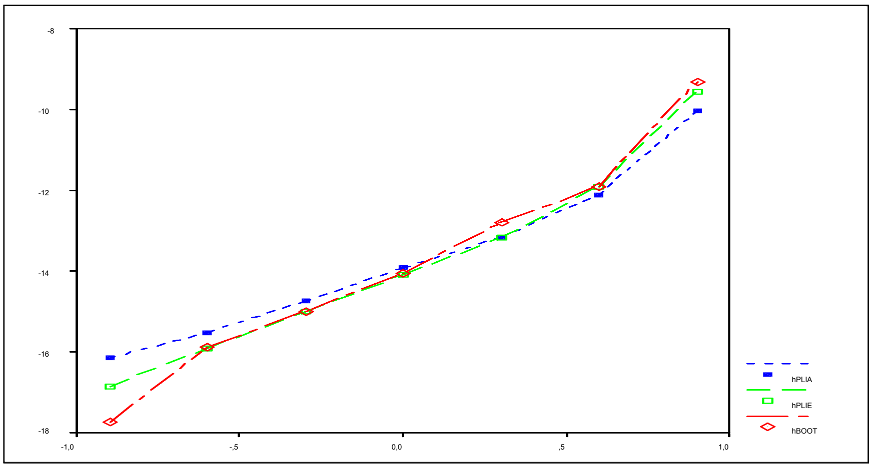

Figure 3: $\ln(\Delta MISE)$ as a function of ρ in Model (1)

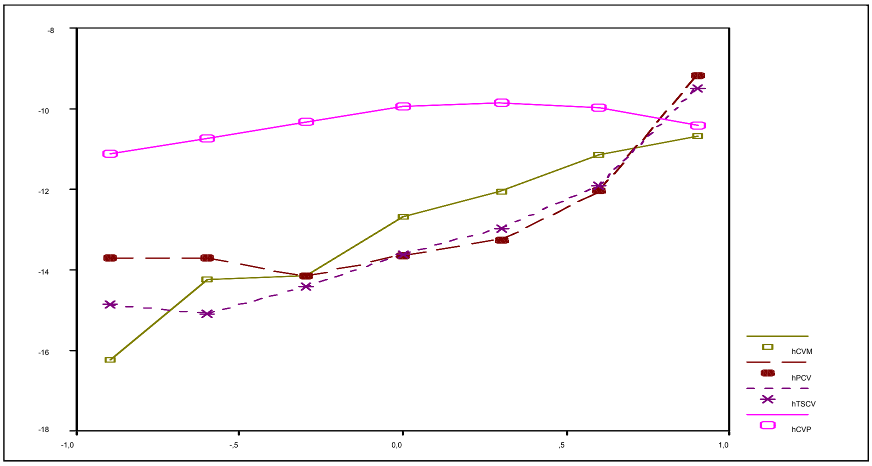

Figure 4: $\ln(\Delta MISE)$ as a function of ρ in Model (1)

This simulation study was repeated for the GLPR estimator. In Figure 5, we compare by means of the $\ln(\Delta MISE)$ the results obtained for $\hat{h}_{PLIA}$ using the LPR estimator with the results for $\hat{h}_{PLIAG}$ and $\hat{h}_{PLIEG}$ using the GLPR estimator.

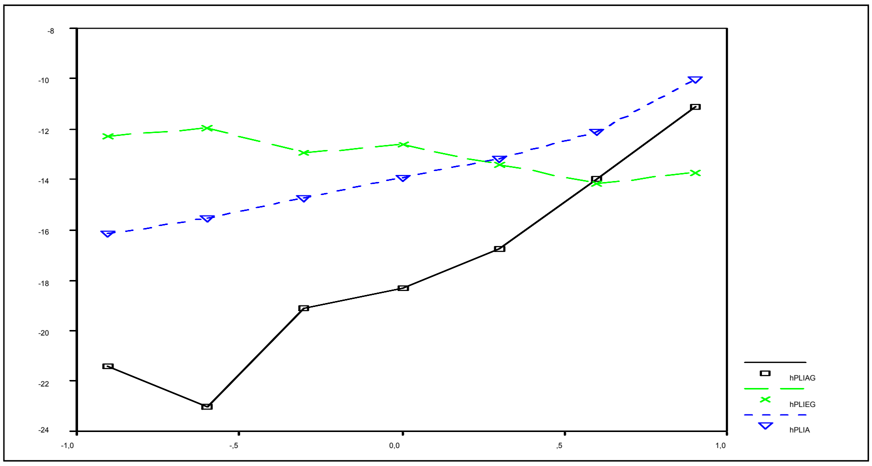

Figure 5: $\ln(\Delta MISE)$ as a function of ρ in Model (1)

Model (2) $m(x) = \sin(5\pi x)$, $n = 200$, $\sigma_\varepsilon = 0.3$ and $\rho = -0.9, -0.6, -0.3, 0, 0.3, 0.6, 0.9$.

In this model the regression function is much more variable than in Model 1. In Figure 6 the seven bandwidths computed and h_{MISE} are represented as a function of ρ. Figure 7 includes $\ln(\Delta MISE)$ as a function of ρ.

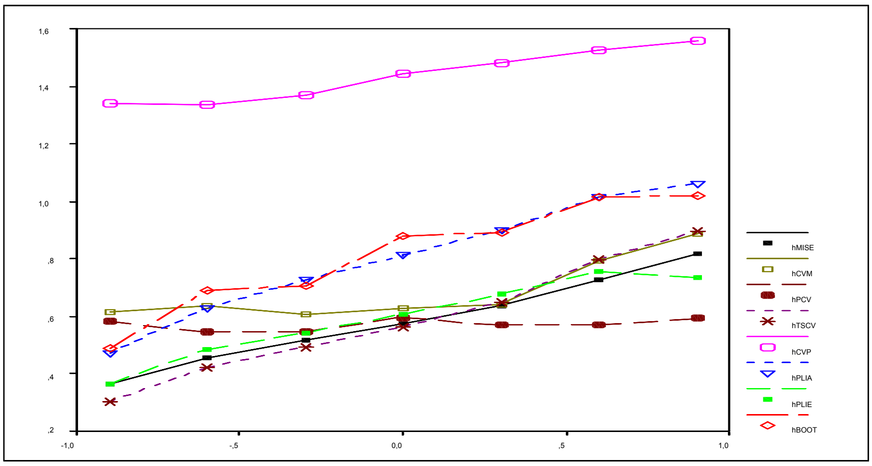

Figure 6: Computed bandwidths and the $MISE$ optimal bandwidth as a function of ρ in Model (2)

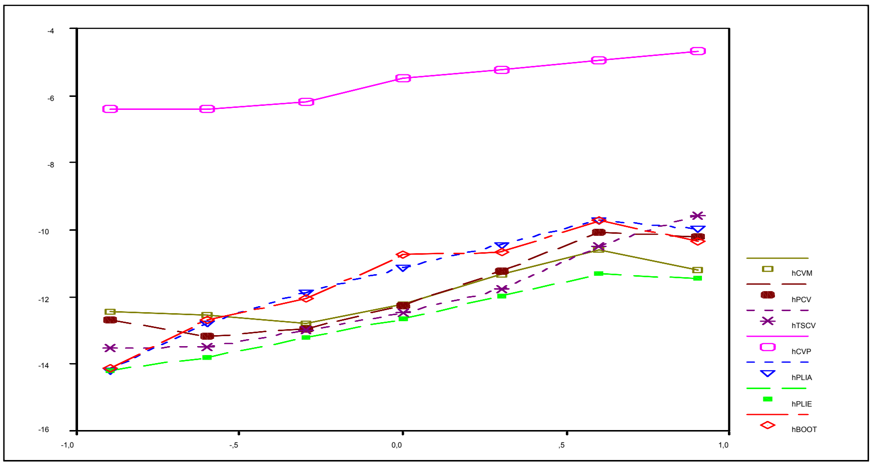

Figure 7: $\ln(\Delta MISE)$ as a function of ρ in Model (2)

As explained in Section 2, some of the methods compared are based on an AR(1) structure of the errors. An interesting point is to observe what happens when the errors are simulated from another dependence structure and

Table 2: h_{MISE}, $\hat{h}_{CVM}$, $\hat{h}_{PCV}$, $\hat{h}_{TSCV}$, $\hat{h}_{CVP}$, $\hat{h}_{PLIA}$, $\hat{h}_{PLIE}$ and $\hat{h}_{BOOT}$ under different correlation specifications in Model (1)

$h \times 10^{-1}$	MA(2)	AR(2)	$(1,1)$	$(1,1)_2$	MA(1)
h_{MISE}	2.929	3.535	3.485	2.121	2.778
$\hat{h}_{CVM}$	3.433	3.349	4.035	3.271	2.612
$\hat{h}_{PCV}$	2.799	2.272	2.662	2.885	2.356
$\hat{h}_{TSCV}$	2.764	1.002	2.734	2.243	2.692
$\hat{h}_{CVP}$	4.403	4.790	4.227	4.020	4.506
$\hat{h}_{PLIA}$	2.809	1.500	2.755	2.106	2.558
$\hat{h}_{PLIE}$	2.786	1.417	2.488	2.240	2.579
$\hat{h}_{BOOT}$	3.328	1.482	2.947	2.175	2.685
$\Delta MISE$					
$\hat{h}_{CVM}$	$6.50E-6$	$5.85E-6$	$6.93E-6$	$3.66E-6$	$5.36E-6$
$\hat{h}_{PCV}$	$1.61E-6$	$4.21E-5$	$2.62E-5$	$9.33E-7$	$6.21E-6$
$\hat{h}_{TSCV}$	$2.43E-5$	$3.30E-4$	$4.65E-6$	$7.00E-9$	$2.90E-8$
$\hat{h}_{CVP}$	$4.31E-5$	$1.63E-5$	$2.28E-5$	$3.88E-5$	$5.23E-5$
$\hat{h}_{PLIA}$	$2.30E-8$	$9.44E-5$	$2.81E-6$	$1.00E-10$	$2.60E-8$
$\hat{h}_{PLIE}$	$8.50E-8$	$1.13E-4$	$8.99E-6$	$2.00E-9$	$3.90E-8$
$\hat{h}_{BOOT}$	$6.60E-6$	$1.05E-4$	$4.89E-5$	$3.06E-7$	$1.39E-6$

the same bandwidths as before are used. To study this situation, several other $ARMA$-type dependence structures for the errors were considered. Specifically, we generated errors using the following models:

- An MA(2) model with normal distribution: $\varepsilon_i = e_i + 1.4e_{i-1} + 0.9e_{i-2}$.
- An AR(2) model with normal distribution: $\varepsilon_i = 0.4\varepsilon_{i-1} + 0.5\varepsilon_{i-2} + e_i$.
- An ARMA$(1,1)$ model with normal distribution: $\varepsilon_i - 0.8\varepsilon_{i-1} = e_i + 0.3e_{i-1}$. (denoted by $(1,1)$).
- An ARMA$(1,1)$ model with normal distribution: $\varepsilon_i - 0.8\varepsilon_{i-1} = e_i - 0.9e_{i-1}$. (denoted by $(1,1)_2$).
- An MA(1) model with normal distribution: $\varepsilon_i = e_i + 0.9e_{i-1}$.

The rest of the parameters used are the same as those used in the AR(1) case. Table 2 shows the results for Model (1).

4 Conclusions

Interesting conclusions can be deduced from the simulation study, some of which are:

1. We observed that the amount of positive dependence makes the optimal bandwidth, h_{MISE}, increase remarkably (see, for example, Table 1). $\hat{h}_{PLIA}$, $\hat{h}_{PLIE}$, $\hat{h}_{TSCV}$ and $\hat{h}_{BOOT}$ had the same tendency as h_{MISE} and their behavior was in general very good, especially when the errors are AR(1) (Figure 1). This is not surprising, taking into account that these selectors are based on an AR(1) structure of the errors. In the studied models with non-AR(1) errors, these bandwidths had reasonable behavior, and except in the AR(2) case, their means were close to the true optimal bandwidth and the efficiency measure $\Delta MISE$ of $\hat{h}_{PLIA}$ and $\hat{h}_{PLIE}$, for example, was smaller than this measure for the bandwidths that do not assume any parametric dependence structure in the errors, $\hat{h}_{CVM}$, $\hat{h}_{PCV}$, $\hat{h}_{CVP}$ (Table 2). Overall, this indicates reasonable robustness against correlation model misspecification in the bandwidths that assume AR(1) errors. In general, if the AR(1) model does not fit the observed correlation structure, this might negatively affect the behavior of the bandwidths based on this assumption for the errors. On the other hand, those selectors, such as $\hat{h}_{CVM}$ or $\hat{h}_{PCV}$, that do not suppose any parametric structure for the errors and simply use a parameter (l_n or g) to handle the amount of dependence, could have better performance in those cases. Obviously, the bandwidths based on AR(1) errors can be modified by assuming another parametric dependence structure for the errors, or in the case of $\hat{h}_{PLIA}$ and $\hat{h}_{PLIE}$, by using alternative ways to estimate $c(\varepsilon)$ or $\Gamma_{(n)} = E\left(\vec{\varepsilon}\vec{\varepsilon}^t\right)$ without making any parametric assumption for the errors.

2. Generally speaking, plug-in bandwidths showed better performance than the cross-validation bandwidths ($\hat{h}_{CVM}$, $\hat{h}_{PCV}$, $\hat{h}_{TSCV}$, $\hat{h}_{CVP}$) and $\hat{h}_{BOOT}$. Considering the bandwidths selected by a cross-validation method, $\hat{h}_{TSCV}$ was, in general, the best because it tracks its target h_{MISE} closely over the whole range of correlation values. Moreover, this bandwidth does not depend on any preliminary parameters.

3. Cross-validation for prediction gave bad results, providing very large bandwidths (Figures 2 and 6) without taking the amount of dependence into account. This is because this parameter is designed to predict a future value using the first b observations but not to estimate the whole regression curve.

4. As mentioned in Section 2, parameters l_n in $\hat{h}_{CVM}$ and g in $\hat{h}_{PCV}$ have a similar role. When the dependence increases the values of these two parameters must be increased in order for the bandwidths obtained by these two techniques to be larger and closer to the optimal bandwidth. In any case, $\hat{h}_{CVM}$ gave better results than $\hat{h}_{PCV}$ when the dependence was strong and positive. This is because modified cross-validation is able to produce larger

bandwidths when l_n increases than $\hat{h}_{PCV}$ with respect to g. As observed in Figures 2 and 6, the best values of $\hat{h}_{PCV}$ are similar when ρ increases, which does not happen for $\hat{h}_{CVM}$. In the following Table the effect of l_n and ρ can be observed for $\hat{h}_{CVM}$ in Model (1).

$\mathbf{h} \times 10^{-1}$	$\mathbf{-0.9}$	$\mathbf{-0.6}$	$\mathbf{-0.3}$	$\boldsymbol{\rho = 0}$	$\mathbf{0.3}$	$\mathbf{0.6}$	$\mathbf{0.9}$
h_{MISE}	1.444	1.848	2.131	2.374	2.697	2.939	3.586
$l_n = 0$	2.901	2.833	2.645	2.178	0.781	0.297	0.266
$l_n = 2$	2.893	2.534	2.346	2.449	2.452	2.218	0.830
$l_n = 5$	1.663	2.507	2.774	3.040	3.171	3.348	2.269
$l_n = 7$	2.181	2.885	3.122	3.345	3.507	3.753	2.894
$l_n = 10$	3.186	3.243	3.494	3.744	3.975	4.153	3.619
$\mathbf{\Delta MISE \times 10^{-6}}$							
$l_n = 0$	2.21	1.11	0.71	19.7	610	1810	1540
$l_n = 2$	2.41	0.65	0.95	3.14	14.8	81.7	727
$l_n = 5$	0.09	0.72	1.42	3.41	5.83	14.4	138
$l_n = 7$	0.32	1.70	3.01	5.46	7.94	15.5	60.1
$l_n = 10$	5.45	4.05	7.08	11.2	15.0	21.6	22.9

5. $\hat{h}_{PLIA}$ and $\hat{h}_{PLIE}$ presented similar and good performance, especially if the regression function is relatively *smooth* (Figure 3). If $m(x)$ is more variable, $\hat{h}_{PLIE}$ works better than $\hat{h}_{PLIA}$ (Figure 7).

6. With respect to the standard deviations of the selectors, first, they are greater when the dependence increases, and second, for each value of ρ, the less variable bandwidths are those obtained using a plug-in method, while the more variable bandwidths are $\hat{h}_{CVM}$, $\hat{h}_{PCV}$ and $\hat{h}_{CVP}$.

6. For the GLPR estimator, $\hat{h}_{PLIEG}$ showed bad behavior, while $\hat{h}_{PLIAG}$ worked really well, even better than $\hat{h}_{PLIA}$ and $\hat{h}_{PLIE}$ (as seen in Figure 5).

7. From this study, we can see that the secondary parameters that must be chosen before applying each method could be more important than initially supposed. We should try to design automatic techniques to select these parameters to avoid a subjective selection, which can produce unsatisfactory results in bandwidth selection and in the corresponding nonparametric estimator. For example, in the case of $\hat{h}_{PLIA}$ (that, in sight of simulations, seems to be a good bandwidth selector), $m''(x)$ appearing in the denominator of (11) when a local linear fit is used can be estimated by a local cubic polynomial fit, using a plug-in bandwidth similar as (11), but with a different constant $C_{2,3}(K)$. Now, in this new plug-in bandwidth, $m^{(iv)}(x)$ must be estimated. This can be done by globally fitting a polynomial of degree six and next calculating the fourth derivative of this fitted curve. As for the estimation of $c(\varepsilon)$ appearing in the numerator of the plug-in bandwidths, we have observed that those methods based on differences of the response variables and those using nonparametric residuals (using, for example, a pilot

bandwidth computed by the TSCV method) provide similar good results. The same rules to choose pilot bandwidths for $\hat{h}_{PLIE}$ can be used.

5 Acknowledgments

The authors wish to thank an Associated Editor and two referees for their helpful comments and suggestions. This work was partially supported by grant PGIDIT03PXIC10505PN and MCyT Grant BFM2002-00265 (European FEDER support included).

References

Cao, R. (1999), 'An overview of bootstrap methods for estimating and predicting in time series', *TEST* **8**(1), 95–116.

Chu, C.K. & Marron, J.S. (1991), 'Comparison of two bandwidth selectors with dependent errors', *The Annals of Statistics* **19**(4), 1906–1918.

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

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

Fan, J., Gijbels, I., Hu, T.C. & Huang, L.S. (1996), 'A study of variable bandwidth selection for local polynomial regression', *Statistica Sinica* **6**, 113–127.

Francisco-Fernández, M. & Vilar-Fernández, J.M. (2001), 'Local polynomial regression estimation with correlated errors', *Communications in Statistics: Theory and Methods* **30**(7), 1271–1293.

Francisco-Fernández, M., Opsomer, J.D. & Vilar-Fernández, J.M. (2004), A plug-in bandwidth selector for local polynomial regression estimator with correlated errors, *Journal of Nonparametric Statistics* **16**(1-2), 127–152.

Hall, P., Lahiri, S.N. & Polzehl, J. (1995), 'On bandwidth choice in nonparametric regression with both short- and long-range dependent errors', *The Annals of Statistics* **23**(6), 1921–1936.

Hall, P. & Van Keilegom, I. (2003), Using difference-based methods for inference in nonparametric regression with time-series errors, *Journal of the Royal Statistical Society, Series B* **65**, 443–456.

Härdle, W., Horowitz, J. & Kneiss, J.P. (2003), 'Bootstrap methods for time series', *International Statistical Review* **71**(2), 435–455.

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

Härdle, W., Tsybakov, A. & Yang, L. (1998), 'Nonparametric vector regression', *Journal of Statistical Planning and Inference* **68**, 221–245.

Härdle, W. & Vieu, P. (1992), 'Kernel regression smoothing of time series', *Journal of Time Series Analysis* **13**, 209–232.

Hart, J. (1994), 'Automated kernel smoothing of dependent data by using time series cross-validation', *Journal of the Royal Statistical Society, Series B* **56**(3), 529–542.

Hart, J. & Vieu, P. (1990), 'Data-driven bandwidth choice for density estimation based on dependent data', *The Annals of Statistics* **18**, 873–890.

Hart, J. & Wehrly, T.E. (1986), 'Kernel regression using repeated measurements data', *Journal of the American Statistical Association* **81**, 1080–1088.

Marron, J.S. (1987), 'Partitioned cross-validation', *Econometric Review* **6**, 271–284.

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

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

Opsomer, J.D., Wang, Y. & Yang, Y. (2001), 'Nonparametric regression with correlated errors', *Statistical Science* **16**, 134–153.

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

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

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

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

Yao, Q. & Tong, H. (1998), 'Cross-validation bandwidth selections for regression estimation based on dependent data', *Journal of Statistical Planning and Inference* **68**, 387–415.