Residual correlation as a bandwidth selection rule in nonparametric regression

Elodie Brunel-Piccinini *, Irène Larramendy-Valverde †

January 7, 2004

Abstract

This paper focuses on the problem of using the structure of residual correlation in a non-parametric model to select an automatic bandwidth when the regression function is estimated by a kernel method. A bandwidth selection rule related to the empirical autocovariance of the residuals is defined and is shown to be asymptotically optimal in the sense of the mean squared error. Simulation trials complete the theoretical backgrounds and illustrate how the criterion performs in practice for different features of the regression curve.

AMS 2000 Subject Classifications: Primary 62G08.

Key words and phrases: Functional estimation, nonparametric regression, kernel estimates, residual autocorrelations, bandwidth selection.

*Corresponding author — MAP 5 FRE CNRS 2428 — Université Paris 5 — 45, rue des Saints Pères, 75 270 Paris cedex 06 — e-mail : brunel@math-info.univ-paris5.fr —

†Equipe de Probabilités et Statistique, Institut de Mathématiques et de Modélisation de Montpellier, UMR 5149, Université Montpellier 2 — Case courrier 051 — Place Eugène Bataillon, 34095 Montpellier Cedex 5 — e-mail : larramen@stat.math.univ-montp2.fr —

1 Introduction

Let us introduce the nonparametric regression model in the case of the fixed design points

$$Y_i = m(x_i) + \varepsilon_i \, , \, i = 1, \dots, n.$$

where m is a smooth deterministic mean function on $[0, 1]$ and the ε_i's are independent and identically distributed errors with zero mean and variance σ^2. The design is fixed e.g. $x_i = i/n$ for $i \in \{1, \dots, n\}$.

Nonparametric estimation of $m(.)$ can be successfully accomplished using kernel methods.

The adequate kernel estimate for the fixed design is defined by Priestley and Chao (1972) as

$$\hat{m}_{h_n}(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - i/n}{h_n}\right) Y_i \tag{1}$$

where K is a kernel function, $h_n \in I\!R^+$ is the bandwidth or smoothing parameter. As well-known, the choice of the bandwidth h_n is one of the crucial point in the practical use of the estimate $\hat{m}_{h_n}$. Suppose that h_n is in some set of interest $H_n \subseteq I\!R^+$. A bandwidth selection rule $\hat{h}_n$ is a H_n-valued function of the data set. In practical situations, it provides a device to build an automatic and objective bandwidth.

Lots of litterature in the two last decades have been devoted to improve automatic bandwidth selection rules. Out of them, the best known are the cross validation and plug-in methods, see Rice (1984), Härdle, Hall and Marron (1988), Chiu (1990), Ruppert, Sheather and Wand (1995), Boente, Fraiman and Meloche (1997). Härdle and Marron (1985) showed that cross-validation provides an asymptotically optimal automatic bandwidth. However, it is known that cross-validation technique leads to an undersmoothing bandwidth. Chiu (1990) proposed a criterion involving penalized functions of the sum of squared residuals to remedy the problem of undersmoothing.

All of these methods are based on the idea that the estimate $\hat{m}_{h_n}(.)$ should be as close as possible to the true function $m(.)$ in some sense.

Instead of focusing on the regression function which is unknown, Braun and Rousson (2000) proposed recently to choose the bandwidth minimizing the Box-Pierce statistic (see Box, 1971). The idea is heuristically different : since the true errors are supposed uncorrelated, it is natural to render the residuals as close as possible to the true errors. Moreover, Rousson and Braun demonstrated via a simulation study, that the method is better than it is by cross-validation to estimate the number of turning points in the true curve. Another motivation to use the structure of correlation of the residuals can be found in Altman (1990). In fact, it is shown that the bandwidth minimizing the asymptotic bias of the residual empirical autocovariance is of the same order than the asymptotic bandwidth $h_n^\star$.

In the same spirit, we propose a criterion that takes into account the correlation structure of the residuals. We establish the first optimality result for this kind of bandwidth selection rule. The paper is organized as follows. The proposed bandwidth selector is presented in Section 2. Notations, assumptions and optimality results are given in Section 3. Some numerical simulations are presented in Section 4. Appendix A is devoted to the proofs of technical Lemmas.

2 The bandwidth selection rule

Consider a kernel estimate $\hat{m}_{h_n}(.)$ of the true regression function $m(.)$. We suppose that the data are available until rank $n+1$ for better readability of mathematical calculations. We define the sample autocovariance at lag $k=1$ by

$$\hat{\gamma}_1(h_n) = \frac{1}{n}\sum_{j=1}^{n}\left[Y_j - \hat{m}_{h_n}^{j}(x_j)\right]\left[Y_{j+1} - \hat{m}_{h_n}^{j+1}(x_{j+1})\right]$$

where $\hat{m}_{h_n}^{j}(x)$ is the standard leave-one-out estimate of $m(x)$ defined for all $j \in \{1,\ldots,n+1\}$ by deleting the j-th observation from the sample e.g.

$$\hat{m}_{h_n}^{j}(x) = \frac{1}{nh_n} \sum_{i \neq j} K\left(\frac{x - x_i}{h_n}\right) Y_i \,. \tag{2}$$

In this paper, the proposed bandwidth selection rule consists in

$$\textit{Choosing } h_n \textit{ so that } |\hat{\gamma}_1(h_n)| \textit{ is minimum.} \tag{3}$$

One can notice that the statistic $\hat{\gamma}_1(h_n)$ is not the sample autocorrelation but the empirical product of moments at lag one of the residual process. It can be seen as a correlation-type criterion. Nevertheless, our approach is close to the more general Braun and Rousson (2000) one's and it is instructive to have a detailed look at their work. Indeed, they proposed to choose h_n that minimizes the Box-Pierce statistic

$$\sum_{j=1}^{m} r_j^2$$

where r_j stands for the sample autocorrelation of the residuals at lag j. This choice really takes into account the fact that the errors are serially uncorrelated in the regression model. The maximal lag m considered to compute the statistitic is arbitrarily choosen. For $m = 1$, the statistic equals the squared sample autocorrelation at lag 1. Our proposal can be seen as a Braun and Rousson-type bandwith selection rule. Further research may investigate the theoretical background of such a general procedure. Hereafter, we only consider the statistic $\hat{\gamma}_1(h_n)$ to summarize the correlation structure of the residuals and we obtain an optimality result for it. But, it would not be surprising that the same results hold for the general Box-Pierce statistic.

Now, in order to evaluate the performance of an automatic kernel estimate, it is usual to consider the following measures of accuracy :

∘ the average squared error d_A defined by

$$d_A(\hat{m}_{h_n}, m) = \frac{1}{n} \sum_{j=1}^{n} \Big(\hat{m}_{h_n}(x_j) - m(x_j) \Big)^2 \tag{4}$$

∘ the mean integrated squared error d_M defined by

$$d_M(\hat{m}_{h_n}, m) = \mathbb{E} \int \Big(\hat{m}_{h_n}(x) - m(x) \Big)^2 \, dx \, .$$

The asymptotic decomposition of d_M is well-known and can be written (see for example Bosq, 1999).

$$d_M(\hat{m}_{h_n}, m) = \frac{\sigma^2}{n h_n} ||K||_2^2 + \frac{h_n^4}{4} \, \sigma^2(K) \, ||m"||_2^2 + o\left(h_n^4 + \frac{1}{n h_n} \right) \tag{5}$$

where $||.||_2$ is the L_2-norm. If K is a positive and symmetric kernel, the only condition required to achieve the convergence rate given by (5) is the existence and boundedness of the first and second order derivatives of m.

Härdle and Marron (1985) showed that d_A and d_M are asymptotically almost surely equivalent so it makes possible to use one of both error criteria. In the following, $h_n^\star$ will denote the minimizer of the asymptotic mean integrated squared error whose expression is easily derived from (5),

$$h_n^\star = \left(\frac{\sigma^2 \, ||K||_2^2}{\sigma^2(K) \, ||m"||_2^2} \right)^{1/5} \times n^{-1/5}.$$

Suppose that $d(\hat{m}_{h_n}, m)$ denotes a given measure of accuracy (for instance d_A or d_M) for the estimate $\hat{m}_{h_n}$, then the bandwidth selection rule $\hat{h}_n$ is said to be asymptotically optimal with respect to d, in the sense defined by Shibata (1981), if

$$\lim_{n\to+\infty} \frac{d(\hat{m}_{\hat{h}_n}, m)}{\inf_{h_n\in H_n} d(\hat{m}_{h_n}, m)} = 1 \quad a.s. \tag{6}$$

3 Asymptotic Optimality of the correlation criterion for the bandwidth selection

3.1 Notations and Assumptions

(A_1) : The bandwidth h_n belongs to H_n, finite subset of the interval $\overline{H}_n = [\alpha\, n^{-1/5}\,;\,\beta\, n^{-1/5}]$, which cardinal increases geometrically *ie* $\sharp\,(H_n) = O\big(n^\rho\big)$; the constant $\rho > 0$ can be chosen arbitrary large. Condition (A_1) reflects practical situations where the optimum $\hat{h}_n$ is obtained by minimizing the correlation-type criterion over a finite grid H_n for fixed n. The grid can be made as tight as desired through the choice of ρ.

(A_2) : The moments of ε are bounded in the sense that there are positive constants $M_1, M_2, \ldots$ so that for all $i \in \mathbb{N}^\star$,

$$\mathbb{E}\,|\varepsilon_i|^k \leq M_k \ , \ k = 1, 2, \ldots .$$

(A_3) : The kernel function K is symmetric, supported on $[-1/2; 1/2]$ and admits a continuous first derivative so that for all x, y in $[-1/2; 1/2]$,

$$|K(y) - K(x)| \leq C(K)\,|y - x|$$

with $C(K) = \sup\limits_{x\in[-1/2;1/2]} |K'(x)|$.

The second order moment of K exists and is denoted by $\sigma^2(K) = \int v^2\,K(v)\,dv$.

(A_4) : The mean function m admits a continuous first derivative on $[0,1]$, so that for all x, y in $[0,1]$,

$$|m(y) - m(x)| \leq C(m)\, |y - x|$$

with $C(m) = \sup\limits_{x \in [0,1]} |m'(x)|$.

The function m admits also a bounded second derivative on $[0,1]$.

3.2 Optimality result

Theorem 1 *Under assumptions* $(\mathrm{A}_1\text{–}\mathrm{A}_4)$ *, the bandwidth selection rule* $\hat{h}_n$ *defined by (3) is asymptotically optimal in the sense that,*

$$\lim_{n \to +\infty} \frac{d_A(\hat{m}_{\hat{h}_n}, m)}{\inf_{h_n \in H_n} d_A(\hat{m}_{h_n}, m)} = 1 \quad a.s. \tag{7}$$

with $\hat{m}_{\hat{h}_n}$ *defined by (1) and* d_A *the measure of accuracy defined by (4).*

Lemma 1 *Under assumptions* $(\mathrm{A}_1\text{–}\mathrm{A}_4)$

$$\sup_{h_n \in H_n} \frac{|I_n|}{d_A(\hat{m}_{h_n}, m)} \to 0 \quad a.s.$$

with

$$I_n \quad = \quad \frac{1}{n} \sum_{j=1}^{n} \left[\Big(\hat{m}_{h_n}(x_j) - m(x_j)\Big)^2 - \Big(\hat{m}^{j+1}_{h_n}(x_j) - m(x_j)\Big)\Big(\hat{m}^{j}_{h_n}(x_{j+1}) - m(x_{j+1})\Big) \right],$$

$\hat{m}^{j}_{h_n}(.)$ *being the leave-one-out estimate of* $m(.)$ *defined by (2).*

Lemma 2 *Under assumptions* $(\mathrm{A}_1\text{–}\mathrm{A}_4)$

$$\sup_{h_n \in H_n} \frac{|J_n|}{d_A(\hat{m}_{h_n}, m)} \to 0 \quad a.s. \qquad and \qquad \sup_{h_n \in H_n} \frac{|K_n|}{d_A(\hat{m}_{h_n}, m)} \to 0 \quad a.s.$$

with

$$J_n = \frac{1}{n}\sum_{j=1}^{n}\Big(\hat{m}^{j}_{h_n}(x_{j+1}) - m(x_{j+1})\Big)\varepsilon_j \quad \textit{and} \quad K_n = \frac{1}{n}\sum_{j=1}^{n}\Big(\hat{m}^{j+1}_{h_n}(x_j) - m(x_j)\Big)\varepsilon_{j+1},$$

$\hat{m}^{j}_{h_n}(.)$ *being the leave-one-out estimate of* $m(.)$ *defined by (2).*

Proof of Theorem : As $\frac{1}{n}\sum_{j=1}^{n}\varepsilon_j\,\varepsilon_{j+1}$ does not depend on h_n, the correlation-type criterion $|\hat{\gamma}_1(h_n)|$ can be replaced by the statistic $|\hat{\gamma}_1(h_n) - \frac{1}{n}\sum_{j=1}^{n}\varepsilon_j\,\varepsilon_{j+1}|$. So, it is known that a sufficient condition to show (7) is that

$$\sup_{h_n,h'_n}\left|\frac{d_A(\hat{m}_{h_n},m) - d_A(\hat{m}_{h'_n},m) - |\hat{\gamma}_1(h_n) - \frac{1}{n}\sum_{j=1}^{n}\varepsilon_j\,\varepsilon_{j+1}| + |\hat{\gamma}_1(h'_n) - \frac{1}{n}\sum_{j=1}^{n}\varepsilon_j\,\varepsilon_{j+1}|}{d_A(\hat{m}_{h_n},m) + d_A(\hat{m}_{h'_n},m)}\right|$$

tends to 0 see Härdle and Marron (1985).

In fact, it is enough to show that

$$\sup_{h_n\in H_n}\frac{\left|d_A(\hat{m}_{h_n},m) - \Big(\hat{\gamma}_1(h_n) - \frac{1}{n}\sum_{j=1}^{n}\varepsilon_j\,\varepsilon_{j+1}\Big)\right|}{d_A(\hat{m}_{h_n},m)} \to 0 \quad a.s.$$

Then, let us decompose $d_A(\hat{m}_{h_n},m) - \Big(\hat{\gamma}_1(h_n) - \frac{1}{n}\sum_{j=1}^{n}\varepsilon_j\,\varepsilon_{j+1}\Big)$ as follows,

$$\begin{aligned}
& d_A(\hat{m}_{h_n},m) - \hat{\gamma}_1(h_n) + \frac{1}{n}\sum_{j=1}^{n}\varepsilon_j\,\varepsilon_{j+1} \\
= \; & \frac{1}{n}\sum_{j=1}^{n}\left[\Big(\hat{m}_{h_n}(x_j) - m(x_j)\Big)^2 - \Big(\hat{m}^{j+1}_{h_n}(x_j) - m(x_j)\Big)\Big(\hat{m}^{j}_{h_n}(x_{j+1}) - m(x_{j+1})\Big)\right] \\
& + \frac{1}{n}\sum_{j=1}^{n}\Big(\hat{m}^{j}_{h_n}(x_{j+1}) - m(x_{j+1})\Big)\varepsilon_j + \frac{1}{n}\sum_{j=1}^{n}\Big(\hat{m}^{j+1}_{h_n}(x_j) - m(x_j)\Big)\varepsilon_{j+1} \\
= \; & I_n + J_n + K_n
\end{aligned}$$

And, the proof will be complete by Lemmas 1 and 2. □

In the following, for the convenience of the reader, let us denote $\hat{m}_n = \hat{m}_{h_n}$ and $\hat{m}_n^j = \hat{m}_{h_n}^j$.

Proof of Lemma 1 : The proof is organized in the following two steps :

i — First, replace $d_A(\hat{m}_n, m)$ by the mean integrated squared error $d_M(\hat{m}_n, m)$ in the statistic involved in Lemma 1 and let $\epsilon > 0$,

$$\mathbb{P}\left(\sup_{h_n \in H_n} \frac{|I_n|}{d_M(\hat{m}_n, m)} > \epsilon\right) \leq \sharp H_n \sup_{h_n \in H_n} \mathbb{P}\Big(|I_n| > \epsilon\, d_M(\hat{m}_n, m)\Big).$$

Using assumption (A_1), we get for some $C > 0$,

$$\mathbb{P}\left(\sup_{h_n \in H_n} \frac{|I_n|}{d_M(\hat{m}_n, m)} > \epsilon\right) \leq C\, n^\rho \sup_{h_n \in H_n} \mathbb{P}\Big(|I_n| > \epsilon\, d_M(\hat{m}_n, m)\Big).$$

By Markov inequality with an arbitrarily great integer k, we obtain

$$\mathbb{P}\left(\sup_{h_n \in H_n} \frac{|I_n|}{d_M(\hat{m}_n, m)} > \epsilon\right) \leq C\, n^\rho\, \epsilon^{-2k} \sup_{h_n \in H_n} \left(\frac{\mathbb{E}\,|I_n|^{2k}}{[d_M(\hat{m}_n, m)]^{2k}}\right).$$

The asymptotic decomposition (5) allows to give a lower bound for $d_M(\hat{m}_n, m)$ for n large enough,

$$d_M(\hat{m}_n, m) \geq 2\,\sigma^2\, ||K||_2^2\, (nh_n)^{-1}.$$

From which, we deduce

$$\mathbb{P}\left(\sup_{h_n \in H_n} \frac{|I_n|}{d_M(\hat{m}_n, m)} > \epsilon\right) \leq C\, \left(2\,\sigma^2\, ||K||_2^2\right)^{-2k} \epsilon^{-2k} \times n^\rho \times \sup_{h_n \in H_n} \Big((nh_n)^{2k}\, \mathbb{E}\,|I_n|^{2k}\Big). \tag{8}$$

The proof will be complete if the above probability is upper bounded by the main term of a convergent serie. That is what we prove hereafter.

ii — The present step is now devoted to give an upper bound for the expectation $\mathbb{E}\,|I_n|^{2k}$. Let us decompose I_n as follows $I_n = I_{1,n} + I_{2,n} + I_{3,n} + r_n + s_n$ with

$$
\begin{aligned}
I_{1,n} &= \frac{1}{n^3 h_n^2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} a_{ij}\, \varepsilon_i \varepsilon_j, \\
I_{2,n} &= \frac{1}{n^3 h_n^2} \sum_{i=1}^{n} a_{ii}\, (\varepsilon_i^2 - \sigma^2),
\end{aligned}
\tag{9}
$$

and $a_{ij} = \sum_{\ell=1}^{n} K\left(\frac{\ell - i}{nh_n}\right)\left[K\left(\frac{\ell - j}{nh_n}\right) - K\left(\frac{\ell - j + 1}{nh_n}\right)\right] + 2K\left(\frac{1}{nh_n}\right) K\left(\frac{i-j}{nh_n}\right)$,

$$
\begin{aligned}
I_{3,n} &= \frac{2}{n^2 h_n} \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{j-i}{nh_n}\right) \left(\frac{1}{nh_n} \sum_{\ell=1}^{n} K\left(\frac{j-\ell}{nh_n}\right) m(x_\ell) - m(x_j)\right) \varepsilon_i \\
&- \frac{1}{n^2 h_n} \sum_{i=1}^{n} \sum_{j \neq i-1} K\left(\frac{j-i}{nh_n}\right) \left(\frac{1}{nh_n} \sum_{\ell \neq j} K\left(\frac{j+1-\ell}{nh_n}\right) m(x_\ell) - m(x_{j+1})\right) \varepsilon_i \\
&- \frac{1}{n^2 h_n} \sum_{i=1}^{n} \sum_{j \neq i} K\left(\frac{j+1-i}{nh_n}\right) \left(\frac{1}{nh_n} \sum_{\ell \neq j+1} K\left(\frac{j-\ell}{nh_n}\right) m(x_\ell) - m(x_j)\right) \varepsilon_i \,,
\end{aligned}
$$

and

$$
\begin{aligned}
r_n &= \frac{\sigma^2}{n^3 h_n^2} \sum_{i=1}^{n} a_{ii} \\
s_n &= \frac{1}{n} \sum_{j=1}^{n} \left(\frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{j-i}{nh_n}\right) m(x_i) - m(x_j)\right)^2 \\
&- \frac{1}{n} \sum_{j=1}^{n} \left(\frac{1}{nh_n} \sum_{i \neq j+1}^{n} K\left(\frac{j-i}{nh_n}\right) m(x_i) - m(x_j)\right) \left(\frac{1}{nh_n} \sum_{i \neq j}^{n} K\left(\frac{j+1-i}{nh_n}\right) m(x_i) - m(x_{j+1})\right).
\end{aligned}
$$

From the C_{2k}-inequality (see Shorack and Wellner, 1986) applied consecutively to the five terms involved in the decomposition of I_n, we get

$$
\begin{aligned}
I_n^{2k} &\leq 2^{2k-1}\, I_{1,n}^{2k} + 2^{4k-2}\, I_{2,n}^{2k} + 2^{6k-3}\, I_{3,n}^{2k} + 2^{6k-3} (s_n + r_n)^{2k} \\
&\leq 2^{6k-3} \left(I_{1,n}^{2k} + I_{2,n}^{2k} + I_{3,n}^{2k} + (r_n + s_n)^{2k}\right) \quad \text{which is always satisfies for } k \geq 1.
\end{aligned}
$$

From this inequality and (8), it is sufficient to provide suitable upper bounds for the expected values of $I_{i,n}^{2k}$, for $i = 1, 2, 3$, and for $(r_n + s_n)^{2k}$.

First, applying Lemma 3 (see Appendix A) with $s = 2k$, and using assumption (A_2) we obtain

$$\begin{aligned} I\!E\left(I_{1,n}^{2k}\right) &\leq 2^{6k} C_{2k}\, (C_{4k}\, M_{4k})^{1/2}\, n^{-6k} h_n^{-4k} \left(\sum_{i=1}^{n} \sum_{j \neq i} a_{ij}^2\right)^k \\ &\leq 2^{6k} C_{2k}\, (C_{4k}\, M_{4k})^{1/2}\, (nh_n)^{-4k} \sup_{i,j} (a_{ij})^{2k} \end{aligned}$$

with a_{ij} defined above and with the constants C_{2k} and C_{4k} defined in Lemma 3.

Now, thanks to lemma 5 with the kernel function K satisfying the assumption (A_3) and for n large enough, and for $1 \leq i, j \leq n$,

$$|a_{ij}| \leq C(K) \left(1 + \frac{2\, C(K)}{nh_n}\right) + 2\, ||K||_\infty \leq C(K)\, (1 + 2\, C(K)) + 2\, ||K||_\infty \tag{10}$$

Hereafter, we denote $A(K) = C(K)\, (1 + 2\, C(K)) + 2\, ||K||_\infty$ the upper bound for the (a_{ij})'s. Hence, we deduce the following upper bound for $I\!E\left(I_{1,n}^{2k}\right)$,

$$I\!E\left(I_{1,n}^{2k}\right) \leq 2^{8k} C_{2k} \sqrt{C_{4k}\, M_{4k}}\, A(K)^{2k}\, (nh_n)^{-4k}\,. \tag{11}$$

For the second term $I_{2,n}$ of the decomposition of I_n, applying Lemma 3, with the coefficient

$$b_i = a_{ii} = \sum_{\ell=1}^{n} K\left(\frac{\ell - i}{nh_n}\right) \left[K\left(\frac{\ell - i}{nh_n}\right) - K\left(\frac{\ell + 1 - i}{nh_n}\right)\right] + 2K(0)\, K\left(\frac{1}{nh_n}\right)$$

and $\varepsilon_i' = \varepsilon_i - \sigma^2$, we get

$$I\!E\left(I_{2,n}^{2k}\right) = n^{-6k} h_n^{-4k}\, I\!E\left(\sum_{i=1}^{n} b_i\, \varepsilon_i'^2\right)^{2k} \leq 2^{2k}\, C_{2k}\, M_{2k}\, n^{-6k} h_n^{-4k} \left(\sum_{i=1}^{n} b_i^2\right)^k.$$

Notice that $b_i = a_{ii}$, and it follows immediately from inequality (10) that,

$$\mathbb{E}\left(I_{2,n}^{2k}\right) \leq 2^{4k}\, C_{2k}\, M_{2k}\, n^{-5k} h_n^{-4k}\, A(K)^{2k}\,. \tag{12}$$

The term $I_{3,n}$ is the sum of two terms $I'_{3,n}$ and $I''_{3,n}$ which have quite similar behaviour. The first one is treated as follows

$$\begin{aligned}
I'_{3,n} &= \frac{1}{n^2 h_n}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{j-i}{nh_n}\right) B_n(x_j)\varepsilon_i \\
&- \frac{1}{n^2 h_n}\sum_{i=1}^{n}\sum_{j\neq i-1} K\left(\frac{j-i}{nh_n}\right)\left(\frac{1}{nh_n}\sum_{\ell\neq j} K\left(\frac{j+1-\ell}{nh_n}\right) m(x_\ell) - m(x_{j+1})\right)\varepsilon_i \\
&= \frac{1}{n^2 h_n}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{j-i}{nh_n}\right)\left(B_n(x_j) - B_n(x_{j+1}) + \frac{1}{nh_n}K\left(\frac{1}{nh_n}\right) m(x_j)\right)\varepsilon_i \\
&+ \frac{1}{n^2 h_n}\sum_{i=1}^{n} K\left(\frac{1}{nh_n}\right)\left(B_n(x_i) - \frac{1}{nh_n}K\left(\frac{1}{nh_n}\right) m(x_{i-1})\right)\varepsilon_i
\end{aligned}$$

where $B_n(x) = \mathbb{E}\left[\hat{m}_n(x) - m(x)\right]$ stands for the ponctual bias of the estimate. By Lemma 3, we have

$$\mathbb{E}(I_{3,n}^{'\,2k}) \leq 2^{2k}\, C_{2k}\, M_{2k}\, n^{-4k} h_n^{-2k}\left(\sum_{i=1}^{n} b_i^2\right)^{k} \tag{13}$$

with the coefficient

$$\begin{aligned}
b_i &= \sum_{j=1}^{n} K\left(\frac{j-i}{nh_n}\right)\left(B_n(x_j) - B_n(x_{j+1}) + \frac{1}{nh_n}K\left(\frac{1}{nh_n}\right) m(x_j)\right) \\
&+ K\left(\frac{1}{nh_n}\right)\left(B_n(x_i) - \frac{1}{nh_n}K\left(\frac{1}{nh_n}\right) m(x_{i-1})\right).
\end{aligned}$$

According to the statement *ii* of Lemma 6 (see Appendix A) and condition (A_4) on the boundedness of m, we get

$$
\begin{aligned}
|b_i| &\le \sum_{j=j_0(i)}^{j_1(i)} ||K||_\infty \left(3\,C(K)\,||m||_\infty (nh_n)^{-1} + C_m\, n^{-1} + ||K||_\infty\; ||m||_\infty (nh_n)^{-1}\right) \\
&+ ||K||_\infty \left(\sup_{1\le i\le n} |B_n(x_i)| + ||K||_\infty\, ||m||_\infty (nh_n)^{-1} \right) \\
&\le (j_1(i) - j_0(i) + 1)\, ||K||_\infty \left((3\,C(K) + ||K||_\infty)\, ||m||_\infty (nh_n)^{-1} + C_m\, n^{-1} \right) \\
&+ ||K||_\infty \left(\sup_{1\le i\le n} |B_n(x_i)| + ||K||_\infty\, ||m||_\infty (nh_n)^{-1} \right).
\end{aligned}
$$

By Lemma 6 it is easily seen that $\sup_{1\le i\le n} |B_n(x_i)| = O(h_n^2 + (nh_n)^{-1})$, therefore there exists a constant $C > 0$ leading to $\sup_{1\le i\le n} |B_n(x_i)| \le C\,(h_n^2 + (nh_n)^{-1})$.

Moreover, as we have already seen $j_1(i) - j_0(i) + 1 \le 2\,nh_n$, for n large enough, so that we obtain the final upper bound,

$$
\begin{aligned}
|b_i| \;\le\; & 2\,||K||_\infty\, ||m||_\infty (3\,C(K) + \,||K||_\infty) + (C(m) + C\,||K||_\infty h_n)\, h_n \\
& + ||K||_\infty\, (||K||_\infty\, ||m||_\infty + C)\, (nh_n)^{-1}.
\end{aligned}
$$

But, the second and the third terms are respectively $O(h_n)$ and $O(nh_n)^{-1}$ so they are negligible in front of the first term which is a constant, therefore it allows us to omit them in the sense that for n large enough,

$$
|b_i| \le 3\,||K||_\infty\, ||m||_\infty (3\,C(K) + \,||K||_\infty)\,.
$$

Finally, using the upper bound for $|b_i|$ in the inequality (13) leads to

$$
I\!E(I_{3,n}^{'\,2k}) \;\le\; 6^{2k}\, C_{2k}\, M_{2k}\, \left(||K||_\infty\, ||m||_\infty (3\,C(K) + \,||K||_\infty)\right)^{2k}\, n^{-3k} h_n^{-2k}\,. \tag{14}
$$

Now, it is straightforward to see that $I''_{3,n}$ can be treated in a similar way to end the study of the term $I_{3,n}$.

From (10), we directly obtain for the term r_n,

$$
|r_n| \;\le\; \frac{\sigma^2}{n^2 h_n^2}\, A(K)\,. \tag{15}
$$

The term s_n can be written

$$
\begin{aligned}
s_n &= \frac{1}{n}\sum_{j=1}^{n} B_n(x_j)^2 - \frac{1}{n}\sum_{j=1}^{n}\left(B_n(x_j) - \frac{1}{nh_n}K\left(\frac{1}{nh_n}\right)\, m(x_{j+1})\right) \\
&\quad \times \left(B_n(x_{j+1}) - \frac{1}{nh_n}K\left(\frac{1}{nh_n}\right)\, m(x_j)\right)
\end{aligned}
$$

It can be expressed as the sum of four simple terms as follows,

$$
\begin{aligned}
s_n &= \frac{1}{n}\sum_{j=1}^{n} B_n(x_j)\,(B_n(x_j) - B_n(x_{j+1})) + \frac{1}{n^2h_n}\sum_{j=1}^{n} B_n(x_j)\,K\left(\frac{1}{nh_n}\right)\, m(x_j) \\
&+ \frac{1}{n^2h_n}\sum_{j=1}^{n} B_n(x_{j+1})\,K\left(\frac{1}{nh_n}\right)\, m(x_{j+1}) - \frac{1}{n^3h_n^2}K^2\left(\frac{1}{nh_n}\right)\sum_{j=1}^{n} m(x_j)\, m(x_{j+1})\,.
\end{aligned}
$$

From assumptions (A_3) and (A_4) that ensure the boundedness of both K and m and using Lemma 6, we get

$$
\begin{aligned}
|s_n| &\leq \left(3\,C(K)||m||_\infty\,(nh_n)^{-1} + C(m)\,n^{-1}\right)\,\frac{1}{n}\sum_{j=1}^{n}|B_n(x_j)| \\
&+ ||m||_\infty\,||K||_\infty\,(nh_n)^{-1}\frac{1}{n}\sum_{j=1}^{n}(|B_n(x_j)| + |B_n(x_{j+1})|) + \frac{1}{n^2h_n^2}||m||_\infty^2\,||K||_\infty^2\,.
\end{aligned}
$$

Thanks to the part *i)* of Lemma 6, there exists $C > 0$ such that for n large enough,

$$
\sup_x |B_n(x)| \leq C\,\left(h_n^2 + (nh_n)^{-1}\right)
$$

which gives

$$
\begin{aligned}
|s_n| &\leq \left(3\,C(K)||m||_\infty\,(nh_n)^{-1} + C(m)\,n^{-1}\right)\,C\,\left(h_n^2 + (nh_n)^{-1}\right) \\
&+ 2\,C\,||m||_\infty\,||K||_\infty\,(nh_n)^{-1}\left(h_n^2 + (nh_n)^{-1}\right) + \frac{1}{n^2h_n^2}||m||_\infty^2\,||K||_\infty^2\,. \qquad (16)
\end{aligned}
$$

From this last inequality and from (11), (12), (14) and (15), the right hand side of (8) is upper bounded by the general term of a convergent Riemann's serie as soon as k is choosen as $k > (1+\rho)$, which concludes the proof of the lemma.

$\square$

Proof of Lemma 2 :

By circular permutation of both j and $j+1$, the terms J_n and K_n are quite similar and will be treated in the same way. So the proof is only given for J_n.

$$\begin{aligned}
J_n &= \frac{1}{n}\sum_{j=1}^{n}\Big(\hat{m}_n^j(x_{j+1}) - m(x_{j+1})\Big)\varepsilon_j \\
&= \frac{1}{n}\sum_{j=1}^{n}\Big(\frac{1}{nh_n}\sum_{i\neq j}^{n} K\left(\frac{j+1-i}{nh_n}\right) Y_i - m(x_{j+1})\Big)\varepsilon_j \\
&= \frac{1}{n^2h_n}\sum_{j=1}^{n}\sum_{i\neq j}^{n} K\left(\frac{j+1-i}{nh_n}\right)\varepsilon_i\,\varepsilon_j + \frac{1}{n}\sum_{j=1}^{n}\left[\frac{1}{nh_n}\sum_{i\neq j}^{n} K\left(\frac{j+1-i}{nh_n}\right) m(x_i) - m(x_{j+1})\right]\varepsilon_j \\
&= J_{n,1} + J_{n,2}.
\end{aligned}$$

Following the sketch of the proof of Lemma 1, it is easily seen that for $\epsilon > 0$, for an arbitrary great integer k and for $i = 1, 2$,

$$\mathbb{P}\left(\sup_{h_n\in H_n}\frac{|J_{n,i}|}{d_M(\hat{m}_n, m)} > \epsilon\right) \leq C\, n^{\rho}\,\epsilon^{-2k}\,(2\,\sigma^2\,||K||_2^2)^{-2k}\sup_{h_n\in H_n}\left((nh_n)^{-2k}\mathbb{E}|J_{n,i}|^{2k}\right). \quad (17)$$

Now, applying Lemma 3, we get

$$\mathbb{E}\,(J_{n,1}^{2k}) \leq 2^{6k}\,C_{2k}\,(C_{4k}\,M_{4k})^{1/2}\,n^{-4k}h_n^{-2k}\left(\sum_{j=1}^{n}\sum_{i\neq j}K^2\left(\frac{j+1-i}{nh_n}\right)\right)^{k}, \quad (18)$$

$$\mathbb{E}\,(J_{n,2}^{2k}) \leq 2^{2k}\,C_{2k}\,M_{2k}\,n^{-2k}\left(\sum_{j=1}^{n}\left[\frac{1}{nh_n}\sum_{i\neq j}^{n}K\left(\frac{j+1-i}{nh_n}\right)m(x_i) - m(x_{j+1})\right]^2\right)^{k} \quad (19)$$

where C_{2k} and C_{4k} are defined in Lemma 3.

— *Upper Bound for* $\mathbb{E}\,(J_{n,1}^{2k})$

$$\sum_{j=1}^{n}\sum_{i\neq j} K^2\left(\frac{j+1-i}{nh_n}\right) \;\leq\; \sum_{j=1}^{n}\sum_{\ell=\ell_0(j)}^{\ell_1(j)} K^2\left(\frac{j+1-\ell}{nh_n}\right)$$
$$\leq\; \sum_{j=1}^{n}(\ell_1(j)-\ell_0(j)+1)\,||K||_\infty^2$$

Now, define $\ell_0(i)$ (resp. $\ell_1(i)$) the smallest integer ℓ such that $\frac{\ell}{n} \geq \frac{i}{n} - \frac{h_n}{2}$ (respectively the biggest integer ℓ such that $\frac{\ell}{n} \leq \frac{i}{n} + \frac{h_n}{2}$), with both $\ell_0(j)$ and $\ell_1(j)$. Since $nh_n \to +\infty$, for n large enough we have already seen that $\ell_1(j)-\ell_0(j)+1 \leq 2\,nh_n$

$$\sum_{j=1}^{n}\sum_{i\neq j} K^2\left(\frac{j+1-i}{nh_n}\right) \;\leq\; 2\,n^2 h_n ||K||_\infty^2$$

and finally, we deduce that

$$\mathbb{E}\,(J_{n,1}^{2k}) \;\leq\; 2^{7k}\,C_{2k}\,\left(C_{4k}\,M_{4k}\right)^{1/2}\,||K||_\infty^{2k}\,n^{-2k}h_n^{-k} \tag{20}$$

— *Upper Bound for* $\mathbb{E}\,(J_{n,2}^{2k})$ in (19) :

$$\frac{1}{nh_n}\sum_{i\neq j}^{n} K\left(\frac{j+1-i}{nh_n}\right)\,m(x_i) - m(x_{j+1}) \;=\; B_n(x_{j+1}) - \frac{1}{nh_n}K\left(\frac{1}{nh_n}\right)\,m(x_j)$$

and thanks to the point $i)$ in Lemma 6, there exists $\mathcal{C}>0$ such that for n large enough,

$$\left|\frac{1}{nh_n}\sum_{i\neq j}^{n} K\left(\frac{j+1-i}{nh_n}\right)\,m(x_i) - m(x_{j+1})\right| \;\leq\; \mathcal{C}\left(h_n^2 + (nh_n)^{-1}\right).$$

Thus, we obtain for $\mathbb{E}\,(J_{n,2}^{2k})$ and n large enough

$$\mathbb{E}\,(J_{n,2}^{2k}) \;\leq\; 2^{3k}\,C_{2k}\,M_{2k}\,\mathcal{C}^{2k}\,n^{-k}\left(h_n^4 + (nh_n)^{-2}\right)^k. \tag{21}$$

It is easily seen from (17), (20) and (21) that the complete convergence occures as soon as k will be choosen as $k > 5(1+\rho)$.

$\square$

4 Simulation study

A Monte-Carlo study is undertaken to investigate the practical performance of the methodology presented in the previous section. Let y_1, ..., y_n be the observations and $x_i = i/n$ for $i = 1, \ldots, n$ the design. The error process $(\varepsilon_n)_{n\geq 0}$ is supposed to be independently and identically distributed $\mathcal{N}(0, \sigma)$. We consider the sample size $n = 100$ as well as two different values for the standard deviation σ of the error process $(\varepsilon_n)_{n\geq 0}$ respectively : $\sigma = 0.1$ and $\sigma = 0.2$. For each model, 30 independent sets of data are generated.

The different regression curves considered are defined by the following models :

$$\begin{aligned}
\textit{Model A : } m(x) &= 16\, x^2(1-x)^2 \\
\textit{Model B : } m(x) &= 0.4\, \cos(5\pi x) \\
\textit{Model C : } m(x) &= 0.4\, \sin(8\pi x)
\end{aligned}$$

To improve the flexibility of the criterion in an exploratory point of view, different behaviour of the regression curve are choosen. Model A admits one turning point, while models B and C which are sinusoidal curves present a lot of oscillations (4 and 8 turning points respectively).

Recall that the automatic bandwidth $\hat{h}_n$ is obtained by minimizing the correlation criterion defined by (3) and that the asymptotic optimal bandwith $h_n^\star$ minimizes the MISE. We choose the Epanechnikov kernel (1969) to compute the estimate, defined by

$$K(t) = \frac{3}{4}(1-t^2)1_{[-1,1]}(t).$$

The absolute relative error $ARE = |d_A(m, \hat{m}_{\hat{h}_n}) - d_A(m, m_{h^\star})|/d_A(m, \hat{m}_{\hat{h}_n})$ is computed as a measure of accuracy of the automatic estimate. The difference between the number of turning points $\hat{N}$ in the automatic estimate $\hat{m}_{\hat{h}_n}$ and the true number of turning points N_0 in the true curve is evaluated via the quantity denoted by TP :

$$TP = \hat{N} - N_0$$

where $\hat{N}$ is the number of changes of signs of the difference $\hat{m}_{\hat{h}_n}(x_i) - \hat{m}_{\hat{h}_n}(x_{i-1})$ for $i = 2, \ldots, n$. Tables 1 - 3 give the asymptotic optimal bandwith $h_n^\star$, the mean, the median and the standard deviation of 30 automatic bandwidths $\hat{h}_n$ obtained by minimization of the correlation criterion for 30 independent sets of generated data. The corresponding values for ARE and TP are also given.

The automatic bandwidth $\hat{h}_n$ does not balance between bias and variance terms as it is usual for the theoretical optimal bandwidth $h^\star$ and consecutively for the cross-validated bandwidth. As an exploratory tool, balance between bias and variance terms is not necessary the best way to obtain an automatic bandwidth selector that detects the main feature of the unknown regression curve. If the regression curve has a lot of turning points, as models C and D, it would be appropriate to undersmooth the data and to have a dominant variance term in order to catch the bulks of the data.

On the other hand, if there are few change of monotony in the regression curve, it would be more adequate to have a dominant bias term. These facts are illustrated by simulations. we can see that for model A, the automatic bandwidth $\hat{h}_n$ is larger than the theoretical optimal bandwidth $h^\star$ which effectively makes balance between bias and variance. For Model B with exactly 4 turning points, the same phenomenom exists but is less pronounced than with Model A. The difference between the automatic bandwidth and the optimal bandwidth (regard to balance bias vs. variance) is not significative. But, when model C is undertaken with 8 turning points,

$\sigma = 0.1$	mean	median	standard deviation
$h_n^\star$	0.068	—	—
$\hat{h}_n$	0.135	0.135	0.006
ARE	0.343	0.337	0.240
TP	0	0	0
$\sigma = 0.2$	mean	median	standard deviation
$h_n^\star$	0.090	—	—
$\hat{h}_n$	0.143	0.146	0.006
ARE	0.182	0.133	0.165
TP	0.067	0	0.365

Table 1: *Model A — Monte-Carlo study for 30 independent data sets of size* $n = 100$.

$\sigma = 0.1$	mean	median	standard deviation
$h_n^\star$	0.036	—	—
$\hat{h}_n$	0.044	0.045	0.003
ARE	0.051	0.040	0.042
TP	2.267	2	0.691
$\sigma = 0.2$	mean	median	standard deviation
$h_n^\star$	0.048	—	—
$\hat{h}_n$	0.049	0.045	0.006
ARE	0.050	0.046	0.039
TP	4.27	4	2.392

Table 2: *Model B — Monte-Carlo study for 30 independent data sets of size* $n = 100$.

$\sigma = 0.1$	mean	median	standard deviation
$h_n^\star$	0.036	—	—
$\hat{h}_n$	0.033	0.035	0.004
ARE	0.074	0.014	0.138
TP	0.4	0	0.076
$\sigma = 0.2$	mean	median	standard deviation
$h_n^\star$	0.048	—	—
$\hat{h}_n$	0.037	0.035	0.006
ARE	0.122	0.084	0.133
TP	2.1	2	2.006

Table 3: *Model C — Monte-Carlo study for 30 independent data sets of size* $n = 100$.

the variance term need to be the dominant term. As a consequence, the automatic bandwidth is smaller than the asymptotic optimal bandwidth.

All these considerations need also to be related to the perturbation introduced by the error process (ε_t) in the model. More important is the variance of the noise, more difficult it is to detect if the variance of the automatic estimate is due to an effective change of monotony in the regression curve. For fixed sample size, the increasing of the variance noise would degrade the identification of the number of turning points as it happens with models having a larger variance noise.

5 Appendix A : Technical Lemmas

The following Lemma is a Whittle's result (1960) and provide bounds for the absolute s-th moments of quadratic and linear forms in independent statistical variables.

Lemma 3 *Assume that $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are independent variables and have zero mean values. Let's denote $\mu_j(s) = \left(\mathbb{E}\,|\varepsilon_j|^s\right)^{1/s}$ provided these quantities exist.*

For any set of real coefficients $\{a_{ij}\}$, $\{b_j\}$, if $B = \sum_{j=1}^n b_j\,\varepsilon_j$, $A = \sum_{i=1}^n \sum_{j=1}^n a_{ij}\,\varepsilon_i\varepsilon_j$ and if $s \geq 2$ then,

$$\mathbb{E}\,|B|^s \leq 2^s\, C_s \left(\sum_{j=1}^n b_j^2\, \mu_j(s)^2\right)^{s/2} \tag{22}$$

$$\mathbb{E}\,|A - \mathbb{E}(A)|^s \leq 2^{3s}\, C_s\, \sqrt{C_{2s}} \left(\sum_{j=1}^n \sum_{k=1}^n a_{jk}^2\, \mu_j(2s)^2\right)^{s/2} \tag{23}$$

with $C_s = \dfrac{2^{s/2}}{\sqrt{\pi}}\Gamma\left(\dfrac{s+1}{2}\right)$.

Lemma 4 *If the kernel function K satisfies assumption $(\mathcal{A}_3)$,*

$$\sup_{x\in[0,1]}\left|\frac{1}{nh_n}\sum_{j=1}^n K\left(\frac{j/n - x}{h_n}\right) - 1\right| \leq 2\,\frac{C(K)}{nh_n}.$$

Proof : As K integrates to 1, by the change of variable $z = \frac{t-x}{h_n}$, we get

$$\int_{-1/2}^{1/2} K(z)\,dz = \int_{x-h_n/2}^{x+h_n/2} \frac{1}{h_n} K\left(\frac{t-x}{h_n}\right)\,dt = 1.$$

Now, with $[0,1] = \cup_{j=1}^n [\dfrac{j-1}{n}, \dfrac{j}{n}]$, let $j_0(x)$ be the smallest integer j such that $x - h_n/2 \leq \frac{j}{n}$ and respectively $j_1(x)$ be the biggest integer j such that $x + h_n/2 \geq \frac{j}{n}$, notice that $j_1(x) - j_0(x) \leq nh_n$, then

$$\int_{x-h_n/2}^{x+h_n/2} \frac{1}{h_n} K\left(\frac{t-x}{h_n}\right)\,dt = \sum_{j=j_0(x)}^{j_1(x)} \int_{(j-1)/n}^{j/n} \frac{1}{h_n} K\left(\frac{t-x}{h_n}\right)\,dt.$$

If $j \leq j_0(x) - 1$, then $\frac{j/n - x}{h_n} < -\frac{1}{2}$ and if $j \geq j_1(x) + 1$ then $\frac{j/n-x}{h_n} > \frac{1}{2}$, and since K is supported on $[-1/2, 1/2]$ we can observe with the same argument that

$$\frac{1}{nh_n}\sum_{j=1}^{n} K\left(\frac{j/n-x}{h_n}\right) = \frac{1}{nh_n}\sum_{j=j_0(x)}^{j_1(x)} K\left(\frac{j/n-x}{h_n}\right) = \sum_{j=j_0(x)}^{j_1(x)} \int_{(j-1)/n}^{j/n} \frac{1}{h_n} K\left(\frac{j/n-x}{h_n}\right)\, dt.$$

Consequently,

$$\begin{aligned}
&\left|\frac{1}{nh_n}\sum_{j=1}^{n} K\left(\frac{j/n-x}{h_n}\right) - \int_{x-h_n/2}^{x+h_n/2} \frac{1}{h_n} K\left(\frac{t-x}{h_n}\right) dt\right| \\
&\quad = \left|\sum_{j=j_0(x)}^{j_1(x)} \int_{(j-1)/n}^{j/n} \frac{1}{h_n}\left[K\left(\frac{j/n-x}{h_n}\right) - K\left(\frac{t-x}{h_n}\right)\right] dt\right| \\
&\quad \leq \frac{1}{h_n}\sum_{j=j_0(x)}^{j_1(x)} \int_{(j-1)/n}^{j/n} C(K)\left|\frac{j/n-t}{h_n}\right| dt \\
&\quad \leq \frac{C(K)}{n^2h_n^2}\left(j_1(x)-j_0(x)+1\right)
\end{aligned}$$

but, thanks to the definition of both $j_0(x)$ and $j_1(x)$, and for n large enough, it is clear that

$$j_1(x) - j_0(x) + 1 \leq nh_n + 1 \leq 2\, nh_n$$

so that finally, for all x

$$\left|\frac{1}{nh_n}\sum_{j=1}^{n} K\left(\frac{j/n-x}{h_n}\right) - 1\right| \;\leq\; 2\,\frac{C(K)}{nh_n}.$$

□

Lemma 5 *If the kernel function K satisfies assumption $(\mathcal{A}_3)$,*

$$\left|\sum_{j=1}^{n} K\left(\frac{i-j}{nh_n}\right)\left[K\left(\frac{i-j}{nh_n}\right) - K\left(\frac{i-j-1}{nh_n}\right)\right]\right| \;\leq\; C(K)\left(1+\frac{2\,C(K)}{nh_n}\right)$$

Proof : By assumption (A_3), we have for $i \in \{1, \ldots, n\}$

$$\begin{aligned}
\left| \sum_{j=1}^{n} K\left(\frac{i-j}{nh_n}\right) \left[K\left(\frac{i-j}{nh_n}\right) - K\left(\frac{i-j-1}{nh_n}\right) \right] \right| &\le \frac{C(K)}{nh_n} \left| \sum_{j=1}^{n} K\left(\frac{i-j}{nh_n}\right) \right| \\
&\le C(K) \left(1 + \sup_x \left| \frac{1}{nh_n} \sum_{j=1}^{n} K\left(\frac{x-j/n}{h_n}\right) - 1 \right| \right) \\
&\le C(K) \left(1 + \frac{2\,C(K)}{nh_n} \right)
\end{aligned}$$

where the last inequality follows directly from Lemma 4.

□

Lemma 6 *i — For $x \in [0,1]$, the ponctual bias $B_n(x)$ is*

$$B_n(x) = \frac{h_n^2}{2} \sigma^2(K)\, m''(x) + o(h_n^2) + O\left(\frac{1}{nh_n}\right)$$

ii — For $x, y \in [0,1]$ such that $y = x + 1/n$

$$|B_n(y) - B_n(x)| \le 3\, C(K)\, ||m||_\infty \frac{1}{nh_n} + C(m)\, n^{-1}$$

where $\sigma^2(K)$ is defined in ($\mathcal{A}_3$).

Proof :

Proof of i)

$$\begin{aligned}
B_n(x) &= \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x-i/n}{h_n}\right) m(\frac{i}{n}) - m(x) \\
&= \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x-i/n}{h_n}\right) m(\frac{i}{n}) - \int_0^1 \frac{1}{h_n} K\left(\frac{x-t}{h_n}\right) m(t)\, dt \\
&+ \int_0^1 \frac{1}{h_n} K\left(\frac{x-t}{h_n}\right) m(t)\, dt - m(x)
\end{aligned}$$

Now, it is wellknown that the ponctual bias of the kernel estimate with $K(\cdot)$ a symetric kernel of order 2 is given by,

$$\int_0^1 \frac{1}{h_n} K\left(\frac{x-u}{h_n}\right) m(u)\, du - m(x) = \frac{h_n^2}{2} m^{''}(x) \int v^2 K(v)\, dv + o(h_n^2)$$

Now, since K is supported on [-1/2, 1/2] and with $j_0(x)$ and $j_1(x)$ defined as in Lemma 4

$$\begin{aligned}
& \left| \sum_{i=1}^{n} \int_{(i-1)/n}^{i/n} \frac{1}{h_n} K\left(\frac{x-i/n}{h_n}\right) m(\frac{i}{n}) - \int_0^1 \frac{1}{h_n} K\left(\frac{x-t}{h_n}\right) m(t)\, dt \right| \\
= \ & \left| \frac{1}{h_n} \sum_{i=j_0(x)}^{j_1(x)} \int_{(i-1)/n}^{i/n} \left[K\left(\frac{x-i/n}{h_n}\right) m(\frac{i}{n}) - K\left(\frac{x-t}{h_n}\right) m(t) \right] dt \right| \\
\leq \ & \frac{1}{h_n} \sum_{i=j_0(x)}^{j_1(x)} \int_{(i-1)/n}^{i/n} \left| K\left(\frac{x-i/n}{h_n}\right) \right| \left| m(\frac{i}{n}) - m(t) \right| + \left| K\left(\frac{x-i/n}{h_n}\right) - K\left(\frac{x-t}{h_n}\right) \right| |m(t)|\, dt \\
\leq \ & \frac{1}{h_n} \sum_{i=j_0(x)}^{j_1(x)} \int_{(i-1)/n}^{i/n} \left| K\left(\frac{x-i/n}{h_n}\right) \right| C(m)\, |i/n - t| + C(K) \left| \frac{i/n-t}{h_n} \right| |m(t)|\, dt \\
\leq \ & \frac{j_1(x) - j_0(x) + 1}{h_n} \left(||K||_\infty \frac{C(m)}{2} n^{-2} + ||m||_\infty \frac{C(K)}{2} n^{-2} h_n^{-1} \right) \\
\leq \ & 2n \left(||K||_\infty \frac{C(m)}{2} n^{-2} + ||m||_\infty \frac{C(K)}{2} n^{-2} h_n^{-1} \right) \\
\leq \ & ||K||_\infty C(m)\, n^{-1} + ||m||_\infty C(K)\, (nh_n)^{-1}.
\end{aligned}$$

Proof of ii) : Let $y = x + \frac{1}{n}$

$$\begin{aligned}
& |B_n(y) - B_n(x)| \\
& = \left| \frac{1}{nh_n} \sum_{j=1}^{n} K\left(\frac{y-j/n}{h_n}\right) m(\frac{j}{n}) - m(y) - \frac{1}{nh_n} \sum_{j=1}^{n} K\left(\frac{x-j/n}{h_n}\right) m(\frac{j}{n}) + m(x) \right| \\
& = \left| \frac{1}{nh_n} \sum_{j=j_0(y)}^{j_1(y)} K\left(\frac{y-j/n}{h_n}\right) m(\frac{j}{n}) - m(y) - \frac{1}{nh_n} \sum_{j=j_0(x)}^{j_1(x)} K\left(\frac{x-j/n}{h_n}\right) m(\frac{j}{n}) + m(x) \right| \\
& \leq \left| \frac{1}{nh_n} \sum_{j=j_0(x)\wedge j_0(y)}^{j_1(x)\vee j_1(y)} \left[K\left(\frac{y-j/n}{h_n}\right) - K\left(\frac{x-j/n}{h_n}\right) \right] m(\frac{j}{n}) \right| + |m(y) - m(x)|
\end{aligned}$$

where $j_0(x), j_1(x), j_0(y)$ and $j_1(y)$ are defined as in Lemma 4. Now applying assumptions $(\mathcal{A}_3)$ on K and $(\mathcal{A}_4)$ on m give finally,

$$|B_n(y) - B_n(x)| \leq \frac{1}{nh_n}\left(j_1(x) \vee j_1(y) - j_0(x) \wedge j_0(y) + 1\right) C(K)\, ||m||_\infty \left|\frac{y-x}{h_n}\right| + C(m)\, |y-x|$$

and by definition of $j_0(x), j_1(x), j_0(y)$ and $j_1(y)$, it is easy to see,

$$j_1(x) \vee j_1(y) - j_0(x) \wedge j_0(y) + 1 = j_1(y) - j_0(x) + 1 \leq nh_n + 2$$

thus, with $y - x = 1/n$, we deduce

$$\begin{aligned} |B_n(y) - B_n(x)| &\leq C(K)\, ||m||_\infty \frac{1}{n^2 h_n^2} (nh_n + 2) + C(m)\, n^{-1} \\ &\leq 3\, C(K)\, ||m||_\infty (nh_n)^{-1} + C(m)\, n^{-1}. \end{aligned}$$

□

Acknowledgements : Authors are very grateful to Pr. Alain Berlinet for having first initiated this work.

References

ALTMAN, N. S., 1990. Kernel smoothing of data with correlated errors. *Journal of the American Statistical Association,* 85, 749-758.

ANDERSON, T. W., 1971, The statistical analysis of time series. *New York : Wiley.*

BOENTE, G. , FRAIMAN, R. , MELOCHE, J., 1997, Robust plug-in bandwidth estimators in nonparametric regression. *Journal of Statistical Planning and Inference,* 57, 109-142.

BOSQ, D. , 1999, Nonparametric statistics for stochastic processes. *Lecture Notes in Statistics* 110, Springer.

BOX, G.E.P. , PIERCE, D.A. , 1970, Distribution of residual autocorrelations in autoregressive-integrated-moving-average time series models. *Journal of the American Statistical Association,* 65, 1509-1526.

Braun, W.J. , Rousson, V. , 2000, An autocorrelation criterion for bandwidth selection in nonparametric regression. *Journal of Statistical Computation and Simulation,* 68, 89-101.

Chiu, S.T. , 1990, Why do bandwidth selectors tend to choose smaller bandwidths and a remedy. *Biometrika,* 77, 222-226.

Epanechnikov, V. , 1969, Nonparametric estimates of a multivariate probability density. *Theory Probab. Appl.* 14, 153-158.

Fan, J. , Gijbels, I. , 1996, Local polynomial modelling and its applications. *Chapman and Hall, London.*

Härdle, W. , Marron, J. S. , 1985, Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Statist.*, 13, 1465-1481.

Härdle, W. , Hall, P. , Marron, J. S. , 1988, How far are automatically chosen regression smoothing parameters from their optimum ? *Journal of the American Statistical Association*, 82-101.

Priestley, M.B. , Chao, M.T. , 1972, Nonparametric function fitting. *Journal of the Royal Statistical Society, Series B*, 34, 385-392.

Rice, J. , 1984, Bandwidth choice for nonparametric regression. *Ann. Statist.*, 12, 1215-1230.

Ruppert, D. , Sheather, S.J. , Wand, M.P. , 1995, An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.*, 90, 1257-1270.

Shibata, R. , 1981, An optimal selection of regression variables. *Biometrika.* 68, 45-54.

Shorack, G.R. , Wellner, J.A. , 1986, Empirical processes with applications to statistics, *Wiley.*

Whittle, P. , 1960, Bounds for the moments of linear and quadratic forms in independent variables. *Th. Probab. Appl.*, 5, n^o 3, 302-305.