ESAIM: PS 18 (2014) 881–899
DOI: 10.1051/ps/2014009

ESAIM: Probability and Statistics
www.esaim-ps.org

LOCAL POLYNOMIAL ESTIMATION OF THE MEAN FUNCTION AND ITS DERIVATIVES BASED ON FUNCTIONAL DATA AND REGULAR DESIGNS

Karim Benhenni[1] and David Degras[2]

Abstract. We study the estimation of the mean function of a continuous-time stochastic process and its derivatives. The covariance function of the process is assumed to be nonparametric and to satisfy mild smoothness conditions. Assuming that n independent realizations of the process are observed at a sampling design of size N generated by a positive density, we derive the asymptotic bias and variance of the local polynomial estimator as n, N increase to infinity. We deduce optimal sampling densities, optimal bandwidths, and propose a new plug-in bandwidth selection method. We establish the asymptotic performance of the plug-in bandwidth estimator and we compare, in a simulation study, its performance for finite sizes n, N to the cross-validation and the optimal bandwidths. A software implementation of the plug-in method is available in the R environment.

Mathematics Subject Classification. 62G08, 62G20.

Received December 20, 2012. Revised March 31, 2014.

1. Introduction

Local polynomial smoothing is a popular method for nonparametric regression. In addition to its fast implementation [10], it enjoys good statistical properties such as design adaptation, optimal convergence rate, and minimax efficiency [8]. Classical examples of local polynomial estimators of order zero are the Nadaraya–Watson and Gasser–Müller kernel estimators. These estimators, however, do not share all the features of higher order local polynomials described above.

There is a vast literature on local polynomial regression under independent measurement errors. Asymptotic bias and variance expressions can be found in [8, 27]. Such results offer key qualitative insights into how the bandwidth affects the estimation. They are also useful in the key problem of selecting the bandwidth. Indeed, they give access to the optimal theoretical bandwidth which in turn can be exploited for data-driven bandwidth selection [9, 25, 26].

In the case of correlated errors, reference [19] give an extensive review of the available asymptotic theory and smoothing parameter selection methods. Local polynomial estimation is studied under mixing conditions in [18], under association in [17], and under (stationary) short-range dependence in [13, 20]. Reference [14]

Keywords and phrases. Local polynomial smoothing, derivative estimation, functional data, sampling density, plug-in bandwidth.

[1] Laboratoire LJK UMR CNRS 5224, Université de Grenoble, 38040 Grenoble, France. karim.benhenni@upmf-grenoble.fr

[2] Department of Mathematical Sciences, DePaul University, Chicago 60614, Illinois, USA. ddegrasv@depaul.edu

Article published by EDP Sciences

develop bootstrap and cross-validation methods to select the bandwidth under short and long-range dependence, while [12] propose a plug-in method for short-range dependent errors.

In the above references, data are essentially modeled as smooth mean functions plus discrete time series errors. In many applications however, data are more adequately represented as observations of continuous-time processes or random functions. This continuous-time framework, which is characteristic of *functional data analysis* [23], it markedly differs from discrete-time models with respect to statistical estimation properties. In particular, consistent estimation of the mean function with functional data entails that the number of observed curves goes to infinity. In addition, the estimation variance primarily depends on the number of curves; it only depends on the size of the observation grid at the second order [6,15]. Studying kernel regression with continuous-time, stationary error processes, reference [15] derive the asymptotic bias and variance of the Gasser–Müller estimator; they select the bandwidth by optimizing an estimate of the integrated mean squared error based on the empirical autocovariance. Reference [11] extend this work to nonstationary error processes with parametric covariance. Reference [2] obtain bias and variance expressions under nonstationary error processes with regular covariance. Reference [3] extend their results to quantized (noisy) observations. Examining smoothing splines in the context of functional data, reference [24] propose a cross-validation method adjusted to functional data that leaves one curve out at the time; the optimality of this method is established in [16]. Limiting distributions of local polynomial regression estimators for longitudinal or functional data can be found in [28]. Reference [7] devises simultaneous confidence bands for local linear estimators based on a functional central limit theorem. However, no result seems available in the functional data setting for the local estimation of *regression derivatives* and with general (not necessarily stationary) autocorrelated processes.

In this paper we consider the usual functional data framework where for each of n statistical units, a curve is observed at the same N sampling points generated by a regular density function. The data-generating process is the sum of a mean function m and a general continuous-time error process. We are interested in the local polynomial estimation of m and its derivatives. We derive second-order asymptotic expressions for the bias and variance of the estimators. From these expressions we deduce optimal sampling densities (see *e.g.* [1,4] for related examples), optimal bandwidths, and asymptotic normality results. Applying these results to bandwidth selection, we develop a plug-in method for the estimation of m and m' and study its convergence properties. We also conduct extensive simulations to compare the performances of local polynomial regression based on different orders of fit and different bandwidths (optimal, plug-in, and cross-validation).

The rest of the paper is organized as follows. The statistical model and local polynomial estimators are defined in Section 2. Theoretical results on the estimation bias and variance are exposed in Section 3. The plug-in method is developed in Section 4. Section 5 presents a simulation study. A discussion is provided in Section 6. Proofs are deferred to the Appendix.

2. Local polynomial regression

We consider the statistical problem of estimating a mean function and its derivatives in a fixed design experiment. Assume that for each of n experimental units, N measurements of the response are available on a regular grid:

$$Y_i(x_j) = m(x_j) + \varepsilon_i(x_j), \qquad i = 1, \dots, n, \quad j = 1, \dots, N, \tag{2.1}$$

where m is the unknown mean function and the ε_i are i.i.d. error processes with mean zero and autocovariance function ρ. The observation points x_j are taken to be regularly spaced quantiles of a continuous positive density f on $[0, 1]$:

$$\int_0^{x_j} f(x)\mathrm{d}x = \frac{j-1}{N-1}, \quad j = 1, \dots, N. \tag{2.2}$$

Note that the uniform density $f \equiv 1$ corresponds to an equidistant design.

Assume that m is at least p times differentiable on $[0, 1]$. Write $\beta_k(x) = m^{(k)}(x)/k!$, $\boldsymbol{\beta}(x) = (\beta_0(x), \dots, \beta_p(x))'$, and $\overline{Y}_{\cdot j} = \frac{1}{n}\sum_{i=1}^n Y_i(x_j)$. Let $0 \le \nu \le p$ be an integer. For each $x \in [0, 1]$, the local polynomial estimator of

$m^{(\nu)}(x)$ of order p is defined as $\widehat{m}_{\nu,p}(x) = \nu!\,\widehat{\beta}_\nu(x)$, where $\widehat{\boldsymbol{\beta}}_N(x) = \big(\widehat{\beta}_0(x), \ldots, \widehat{\beta}_p(x)\big)'$ is the solution to the minimization problem

$$\min_{\boldsymbol{\beta}(x)} \sum_{j=1}^{N} \left(\overline{Y}_{\cdot j} - \sum_{k=0}^{p} \beta_k(x)\,(x_j - x)^k \right)^2 K\left(\frac{x_j - x}{h}\right). \tag{2.3}$$

In (2.3), h denotes a positive bandwidth and K is a kernel function. Let $\overline{\mathbf{Y}}_N = (\overline{Y}_{\cdot 1}, \ldots, \overline{Y}_{\cdot N})'$ and denote the canonical basis of $\mathbb{R}^{p+1}$ by $(\mathbf{e}_k)_{k=0,\ldots,p}$ ($\mathbf{e}_k$ has a 1 in the $(k+1)$th position and 0 elsewhere). Finally define the matrix

$$\mathbf{X}_N = \begin{pmatrix} 1 & (x_1 - x) & \cdots & (x_1 - x)^p \\ \vdots & \vdots & & \vdots \\ 1 & (x_N - x) & \cdots & (x_N - x)^p \end{pmatrix}$$

and $\mathbf{W}_N = \operatorname{diag}\big((1/Nh)K\big((x_j - x)/h\big)\big)$. The estimator $\widehat{m}_{\nu,p}(x)$ expresses as

$$\widehat{m}_{\nu,p}(x) = \nu!\,\mathbf{e}_\nu'\,\widehat{\boldsymbol{\beta}}_N(x), \quad \text{with } \widehat{\boldsymbol{\beta}}_N(x) = (\mathbf{X}_N'\mathbf{W}_N\mathbf{X}_N)^{-1}\mathbf{X}_N'\mathbf{W}_N\overline{\mathbf{Y}}_N. \tag{2.4}$$

3. Asymptotic study

3.1. Asymptotic bias and variance

We make the following hypotheses for the asymptotic study:

- (H1) The kernel K is a Lipschitz-continuous, symmetric density function with compact support.
- (H2) The bandwidth $h = h(n, N)$ satisfies $h \to 0$, $Nh^2 \to \infty$, and $nh^{2\nu} \to \infty$ as $n, N \to \infty$.
- (H3) The regression function m is $(p+2)$ times continuously differentiable on $[0, 1]$.
- (H4) The sampling density f is continuously differentiable on $[0, 1]$.
- (H5) The covariance function ρ is continuous on the unit square $[0, 1]^2$ and has continuous first-order partial derivatives outside the main diagonal. These derivatives have left and right limits on the main diagonal determined by $\rho^{(0,1)}(x, x^-) = \lim_{y \nearrow x} \rho^{(0,1)}(x, y)$ and $\rho^{(0,1)}(x, x^+) = \lim_{y \searrow x} \rho^{(0,1)}(x, y)$.

The kernel condition (H1) and regularity conditions (H3) and (H4) are usual in nonparametric regression. (H2) ensure that the bandwidth h goes to zero slowly enough. More precisely, the condition $nh^{2\nu} \to \infty$ guarantees that the variance of $\widehat{m}_{\nu,p}(x)$ goes to zero as $n, N \to \infty$ and the condition $Nh^2 \to \infty$ is required to obtain second-order bias and variance expansions. (H5) is a mild smoothness assumption satisfied by many processes, for example Wiener and Ornstein−Uhlenbeck processes where ρ admits uniformly bounded second partial derivatives outside the main diagonal.

Define the vectors $\mathbf{c} = (\mu_{p+1}, \ldots, \mu_{2p+1})'$ and $\tilde{\mathbf{c}} = (\mu_{p+2}, \ldots, \mu_{2p+2})'$, where $\mu_k = \int_{-\infty}^{+\infty} u^k K(u)\mathrm{d}u$ denotes the kth moment of K. Let $\mathbf{S} = (\mu_{k+l})$, $\tilde{\mathbf{S}} = (\mu_{k+l+1})$, $\mathbf{S}^* = (\mu_k \mu_l)$, and $\mathbf{A} = \big(\frac{1}{2}\iint_{\mathbb{R}^2} |u - v| u^k v^l K(u)K(v)\mathrm{d}u\mathrm{d}v\big)$ be matrices of size $(p+1) \times (p+1)$ with elements indexed by $k, l = 0, \ldots, p$. For $x \in (0, 1)$, define the jump function

$$\alpha(x) = \rho^{(0,1)}(x, x^-) - \rho^{(0,1)}(x, x^+)\,.$$

The asymptotic bias and variance of the estimator $\widehat{m}_{\nu,p}(x)$ are established hereafter.

Theorem 3.1. *Assume* (H1)–(H5). *Then as* $n, N \to \infty$,

$$\mathbb{E}\big(\widehat{m}_{\nu,p}(x)\big) - m^{(\nu)}(x) = \frac{\nu!\, m^{(p+1)}(x)}{(p+1)!}\left(\mathbf{e}_\nu' \mathbf{S}^{-1}\mathbf{c}\right) h^{p+1-\nu} + o(h^{p+2-\nu})$$
$$+ \nu! \left\{ \frac{m^{(p+2)}(x)}{(p+2)!} \mathbf{e}_\nu' \mathbf{S}^{-1} \tilde{\mathbf{c}} + \frac{m^{(p+1)}(x)}{(p+1)!} \frac{f'(x)}{f(x)} \left(\mathbf{e}_\nu' \mathbf{S}^{-1} \tilde{\mathbf{c}} - \mathbf{e}_\nu' \mathbf{S}^{-1} \tilde{\mathbf{S}} \mathbf{S}^{-1} \mathbf{c} \right) \right\} h^{p+2-\nu}$$

and

$$\operatorname{Var}\big(\widehat{m}_{\nu,p}(x)\big) = \frac{(\nu!)^2 \rho(x,x)}{nh^{2\nu}}\, \mathbf{e}_\nu' \mathbf{S}^{-1} \mathbf{S}^* \mathbf{S}^{-1} \mathbf{e}_\nu - \frac{(\nu!)^2\, \alpha(x)}{nh^{2\nu-1}}\, \mathbf{e}_\nu' \mathbf{S}^{-1} \mathbf{A} \mathbf{S}^{-1} \mathbf{e}_\nu + o\left(\frac{1}{nh^{2\nu-1}}\right).$$

Remark 3.2. The asymptotic bias in Theorem 3.1 does not depend on the stochastic structure of the data (here, continuous-time process). A similar bias expansion (but only at the first order) can be found in [8] in the context of independent errors. Also, the asymptotic variance in Theorem 3.1 extends the results of [3] and [15] on the kernel estimation of m.

Remark 3.3. The reason for presenting second-order expansions in Theorem 3.1 is that first-order terms may vanish due to the symmetry of K which causes its odd moments to be zero. For instance, the first-order term in the bias vanishes if $(p-\nu)$ is even and the first-order term in the variance vanishes if ν is odd. In both cases the second-order terms generally allow to find exact rates of convergence and asymptotic optimal bandwidths.

If the covariance ρ has continuous first derivatives at (x,x), the second-order variance term in Theorem 3.1 vanishes since $\rho^{(0,1)}(x,x^-) = \rho^{(0,1)}(x,x^+)$. In this case the variance expansion does not depend on h if $\nu = 0$ or ν is odd (see Rem. 3.3). This makes it impossible to assess $\operatorname{Var}(\widehat{m}_{\nu,p}(x))$ with Theorem 3.1 alone. This issue can be solved by deriving higher-order variance expansions under stronger differentiability assumptions on f and ρ. For simplicity we restrict our attention to the case of an equidistant sampling grid. Define the matrices $\mathbf{A}_1 = \left(\frac{1}{2}(\mu_k\mu_{l+2} + \mu_{k+2}\mu_l)\right)$, $\mathbf{A}_2 = (\mu_{k+1}\mu_{l+1})$, and $\mathbf{A}_3 = \left(\frac{1}{6}(\mu_{k+3}\mu_{l+1} + \mu_{k+1}\mu_{l+3})\right)$ indexed by $k,l = 0,\dots,p$.

Theorem 3.4. *Assume* (H1)–(H5) *with* $f \equiv 1$ *on* $[0,1]$.

- *Case ν even. Assume further that ρ is twice continuously differentiable at (x,x) and $Nh^3 \to \infty$ as $n, N \to \infty$. Then*

$$\operatorname{Var}\big(\widehat{m}_\nu(x)\big) = \frac{(\nu!)^2 \rho(x,x)}{nh^{2\nu}} \mathbf{e}_\nu' \mathbf{S}^{-1} \mathbf{S}^* \mathbf{S}^{-1} \mathbf{e}_\nu + \frac{(\nu!)^2 \rho^{(0,2)}(x,x)}{nh^{2\nu-2}} \mathbf{e}_\nu' \mathbf{S}^{-1} \mathbf{A}_1 \mathbf{S}^{-1} \mathbf{e}_\nu + o\left(\frac{1}{nh^{2\nu-2}}\right).$$

- *Case ν odd. Assume further that ρ is four times continuously differentiable at (x,x) and $Nh^5 \to \infty$ as $n, N \to \infty$. Then*

$$\operatorname{Var}\big(\widehat{m}_\nu(x)\big) = \frac{(\nu!)^2 \rho^{(1,1)}(x,x)}{nh^{2\nu-2}} \mathbf{e}_\nu' \mathbf{S}^{-1} \mathbf{A}_2 \mathbf{S}^{-1} \mathbf{e}_\nu + \frac{(\nu!)^2 \rho^{(1,3)}(x,x)}{nh^{2\nu-4}} \mathbf{e}_\nu' \mathbf{S}^{-1} \mathbf{A}_3 \mathbf{S}^{-1} \mathbf{e}_\nu + o\left(\frac{1}{nh^{2\nu-4}}\right).$$

Remark 3.5. Looking at Theorems 3.1 and 3.4, it is not clear whether the variance of $\widehat{m}_{\nu,p}(x)$ is a decreasing function of h. In other words, smoothing more may not always reduce the variance of the estimator. See [5] for a similar observation in the context of functional principal components analysis.

3.2. Optimal sampling densities and bandwidths

In this section we discuss the optimization of the (asymptotic) mean squared error

$$\begin{aligned}\mathrm{MSE} &= \mathbb{E}\big(\widehat{m}_{\nu,p}(x) - m^{(\nu)}(x)\big)^2\\ &= \mathrm{Bias}\big(\widehat{m}_{\nu,p}(x)\big)^2 + \mathrm{Var}\big(\widehat{m}_{\nu,p}(x)\big)\end{aligned}$$

in Theorem 3.1 with respect to the sampling density f and the bandwidth h. A similar optimization can be carried out in Theorem 3.4 where the covariance ρ is assumed to be more regular (twice or four times differentiable).

We first examine the choice of f that minimizes the asymptotic squared bias of $\widehat{m}_\nu(x)$. Indeed the asymptotic variance of $\widehat{m}_\nu(x)$ is independent of f, as can be seen in Theorem 3.1. This optimization is particularly useful in practice when the grid size N is not too large and subject to a sampling cost constraint.

For $p-\nu$ even, $\mathbf{e}_\nu'\mathbf{S}^{-1}\mathbf{c} = 0$ so that the first-order term in the bias vanishes, as noted in Remark 3.3. Moreover, the second-order term can be rendered equal to zero (except at zeros of $m^{(p+1)}(x)$) by taking a sampling density f such that $g_{p,\nu}(x) = 0$, where

$$g_{p,\nu}(x) = \frac{m^{(p+2)}(x)}{(p+2)!}\mathbf{e}_\nu'\mathbf{S}^{-1}\tilde{\mathbf{c}} + \frac{m^{(p+1)}(x)}{(p+1)!}\frac{f'(x)}{f(x)}\left(\mathbf{e}_\nu'\mathbf{S}^{-1}\tilde{\mathbf{c}} - \mathbf{e}_\nu'\mathbf{S}^{-1}\tilde{\mathbf{S}}\mathbf{S}^{-1}\mathbf{c}\right).$$

The solution of the previous equation is

$$f^*(x) = d_0^{-1}\left|m^{(p+1)}(x)\right|^{\gamma/(p+2)}, \tag{3.1}$$

with d_0 such that $\int_0^1 f^*(x)\mathrm{d}x = 1$ and $\gamma = \mathbf{e}_\nu'\mathbf{S}^{-1}\tilde{\mathbf{c}}/(\mathbf{e}_\nu'\mathbf{S}^{-1}\tilde{\mathbf{S}}\mathbf{S}^{-1}\mathbf{c} - \mathbf{e}_\nu'\mathbf{S}^{-1}\tilde{\mathbf{c}})$. Observe that $f^*(x)$ is well-defined over $[0,1]$ if and only if $m^{(p+1)}(x) \neq 0$ for all $x \in [0,1]$. With the choice $f = f^*$, the bias of $\widehat{m}_{\nu,p}(x)$ is of order $o(h^{p+2-\nu})$, so that a higher order expansion would be required to get the exact rate of convergence. In practice, the density f^* depends on the unknown quantity $m^{(p+1)}(x)$. It can be approximated by replacing $m^{(p+1)}(x)$ in (3.1) with $\widehat{m}_{p+1,p+k}(x)$ for some $k \geq 1$.

For $p-\nu$ odd, the first-order term in the bias is non zero (if $m^{(p+1)}(x) \neq 0$) but does not depend on f. On the other hand, the second-order term vanishes for any sampling density f. Therefore, a higher order expansion of the bias would be required to optimize the MSE with respect to f.

We turn to the optimization of the bandwidth h and start with a lemma.

Lemma 3.6. *Assume* (H5). *Then the jump function satisfies* $\alpha(x) \geq 0$.

This lemma is easily checked for covariance-stationary processes [see *e.g.* 15] but is less intuitive for general covariance functions ρ. It is helpful for determining whether the asymptotic variance of $\widehat{m}_{\nu,p}(x)$ is a decreasing function of h, in which case the MSE can be optimized. To make this optimization possible, we assume that $\alpha(x) > 0$ (or higher order differentiability for ρ if $\alpha(x) = 0$; see Rem. 3.9) throughout this section.

When estimating the regression function itself ($\nu = 0$), the leading variance term in Theorem 3.1 does not depend on h. If $\mathbf{e}_0'\mathbf{S}^{-1}\mathbf{A}\mathbf{S}^{-1}\mathbf{e}_0 \leq 0$, then the second-order variance term (in h/n) is nonnegative and the optimization of the MSE yields the solution $h = 0$, which is not admissible. In fact, we suspect that $\mathbf{e}_0'\mathbf{S}^{-1}\mathbf{A}\mathbf{S}^{-1}\mathbf{e}_0 > 0$ for all kernels K satisfying (H1) and all integers $p \geq 0$, although we have only checked it for the special cases $p \leq 2$ (local constant, linear, or quadratic fit). Under this conjecture, the optimal bandwidth for the MSE exists and can be obtained from Theorem 3.1. The cases p even and p odd must be treated separately as they correspond to different bias expressions (see Rem. 3.3 and the optimization of f above). More precisely, if $\nu = 0$ and p is odd, then the asymptotic optimal bandwidth is

$$h_{\mathrm{opt}} = \left\{\frac{(p+1)!^2\,(\mathbf{e}_0'\mathbf{S}^{-1}\mathbf{A}\mathbf{S}^{-1}\mathbf{e}_0)\,\alpha(x)}{(2p+2)\left(m^{(p+1)}(x)\right)^2(\mathbf{e}_0'\mathbf{S}^{-1}\mathbf{c})^2}\right\}^{1/(2p+1)} n^{-1/(2p+1)}.$$

In the case where $\nu = 0$ and p is even, the asymptotic optimal bandwidth becomes

$$h_{\text{opt}} = \left\{ \frac{(\mathbf{e}_0' \mathbf{S}^{-1} \mathbf{A} \mathbf{S}^{-1} \mathbf{e}_0)\, \alpha(x)}{(2p+4)\, g_{p,0}(x)^2} \right\}^{1/(2p+3)} n^{-1/(2p+3)}.$$

Note that we have assumed that $g_{p,0}(x) \neq 0$ in the previous optimization, which means that f differs from the optimal sampling density f^*. To optimize the MSE when $f = f^*$, it would be necessary to derive a higher order expansion for the bias; the optimal bandwidth would then be of order at least $n^{-1/(2p+5)}$.

In the following corollary, we give the optimal bandwidth h in two important cases ($\nu = 0, 1$), using, for simplicity, a uniform sampling density $f \equiv 1$ on $[0, 1]$. Optimal bandwidths can be obtained similarly for $\nu \geq 2$.

Corollary 3.7. *Assume* (H1)–(H5) *with* $f \equiv 1$ *on* $[0, 1]$ *and* $\alpha(x) > 0$.

1. *Local constant or linear estimation of* m ($\nu = 0$, $p \in \{0, 1\}$). *Assume further that* $m''(x) \neq 0$ *and* $Nn^{-2/3} \to \infty$ *as* $n, N \to \infty$. *Then the optimal bandwidth for the asymptotic MSE of* $\widehat{m}_0(x)$ *is*

$$h_{\text{opt}} = \left(\frac{\alpha(x)}{2\mu_2^2\, m''(x)^2} \iint_{\mathbb{R}^2} |u - v| K(u) K(v) \mathrm{d}u \mathrm{d}v \right)^{1/3} n^{-1/3}.$$

2. *Local linear or quadratic estimation of* m' ($\nu = 1$, $p \in \{1, 2\}$). *Assume further that* $m^{(3)}(x) \neq 0$ *and* $Nn^{-2/5} \to \infty$ *as* $n, N \to \infty$. *Then the optimal bandwidth for the asymptotic MSE of* $\widehat{m}_1(x)$ *is*

$$h_{\text{opt}} = \left(-\frac{9\, \alpha(x)}{2\mu_4^2\, m^{(3)}(x)^2} \iint_{\mathbb{R}^2} |u - v| uv K(u) K(v) \mathrm{d}u \mathrm{d}v \right)^{1/5} n^{-1/5}.$$

Corollary 3.7 provides the theoretical basis for the plug-in bandwidth selection method to be developed in Section 4.

Remark 3.8. In the cases $(\nu, p) = (0, 1)$ and $(\nu, p) = (1, 2)$ of Corollary 3.7, the results actually hold for any sampling density f satisfying (H4). Also, the first part of the corollary corresponds to Theorem 3 of [15] and Corollary 2.1 of [3] when the Gasser−Muller estimator is used with an equidistant design.

Remark 3.9. Under the assumptions of Theorem 3.4, in case 1 of Corollary 3.7, the optimal bandwidth is $h_{\text{opt}} = |2\, \rho^{(0,2)}(x, x)|^{1/2}\, (\mu_2 m''(x)^2)^{-1/2}\, n^{-1/2}$ provided that $\rho^{(0,2)}(x, x) < 0$ and $Nn^{-3/2} \to \infty$ as $n, N \to \infty$. In case 2, the optimal bandwidth is $h_{\text{opt}} = |6\mu_2 \rho^{(1,3)}(x, x)|^{1/2}\, (\mu_4 m^{(3)}(x)^2)^{-1/2}\, n^{-1/2}$ provided that $\rho^{(1,3)}(x, x) < 0$ and $Nn^{-5/2} \to \infty$ as $n, N \to \infty$.

Theorem 3.1 also provides optimal bandwidths for global error measures such as the integrated mean squared error

$$\text{MISE} = \int_0^1 \mathbb{E}\big(\widehat{m}_{\nu,p}(x) - m^{(\nu)}(x)\big)^2 \mathrm{d}x. \tag{3.2}$$

More precisely, denoting by $[-\tau, \tau]$ the support of K, the bias and variance expressions in Theorem 1 hold uniformly over $[\tau h, 1 - \tau h]$, and their orders are the same near the boundariesy regions $[0, \tau h)$ and $(1 - \tau h, 1]$ (only the multiplicative constants are lost). As $n, N \to \infty$, the MISE is therefore equivalent to the weighted integral over $[0, 1]$ of the (squared) bias plus variance expansions of Theorem 3.1. One can thus replace the terms $\alpha(x)$ and $(m^{(\nu)}(x))^2$ in Corollary 3.7 by $\int_0^1 \alpha(x) \mathrm{d}x$ and $\int_0^1 (m^{(\nu)}(x))^2 \mathrm{d}x$, respectively, to obtain global optimal bandwidths.

3.3. Asymptotic normality

Theorems 3.1 and 3.4 provide the normalization for the limiting distribution of $\widehat{m}_{\nu,p}(x)$. Further, $\widehat{m}_{\nu,p}(x) = \frac{1}{n}\sum_{i=1}^{n}\widehat{m}_i(x)$, where $\widehat{m}_i$ is the local polynomial smoother applied to the $Y_i(x_j)$, $j = 1,\dots,N$. Since the $\widehat{m}_i$ are i.i.d. with finite variance like the Y_i, $i = 1,\dots,n$, the central limit theorem applies. We now present the asymptotic distribution of $\widehat{m}_\nu(x)$ according to the parity of ν and p (see Rem. 3.3 on the vanishing terms in the asymptotic variance). Denote convergence in distribution by $\longrightarrow$ and the centered normal distribution with variance σ^2 by $N(0,\sigma^2)$. Imposing extra conditions on the bandwidth h to ensure negligibility of the bias, we have the following result.

Theorem 3.10. *Assume* (H1)–(H5).

- *Case ν even. Assume further that $nh^{2p+4} \to 0$ if p is even, resp. $nh^{2p+2} \to 0$ if p is odd, as $n, N \to \infty$. Then*

$$\sqrt{nh^{2\nu}}\left(\widehat{m}_{\nu,p}(x) - m^{(\nu)}(x)\right) \longrightarrow N\big(0, (\nu!)^2\,\rho(x,x)\left(\mathbf{e}_\nu'\mathbf{S}^{-1}\mathbf{S}^*\mathbf{S}^{-1}\mathbf{e}_\nu\right)\big).$$

- *Case ν odd and $\alpha(x) > 0$. Assume further that $nh^{2p+1} \to 0$ if p is even, resp. $nh^{2p+3} \to 0$ if p is odd, as $n, N \to \infty$. Then*

$$\sqrt{nh^{2\nu-1}}\left(\widehat{m}_{\nu,p}(x) - m^{(\nu)}(x)\right) \longrightarrow N\big(0, (\nu!)^2\,\alpha(x)\left|\mathbf{e}_\nu'\mathbf{S}^{-1}\mathbf{A}\mathbf{S}^{-1}\mathbf{e}_\nu\right|\big).$$

- *Case ν odd and $\alpha(x) = 0$. Assume further that $f \equiv 1$, ρ is four times differentiable at (x,x), $Nh^5 \to \infty$, and that $nh^{2p} \to 0$ if p is even, resp. $nh^{2p+2} \to 0$ if p is odd, as $n, N \to \infty$. Then*

$$\sqrt{nh^{2\nu-2}}\left(\widehat{m}_{\nu,p}(x) - m^{(\nu)}(x)\right) \longrightarrow N\big(0, (\nu!)^2\,\rho^{(1,1)}(x,x)\left(\mathbf{e}_\nu'\mathbf{S}^{-1}\mathbf{A}_2\mathbf{S}^{-1}\mathbf{e}_\nu\right)\big).$$

4. Plug-in bandwidth selection

In this section we consider the local polynomial estimation of the regression function $m(x)$. We propose a plug-in estimator for the optimal global bandwidth and determine its convergence rate. Extensions of the plug-in methodology to the estimation of derivatives $m^{(\nu)}$, $\nu \geq 1$, are straightforward. For reasons of space, we do not discuss them in this paper.

In case 1 of Corollary 3.7, the optimal global bandwidth is

$$h_{\text{opt}} = \left(\frac{I_\alpha}{2\mu_2^2\theta_{2,2}}C_1(K)\right)^{1/3} n^{-1/3} \tag{4.1}$$

with $I_\alpha = \int_0^1 \alpha(x)\mathrm{d}x$, $\theta_{r,s} = \int_0^1 m^{(r)}(x)m^{(s)}(x)\mathrm{d}x$, and $C_1(K) = \int_{\mathbb{R}^2}|u-v|K(u)K(v)\mathrm{d}u\mathrm{d}v$. The plug-in bandwidth is obtained by replacing I_α and $\theta_{2,2}$ with suitable estimators in this expression.

The integral I_α is well-estimated by averaging the quadratic variations of the sampled processes Y_i, $i = 1,\dots,n$:

$$\widehat{I}_\alpha = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=2}^{N}\left(Y_i(x_j) - Y_i(x_{j-1})\right)^2. \tag{4.2}$$

Under the regularity condition (H5), $\widehat{I}_\alpha$ is asymptotically unbiased as $N \to \infty$. Noting that $\mathbb{E}(\widehat{I}_\alpha)$ does not depend on n, it follows from the Strong Law of Large Numbers that $\widehat{I}_\alpha$ is strongly consistent as $n, N \to \infty$. A similar result for Gaussian processes can be found in [21].

Turning to the estimation of $\theta_{2,2}$, we first construct a local polynomial estimator $\widehat{m}_{2,p_\theta}$ of $m^{(2)}$ with degree p_θ and bandwidth g, and then take its L_2 norm to obtain the estimator

$$\hat{\theta}_{2,2}(g) = \int_0^1 \left(\widehat{m}_{2,p_\theta}(x)\right)^2 \mathrm{d}x. \tag{4.3}$$

The bandwidth g is selected using the cross-validation method of [24] which consists in minimizing the prediction score

$$\mathrm{CV}(g) = \frac{1}{nN} \sum_{i=1}^{n} \sum_{j=1}^{N} \left(\widehat{m}_{0,p}^{(-i)}(x_j) - Y_i(x_j) \right)^2, \tag{4.4}$$

where $\widehat{m}_{0,p}^{(-i)}(x)$ is the estimator (2.4) having degree $p = p_\theta$ and bandwidth g applied to the $(n-1)$ curves Y_k, $k \neq i$, at the design points x_j. Although this cross-validation method is designed for the estimation of m and enjoys optimality properties only in this case (see *e.g.*, [16]), it also gives reasonable bandwidths for the estimation of $m^{(\nu)}$, $\nu \geq 1$, as shown by our simulations.

We now establish the consistency of the plug-in global bandwidth for the estimation of the regression function $m(x)$. Similar results can be obtained for the general case $m^{(\nu)}(x)$ at the cost of heavy computations. For convenience, we assume that the design points are equidistant ($f \equiv 1$). We also require that the error process ε is Gaussian so that, by Isserlis' theorem, it satisfies $\mathbb{E}\left(\varepsilon(x_i)\varepsilon(x_j)\varepsilon(x_k)\right) = 0$ and

$$\mathrm{Cov}\left(\varepsilon(x_i)\varepsilon(x_j), \varepsilon(x_k)\varepsilon(x_l)\right) = \rho(x_i, x_k)\rho(x_j, x_l) + \rho(x_i, x_l)\rho(x_j, x_k)$$

for all indexes $i,\ j,\ k,\ l$. The following result gives the rate of convergence of the plug-in bandwidth estimator $\hat{h}_{\mathrm{opt}}$ to the optimal bandwidth h_{opt} for the estimation of the regression function $m(x)$.

Theorem 4.1. *Assume* (H1), (H3) *and* (H5). *In addition, assume that ε is Gaussian and that $n = O(N^{(p_\theta+3)/(p_\theta+4)})$ for some $p_\theta \in \{3,5\}$ as $n, N \to \infty$. Let $g = Gn^{-1/(p_\theta+3)}$ be a pilot bandwidth, with $G > 0$ a constant. Then the plug-in bandwidth $\hat{h}_{\mathrm{opt}}$ based on* (4.1)–(4.3) *converges in probability to h_{opt} at an optimal rate as $n, N \to \infty$ and its relative error satisfies*

$$n^{(p_\theta-1)/(p_\theta+3)} \left(\frac{\hat{h}_{\mathrm{opt}} - h_{\mathrm{opt}}}{h_{\mathrm{opt}}} \right) \longrightarrow D,$$

with $D = -\frac{4}{3}\theta_{2,2}^{-1} \left\{ \frac{\theta_{2,p_\theta+1}}{(p_\theta+1)!} \left(\mathbf{e}_2' \mathbf{S}^{-1} \mathbf{c}\right) G^{p_\theta-1} + (\mathbf{e}_2' \mathbf{S}^{-1} \mathbf{S}^* \mathbf{S}^{-1} \mathbf{e}_2) G^{-4} \int_0^1 \rho(x,x) \mathrm{d}x \right\}$.

5. Numerical study

In this section we compare the numerical performances of local polynomial estimators based on different orders of fit p and bandwidths h. The three bandwidth choices under scrutiny are the bandwidth h_{ex} that minimises the (exact, finite-sample) MISE (3.2); the plug-in bandwidth h_{plug} of Section 4; and the cross-validation bandwidth h_{cv} that minimizes the prediction score (4.4). The bandwidth h_{ex} serves as a benchmark and cannot be computed in practice.

The regression functions chosen for simulating model (2.1) are

$$\begin{cases} m_1(x) = 16(x-0.5)^4, \\ m_2(x) = \dfrac{1}{1+e^{-10(x-0.5)}} + 0.03\sin(6\pi x). \end{cases} \tag{5.1}$$

The polynomial function m_1 has unit range and has relatively high curvature away from its minimum at $x = 0.5$. The function m_2 is a linear combination of a logistic function and a rapidly varying sine function. The factor 0.03 is chosen so that the sine function has small influence on m_2 but a much larger on m_2'. These functions and their derivatives are displayed in Figure 1.

For the stochastic part of model (2.1) we use Gaussian processes with mean zero and covariance functions

$$\begin{cases} \rho_1(x,y) = \min(x,y), \\ \rho_2(x,y) = e^{-15|x-y|}. \end{cases} \tag{5.2}$$

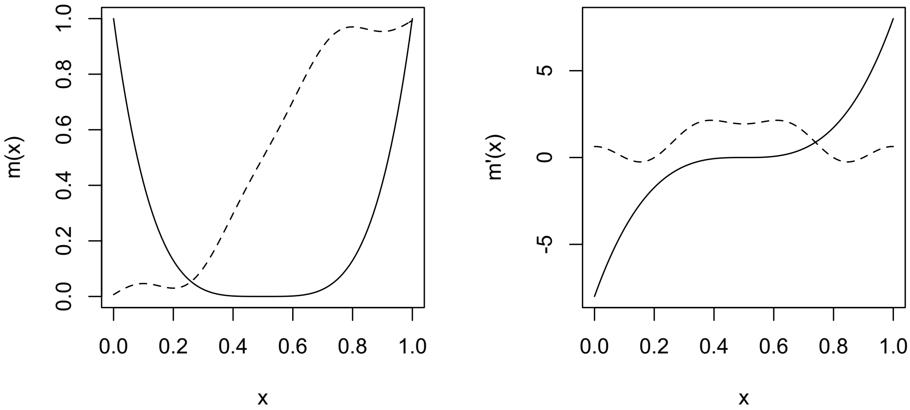

FIGURE 1. Left panel: regression functions m_1 (solid line) and m_2 (dashed). Right panel: first derivatives m_1' (solid line) and m_2' (dashed).

The first error process is a standard Wiener process on $[0, 1]$; the second is a stationary Ornstein−Uhlenbeck process. The parameter $\lambda = 15$ in ρ_2 yields correlation levels between two consecutive measurements ranging from 0.22 for $N = 10$ to 0.86 for $N = 100$. Similarly, the variance of these processes is such that when estimating $m_i, i \in \{1, 2\}$, the signal-to-noise ratio is fairly low for n small and high for n large.

The simulations are conducted in the R environment [22] using the package `SCBmeanfd` contributed by the second author. This package is available on the CRAN website `http://www.cran.r-project.org`. We consider all combinations of m_1, m_2 and ρ_1, ρ_2 with the sample size n and grid size N varying in $\{10, 20, 50, 100\}$. We examine different targets $m_i^{(\nu)}$ and estimators $\widehat{m}_{\nu,p}$ ($p = 0, 1$ for $\nu = 0$, *i.e.* local constant and linear fits, and $p = 1, 2$ for $\nu = 1$, *i.e.* local linear and quadratic fits). In each case model (2.1) is simulated 1000 times. The kernel used for the estimation is a truncated Gaussian density and the bandwidths under study are h_{ex}, h_{plug}, and h_{cv}.

Some of the extensive simulations are presented in Tables 1, 2 and 3. The Columns 1–4 of each table contain n, N, h_{ex}, the median h_{plug}, and the median h_{cv} over the 1000 simulations. The Columns 6-7-8 show the median integrated squared error $\int_0^1 (\widehat{m}_{\nu,p}(x) - m^{(\nu)}(x))^2 \mathrm{d}x$ over the 1000 simulations for h_{ex}, h_{plug}, and h_{cv}, with the interquartile range shown in brackets.

We first comment the estimation of m. Looking at Table 1 (local linear estimation of m_1 with covariance ρ_1) it appears that the bandwidths h_{ex}, h_{plug}, h_{cv} are very close and yield similar performances for almost all n, N. Similar observations hold for the local linear estimation of m_1 with covariance ρ_2. However, note in Table 1 that h_{plug} yields smaller performances when $n \in \{50, 100\}$ and $N = 10$, which can be expected since this bandwidth is only optimal for large N. Local constant estimation of m_1 and m_2 (not displayed here) yields very similar results.

The results for the estimation of m' are summarized in Tables 2 and 3 and Figure 2. Over all combinations of n, N, m_i, and ρ_i, the median efficiency of the plug-in method (resp. of the cross-validation method) relative to the optimal bandwidth h_{ex} is 80% (resp. 85%) for both the local linear and local quadratic estimators. Considering only local quadratic estimation with $N \geq 20$ (resp. $n, N \geq 50$) the median efficiency of the plug-in method increases to 86% (resp. 89%) while that of cross-validation decreases to 80% (resp. 76%). Tables 2 and 3 seem to indicate that cross-validation yields better results than plug-in under local linear estimation and worse results under local quadratic estimation. In fact this is not longer true in simulations with m_2 or ρ_2. Overall, the plug-in method yields comparable or better performances than cross-validation for large n, N but should not be used for small n, N. Cross-validation gives reasonable results in all cases.

In comparison to local linear estimation, local quadratic estimation has an extra fitting parameter, which reduces the bias at the expense of increasing the variance. Which order of local polynomial fit achieves better

TABLE 1. Local linear estimation of m_1 with Wiener process noise.

n	N	h_{ex}	h_{plug}	h_{cv}	ISE$_{ex}$	ISE$_{plug}$	ISE$_{cv}$
10	10	0.07	0.08	0.08	0.031 (0.051)	0.031 (0.050)	0.032 (0.051)
10	20	0.06	0.06	0.06	0.028 (0.048)	0.028 (0.047)	0.028 (0.048)
10	50	0.07	0.06	0.07	0.026 (0.049)	0.026 (0.050)	0.027 (0.050)
10	100	0.07	0.06	0.07	0.027 (0.047)	0.027 (0.047)	0.027 (0.048)
20	10	0.06	0.06	0.06	0.019 (0.026)	0.020 (0.026)	0.020 (0.025)
20	20	0.05	0.05	0.05	0.014 (0.025)	0.014 (0.025)	0.014 (0.025)
20	50	0.05	0.04	0.05	0.014 (0.024)	0.014 (0.024)	0.014 (0.024)
20	100	0.05	0.04	0.05	0.014 (0.026)	0.014 (0.027)	0.014 (0.027)
50	10	0.06	0.05	0.06	0.010 (0.011)	0.015 (0.012)	0.010 (0.011)
50	20	0.04	0.03	0.04	0.006 (0.011)	0.006 (0.011)	0.006 (0.011)
50	50	0.03	0.03	0.03	0.006 (0.011)	0.006 (0.010)	0.006 (0.011)
50	100	0.03	0.03	0.03	0.005 (0.010)	0.005 (0.010)	0.005 (0.010)
100	10	0.06	0.05	0.06	0.007 (0.006)	0.012 (0.006)	0.007 (0.006)
100	20	0.03	0.03	0.03	0.003 (0.005)	0.003 (0.005)	0.003 (0.005)
100	50	0.03	0.03	0.03	0.003 (0.005)	0.003 (0.005)	0.003 (0.005)
100	100	0.03	0.02	0.03	0.003 (0.005)	0.003 (0.005)	0.003 (0.005)

TABLE 2. Local linear estimation of m_1' with Wiener process noise.

n	N	h_{ex}	h_{plug}	h_{cv}	ISE$_{ex}$	ISE$_{plug}$	ISE$_{cv}$
10	10	0.07	0.08	0.08	1.96 (1.09)	2.06 (1.12)	2.02 (1.09)
10	20	0.05	0.08	0.06	1.01 (0.62)	1.29 (0.79)	1.14 (0.73)
10	50	0.05	0.08	0.07	0.84 (0.52)	1.12 (0.66)	1.06 (0.73)
10	100	0.05	0.07	0.07	0.83 (0.50)	1.08 (0.64)	1.13 (0.78)
20	10	0.06	0.07	0.06	1.75 (0.72)	1.82 (0.73)	1.78 (0.72)
20	20	0.04	0.07	0.05	0.70 (0.43)	0.96 (0.53)	0.73 (0.43)
20	50	0.04	0.07	0.05	0.52 (0.31)	0.77 (0.43)	0.60 (0.38)
20	100	0.04	0.06	0.05	0.52 (0.30)	0.76 (0.44)	0.60 (0.39)
50	10	0.06	0.06	0.06	1.65 (0.46)	1.67 (0.46)	1.66 (0.45)
50	20	0.04	0.06	0.04	0.47 (0.25)	0.65 (0.27)	0.48 (0.25)
50	50	0.03	0.06	0.03	0.29 (0.15)	0.49 (0.23)	0.30 (0.15)
50	100	0.03	0.06	0.04	0.27 (0.13)	0.45 (0.21)	0.29 (0.14)
100	10	0.06	0.05	0.06	1.60 (0.31)	1.88 (0.38)	1.61 (0.30)
100	20	0.03	0.05	0.03	0.38 (0.15)	0.49 (0.19)	0.39 (0.15)
100	50	0.02	0.05	0.03	0.18 (0.08)	0.33 (0.13)	0.18 (0.08)
100	100	0.02	0.05	0.03	0.17 (0.08)	0.31 (0.13)	0.17 (0.08)

performances in a given scenario depends on the balance between bias and variance. In Tables 2 and 3 it can be seen that with the optimal bandwidth h_{ex}, the local quadratic estimator yields sensibly better results than the local linear when the target is m_1' (due to the high curvature of m_1' which makes the bias large in comparison to the variance). The situation is however reversed when the target is m_2' (with relatively low curvature). As shown in Figure 2, the local quadratic estimator has a slightly smaller squared bias than the local linear for small h (see the left panel) but a much larger variance (see the middle panel). As a result the optimal MISE is smaller for the local linear estimator and the optimal bandwidths are quite different: $h_{ex} = 0.13$ for the linear fit and $h_{ex} = 0.30$ for the quadratic fit (see the right panel of Fig. 2).

6. Discussion

Considering independent realizations of a continuous-time stochastic process observed on a regular grid, we have derived asymptotic expansions for the bias and variance of local polynomial estimators of the mean

TABLE 3. Local quadratic estimation of m_1' with Wiener process noise.

n	N	h_{ex}	h_{plug}	h_{cv}	ISE$_{ex}$	ISE$_{plug}$	ISE$_{cv}$
10	10	0.11	0.08	0.16	0.99 (1.12)	2.74 (3.57)	1.10 (1.12)
10	20	0.10	0.08	0.14	0.61 (0.51)	0.71 (0.55)	0.71 (0.66)
10	50	0.09	0.08	0.15	0.53 (0.41)	0.60 (0.45)	0.76 (0.72)
10	100	0.09	0.07	0.15	0.52 (0.41)	0.58 (0.44)	0.76 (0.79)
20	10	0.10	0.07	0.11	0.70 (0.78)	5.08 (3.19)	0.76 (0.79)
20	20	0.08	0.07	0.11	0.37 (0.31)	0.41 (0.35)	0.40 (0.34)
20	50	0.08	0.07	0.12	0.30 (0.23)	0.33 (0.24)	0.38 (0.34)
20	100	0.08	0.06	0.12	0.30 (0.22)	0.32 (0.23)	0.42 (0.34)
50	10	0.09	0.06	0.11	0.52 (0.47)	7.55 (1.34)	0.57 (0.52)
50	20	0.06	0.06	0.09	0.19 (0.16)	0.20 (0.17)	0.20 (0.15)
50	50	0.06	0.06	0.10	0.15 (0.10)	0.16 (0.10)	0.20 (0.15)
50	100	0.06	0.06	0.10	0.14 (0.10)	0.15 (0.10)	0.20 (0.15)
100	10	0.09	0.05	0.10	0.48 (0.35)	8.27 (1.40)	0.50 (0.39)
100	20	0.05	0.05	0.08	0.12 (0.10)	0.12 (0.10)	0.13 (0.10)
100	50	0.05	0.05	0.08	0.08 (0.06)	0.09 (0.06)	0.12 (0.08)
100	100	0.05	0.05	0.09	0.08 (0.05)	0.08 (0.06)	0.12 (0.08)

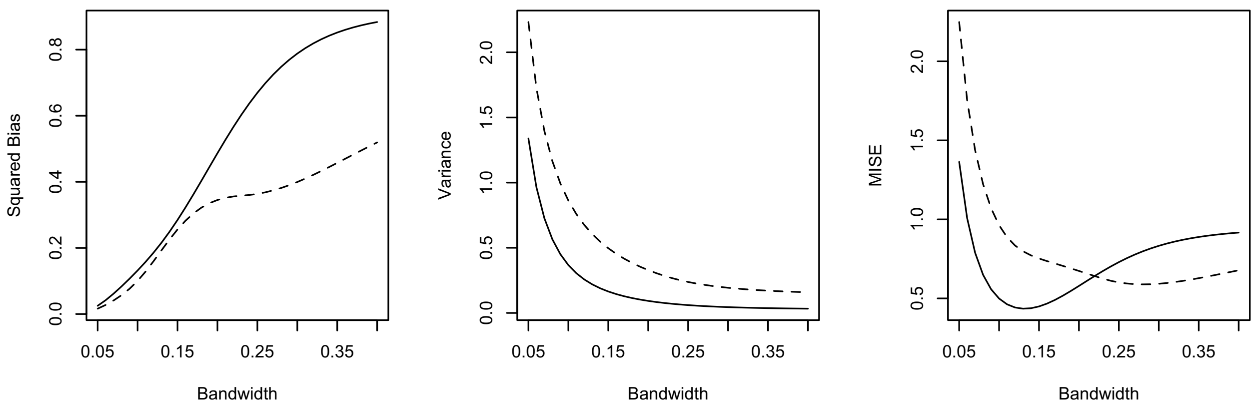

FIGURE 2. Comparison of local linear (solid line) and local quadratic fitting (dashed line) for the estimation of derivatives. The estimation target is m_2' and the covariance function is ρ_2, with $n = N = 50$ in (2.1).

function and its derivatives. Using the same arguments as in the present paper, these results can be extended to noisy observations of continuous processes defined on multivariate domains. Based on these results, we have deduced optimal sampling densities and bandwidths and devised a plug-in bandwidth selection method. In simulations, this plug-in method appears as a valid alternative to cross-validation when the observation grid is moderate to large (this is typically the case for functional data). Given that cross-validation produces nearly optimal results in our simulation setup, any improvement brought by another bandwidth selection method is bound to be small. In this light, the fact that the proposed plug-in method yields comparable performances to cross-validation is a positive finding. Furthermore, the computation of the plug-in bandwidth is much faster than the cross-validation procedure, especially for large data sets. Another salient result of the numerical study is that although cross-validation is primarily intended for estimating the mean function, it also gives satisfactory results in derivative estimation.

Appendix A. Proofs

A.1. Proof of Theorem 3.1

Bias term.

Write $\mathbf{m}_N = (m(x_1), \ldots, m(x_N))'$ and define the $(p+1) \times (p+1)$ matrix $\mathbf{S}_N = \mathbf{X}_N' \mathbf{W}_N \mathbf{X}_N$ with (k,l)th element $(0 \le k, l \le p)$ given by

$$s_{k+l,N} = \frac{1}{Nh} \sum_{j=1}^{N} (x_j - x)^{k+l} K\left(\frac{x_j - x}{h}\right).$$

It follows from (2.4) that

$$\mathbb{E}(\widehat{\boldsymbol{\beta}}_N(x)) = \mathbf{S}_N^{-1} \mathbf{X}_N' \mathbf{W}_N \mathbf{m}_N. \tag{A.1}$$

In view of (H3), the Taylor expansion of $m(x_j)$ at the order $(p+2)$ is

$$m(x_j) = \sum_{k=0}^{p+2} (x_j - x)^k \beta_k(x) + o\left((x_j - x)^{p+2}\right)$$

and thus

$$\mathbf{m}_N = \mathbf{X}_N \boldsymbol{\beta}(x) + \beta_{p+1}(x) \begin{pmatrix} (x_1 - x)^{p+1} \\ \vdots \\ (x_N - x)^{p+1} \end{pmatrix} + (\beta_{p+2}(x) + o(1)) \begin{pmatrix} (x_1 - x)^{p+2} \\ \vdots \\ (x_N - x)^{p+2} \end{pmatrix}.$$

As a result the bias in the estimation of $\boldsymbol{\beta}(x)$ is

$$\mathbb{E}(\widehat{\boldsymbol{\beta}}_N(x)) - \boldsymbol{\beta}(x) = \beta_{p+1} \mathbf{S}_N^{-1} \mathbf{c}_N + (\beta_{p+2} + o(1))\, \mathbf{S}_N^{-1} \tilde{\mathbf{c}}_N, \tag{A.2}$$

where $\mathbf{c}_N = (s_{p+1,N}, \ldots, s_{2p+1,N})'$ and $\tilde{\mathbf{c}}_N = (s_{p+2,N}, \ldots, s_{2p+2,N})'$.

Using (H1)$-$(H4), straightforward calculations yield the following approximation to the elements of the matrix $\mathbf{S}_N$:

$$s_{k+l,N} = h^{k+l} \left(\int_{-\infty}^{\infty} u^{k+l} K(u) f(x + hu) du + O\left((Nh)^{-1}\right) \right), \tag{A.3}$$

the $O((Nh)^{-1})$ being the error in the integral approximation of a Riemann sum.

Based on (H4), a Taylor expansion of $f(x + hu)$ at order 1 yields

$$s_{k+l,N} = h^{k+l} \big(\mu_{k+l} f(x) + h \mu_{k+l+1} f'(x) + o(h) \big) \tag{A.4}$$

under the condition $Nh^2 \to \infty$ in (H2). The last relation stands in matrix form as

$$\mathbf{S}_N = \mathbf{H} \left(f(x) \mathbf{S} + h f'(x) \tilde{\mathbf{S}} + o(h) \right) \mathbf{H}, \tag{A.5}$$

where $\mathbf{H} = diag(1, h, \ldots, h^p)$. In particular, it holds that

$$\begin{cases} \mathbf{c}_N = h^{p+1} \mathbf{H} \big(f(x) \mathbf{c} + (h f'(x) + o(h))\, \tilde{\mathbf{c}} \big), \\ \tilde{\mathbf{c}}_N = h^{p+2} \mathbf{H}\, (f(x) + o(1))\, \tilde{\mathbf{c}}. \end{cases} \tag{A.6}$$

With the relation $(\mathbf{A} + h\mathbf{B})^{-1} = \mathbf{A}^{-1} - h \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1} + o(h)$ holding for any invertible matrices $\mathbf{A}, \mathbf{B}$ of compatible dimensions, we have

$$\mathbf{S}_N^{-1} = \mathbf{H}^{-1} \left(\frac{1}{f(x)} \mathbf{S}^{-1} - h \frac{f'(x)}{f^2(x)} \mathbf{S}^{-1} \tilde{\mathbf{S}} \mathbf{S}^{-1} + o(h) \right) \mathbf{H}^{-1}. \tag{A.7}$$

Plugging (A.6) and (A.7) in (A.2) and truncating the expansion to the second order, the bias expression of Theorem 3.1 follows.

Variance term.

Define the $N \times N$ matrix $\mathbf{V}_N = (\rho(x_i, x_j))$ and the $(p+1) \times (p+1)$ matrix $\mathbf{S}_N^* = \mathbf{X}_N' \mathbf{W}_N \mathbf{V}_N \mathbf{W}_N \mathbf{X}_N$. Noting that $\mathrm{Var}(\overline{\mathbf{Y}}_N) = n^{-1}\mathbf{V}_N$ and considering (2.4), it can be seen that

$$\mathrm{Var}\big(\widehat{\boldsymbol{\beta}}_N(x)\big) = n^{-1}\mathbf{S}_N^{-1}\mathbf{S}_N^*\mathbf{S}_N^{-1}. \tag{A.8}$$

The asymptotic behavior of $\mathbf{S}_N^*$ is characterized in the following lemma.

Lemma A.1. *Assume* (H1) *and* (H5)*. Then as $n, N \to \infty$,*

$$\begin{aligned}\mathbf{S}_N^* = \mathbf{H}\Big\{&\phi(x,x)\mathbf{S}^* + h(\phi^{(0,1)}(x,x^+) - \phi^{(0,1)}(x,x^-))\mathbf{A}\\ &+ h(\phi^{(0,1)}(x,x^+) + \phi^{(0,1)}(x,x^-))\mathbf{B} + o(h)\Big\}\mathbf{H}\end{aligned}$$

with the matrices $\mathbf{A}$, $\mathbf{S}^*$ *as in Section 3,* $\mathbf{B} = \big(\frac{1}{2}(\mu_{k+1}\mu_l + \mu_k\mu_{l+1})\big)$*, and* $\phi(y,z) = \rho(y,z)f(y)f(z)$*.*

Plugging Lemma A.1 and (A.7) in (A.8), we have

$$\begin{aligned} nf(x)^2\,\mathbf{H}\mathrm{Var}(\widehat{\boldsymbol{\beta}}_N(x))\mathbf{H} &= \phi(x,x)\,\mathbf{S}^{-1}\mathbf{S}^*\mathbf{S}^{-1} + o(h)\\ &\quad - h\phi(x,x)\,\frac{f'(x)}{f(x)}\,\Big(\mathbf{S}^{-1}\tilde{\mathbf{S}}\mathbf{S}^{-1}\mathbf{S}^*\mathbf{S}^{-1} + \mathbf{S}^{-1}\mathbf{S}^*\mathbf{S}^{-1}\tilde{\mathbf{S}}\mathbf{S}^{-1}\Big)\\ &\quad + h\,\Big(\phi^{(0,1)}(x,x^+) - \phi^{(0,1)}(x,x^-)\Big)\,\mathbf{S}^{-1}\mathbf{A}\mathbf{S}^{-1}\\ &\quad + h\,\Big(\phi^{(0,1)}(x,x^+) + \phi^{(0,1)}(x,x^-)\Big)\,\mathbf{S}^{-1}\mathbf{B}\mathbf{S}^{-1}. \end{aligned} \tag{A.9}$$

Note that the $o(h)$ above stands for a matrix whose coefficients are negligible compared to h as $h \to 0$.

Expressing $\phi^{(0,1)}(x,x^{\pm})$ in terms of $\rho^{(0,1)}(x,x^{\pm})$, we get

$$\phi^{(0,1)}(x,x^{\pm}) = f(x)f'(x)\rho(x,x) + f^2(x)\rho^{(0,1)}(x,x^{\pm}) \tag{A.10}$$

and

$$\begin{aligned} n\,\mathbf{H}\,\mathrm{Var}\big(\widehat{\boldsymbol{\beta}}_N(x)\big)\mathbf{H} &= \rho(x,x)\mathbf{S}^{-1}\mathbf{S}^*\mathbf{S}^{-1} + o(h)\\ &\quad - h\rho(x,x)\,\frac{f'(x)}{f(x)}\,\Big(\mathbf{S}^{-1}\tilde{\mathbf{S}}\mathbf{S}^{-1}\mathbf{S}^*\mathbf{S}^{-1} + \mathbf{S}^{-1}\mathbf{S}^*\mathbf{S}^{-1}\tilde{\mathbf{S}}\mathbf{S}^{-1}\Big)\\ &\quad + h\,(\rho^{(0,1)}(x,x^+) - \rho^{(0,1)}(x,x^-))\,\mathbf{S}^{-1}\mathbf{A}\mathbf{S}^{-1}\\ &\quad + h\left(2\frac{f'(x)}{f(x)}\rho(x,x) + \Big(\rho^{(0,1)}(x,x^+) + \rho^{(0,1)}(x,x^-)\Big)\right)\mathbf{S}^{-1}\mathbf{B}\mathbf{S}^{-1}. \end{aligned} \tag{A.11}$$

This variance expression can be simplified further due to the fact that

$$\begin{cases} \mathbf{e}_\nu'\mathbf{S}^{-1}\tilde{\mathbf{S}}\mathbf{S}^{-1}\mathbf{S}^*\mathbf{S}^{-1}\mathbf{e}_\nu = 0\\ \mathbf{e}_\nu'\mathbf{S}^{-1}\mathbf{B}\mathbf{S}^{-1}\mathbf{e}_\nu = 0 \end{cases} \tag{A.12}$$

for all $\nu = 0, \ldots, p$. Indeed, by the symmetry of K, $\mathbf{S} = (\mu_{k+l})$ has its (k,l)th entry equal to zero if k,l are of different parity. The same property can be established for $\mathbf{S}^{-1}$ by cofactor arguments. For $\tilde{\mathbf{S}} = (\mu_{k+l+1})$, the (k,l)th entry is zero if k,l are of the same parity. For $\mathbf{S}^* = (\mu_k\mu_l)$, the sparsity is even stronger: all rows and columns of odd order (recall that the indexing starts at 0) have their entries equal to zero. Basic matrix algebra then shows that the matrices $\mathbf{S}^{-1}\tilde{\mathbf{S}}\mathbf{S}^{-1}$ and $\mathbf{S}^*\mathbf{S}^{-1}$ have the same sparsity structures as $\tilde{\mathbf{S}}$ and $\mathbf{S}^*$, respectively. As a result $\mathbf{S}^{-1}\tilde{\mathbf{S}}\mathbf{S}^{-1}\mathbf{S}^*\mathbf{S}^{-1}$ has its diagonal coefficients equal to zero and the first part of (A.12) follows. The second part is derived along the same lines. It suffices to notice that $B_{kl} = \mu_k \mu_{l+1} + \mu_{k+1}\mu_l = 0$ if k and l are the same parity, so that $\mathbf{S}^{-1}\mathbf{B}\mathbf{S}^{-1}$ has all its diagonal coefficients equal to zero.

Finally, we deduce from (A.11) and (A.12) that

$$\begin{aligned} n\mathrm{Var}(\widehat{m}_\nu(x)) &= n(\nu!)^2\, \mathbf{e}_\nu' \,\mathrm{Var}(\widehat{\boldsymbol{\beta}}_N(x))\, \mathbf{e}_\nu \\ &= (\nu!)^2 h^{-2\nu}\, \rho(x,x)\, \mathbf{e}_\nu' \mathbf{S}^{-1}\mathbf{S}^*\mathbf{S}^{-1}\mathbf{e}_\nu + o(h^{-2\nu+1}) \\ &\quad + (\nu!)^2 h^{-2\nu+1}\, (\rho^{(0,1)}(x,x^+) - \rho^{(0,1)}(x,x^-))\, \mathbf{e}_\nu' \mathbf{S}^{-1}\mathbf{A}\mathbf{S}^{-1}\mathbf{e}_\nu, \end{aligned}$$

which completes the Proof of Theorem 3.1.

A.2. Proof of Lemma A.1.

The arguments used to approximate $\mathbf{S}_N$ in (A.3) can be applied again to show that, under (H1) and (H2) and (H4) and (H5),

$$\begin{aligned} s^*_{kl,N} &= \frac{1}{(Nh)^2} \sum_{i=1}^N \sum_{j=1}^N (x_i - x)^k (x_j - x)^l K\left(\frac{x_i - x}{h}\right) K\left(\frac{x_j - x}{h}\right) \rho\,(x_i, x_j) \\ &= \frac{1}{h^2} \iint_{[-1,1]^2} (u-x)^k (v-x)^l K\left(\frac{u-x}{h}\right) K\left(\frac{v-x}{h}\right) \rho(u,v) f(u) f(v) \mathrm{d}u\mathrm{d}v + O\left(\frac{h^{k+l}}{Nh}\right) \\ &= h^{k+l} \iint_{[-1,1]^2} u^k v^l \phi(x+hu, x+hv) K(u)K(v) \mathrm{d}u\mathrm{d}v + o\left(h^{k+l+1}\right). \end{aligned} \tag{A.13}$$

Performing Taylor expansions and using (H4) and (H5), one can show that

$$\phi(x+hu, x+hv) = \phi(x,x) + hu\phi^{(0,1)}(x,x^-) + hv\phi^{(0,1)}(x,x^+) + o(h)$$

for all $-1 \le u \le v \le 1$. This result is obtained by introducing a pivotal point $(x+hu, x)$ or $(x, x+hv)$ such that the lines connecting this point to $(x+hu, x+hv)$ and (x,x) do not cross the main diagonal of $[0,1]^2$. Since ϕ is differentiable on each side of the diagonal, one can then perform Taylor expansions along the connecting lines. The above result also relies on the identities $\phi^{(1,0)}(x^+,x) = \phi^{(0,1)}(x,x^+) = \phi^{(0,1)}(x^-,x)$ and $\phi^{(1,0)}(x^-,x) = \phi^{(0,1)}(x,x^-) = \phi^{(0,1)}(x^+,x)$ (thanks to the symmetry of ϕ and the continuity of the first partial derivatives of ϕ on either side of the diagonal). By symmetry considerations, it holds for all $u, v \in [-1,1]$ that

$$\begin{aligned} \phi(x+hu, x+hv) =& \phi(x,x) + h\,(u \wedge v)\, \phi^{(0,1)}(x,x^-) \\ &+ h\,(u \vee v)\, \phi^{(0,1)}(x,x^+) + o(h). \end{aligned} \tag{A.14}$$

Using the fact that $(u \wedge v) + (u \vee v) = u + v$ and $(u \vee v) - (u \wedge v) = |u - v|$ and writing $\phi^{(0,1)}(x,x^\pm) = \frac{1}{2}\,(\phi^{(0,1)}(x,x^+) + \phi^{(0,1)}(x,x^-)) \pm \frac{1}{2}\,(\phi^{(0,1)}(x,x^+) - \phi^{(0,1)}(x,x^-))$, one concludes with the dominated convergence theorem that

$$\begin{aligned} s^*_{kl,N} &= h^{k+l} \iint_{[-1,1]^2} u^k v^l K(u)K(v) \phi(x+hu, x+hv) \mathrm{d}u\mathrm{d}v + o(h^{k+l+1}) \\ &= h^{k+l} \Big\{ \phi(x,x)\mu_k\mu_l + \frac{h}{2}\left(\phi^{(0,1)}(x,x^+) + \phi^{(0,1)}(x,x^-)\right)(\mu_{k+1}\mu_l + \mu_k\mu_{l+1}) \\ &\quad + \frac{h}{2}\left(\phi^{(0,1)}(x,x^+) - \phi^{(0,1)}(x,x^-)\right) \iint_{[-1,1]^2} |u-v| u^k v^l K(u)K(v) \mathrm{d}u\mathrm{d}v \Big\} \\ &\quad + o(h^{k+l+1}). \end{aligned}$$

$\square$

A.3. Proof of Theorem 3.4.

This result is obtained along the same lines as Theorem 3.1. More precisely, it suffices to push the matrix expansions of $\mathbf{S}_N^{-1}$ in (A.7) and $\mathbf{S}_N^*$ in Lemma A.1 to a higher order d. First, since $f \equiv 1$, it is easily seen that $\mathbf{S}_N = \{1+o(h^d)\}\mathbf{HSH}$ provided that $Nh^{d+1} \to \infty$. Therefore, (A.7) simply extends in $\mathbf{S}_N^{-1} = \{1+o(h^d)\}\mathbf{H}^{-1}\mathbf{S}^{-1}\mathbf{H}^{-1}$. Second, if the covariance ρ is d times differentiable at (x,x), then a Taylor expansion of order d can be performed for $\rho(x+hu, x+hv)$, followed by an application of the dominated convergence theorem over $[-1,1]^2$ as $h \to 0$. For $d = 4$, we get for instance (see the Proof of Lem. A.1):

$$
\begin{aligned}
s^*_{kl,N} &= h^{k+l} \iint_{[-1,1]^2} u^k v^l K(u)K(v)\rho(x+hu, x+hv)\mathrm{d}u\mathrm{d}v + o(h^{k+l+4}) \\
&= h^{k+l}\Big\{\rho(x,x)\mu_k\mu_l + h\rho^{(0,1)}(x,x)\left(\mu_{k+1}\mu_l + \mu_k\mu_{l+1}\right) \\
&\quad + h^2\left(\rho^{(0,2)}(x,x)\frac{\mu_{k+2}\mu_l + \mu_k\mu_{l+2}}{2!} + \rho^{(1,1)}(x,x)\mu_{k+1}\mu_{l+1}\right) \\
&\quad + h^3\left(\rho^{(0,3)}(x,x)\frac{\mu_{k+3}\mu_l + \mu_k\mu_{l+3}}{3!} + \rho^{(1,2)}(x,x)\frac{\mu_{k+2}\mu_{l+1} + \mu_{k+1}\mu_{l+2}}{2!}\right) \\
&\quad + h^4\left(\rho^{(0,4)}(x,x)\frac{\mu_{k+4}\mu_l + \mu_k\mu_{l+4}}{4!} + \rho^{(1,3)}(x,x)\frac{\mu_{k+3}\mu_{l+1} + \mu_{k+1}\mu_{l+3}}{3!} + \rho^{(2,2)}(x,x)\frac{\mu_{k+2}\mu_{l+2}}{2!\,2!}\right) + o\left(h^4\right)\Big\}.
\end{aligned}
\tag{A.15}
$$

The arguments used in Theorem 3.1 relative to the sparsity structure of $\mathbf{S}^{-1}$ and the limit matrix of $\mathbf{S}_N^*$ still apply here. In a nutshell, the matrices of the form $(\mu_{k+a}\mu_{l+b})$ in (A.15) that do contribute to the limit variance of $\widehat{m}_{\nu,p}(x)$ are those for which both $\nu+a$ and $\nu+b$ are even (nonzero moments of the kernel K). Therefore, the terms of order h and h^3 inside the brackets of (A.15) do not contribute to the limit variance of $\widehat{m}_{\nu,p}(x)$. For ν even, the terms $\mu_{k+1}\mu_{l+1}$ in (A.15) do not contribute either but the terms $\mu_k\mu_l$ and $h^2(\mu_{k+2}\mu_l + \mu_k\mu_{l+2})$ do. An expansion to order $d = 2$ is thus sufficient. For ν odd, only the terms $h^2\mu_{k+1}\mu_{l+1}$ and $h^4(\mu_{k+3}\mu_{l+1} + \mu_{k+1}\mu_{l+3})$ contribute to the limit variance of $\widehat{m}_{\nu,p}(x)$ up to order 4. In this case the expansion to order $d = 4$ is necessary, as an expansion to order 2 only results in a variance term of order $1/n$ (independent of h) when $\nu = 1$. Theorem 3.4 immediately follows from these arguments. □

A.4. Proof of Lemma 3.6.

Starting from the Taylor expansion (A.14) and the subsequent argument in the Proof of Lemma A.1, it can be shown that for all $u, v \in [-1,1]^2$,

$$
\begin{aligned}
\rho(x+hu, x+hv) = \rho(x,x) &+ \frac{h}{2}\left(\rho^{(0,1)}(x,x^+) + \rho^{(0,1)}(x,x^-)\right)(u+v) \\
&+ \frac{h}{2}\left(\rho^{(0,1)}(x,x^+) - \rho^{(0,1)}(x,x^-)\right)|u-v| + o(h)\,.
\end{aligned}
\tag{A.16}
$$

Let us write $a = \left(\rho^{(0,1)}(x,x^+) + \rho^{(0,1)}(x,x^-)\right)/2$ and $b = \left(\rho^{(0,1)}(x,x^+) - \rho^{(0,1)}(x,x^-)\right)/2$ for brevity. The dominated convergence theorem and (H5) imply that for any bounded, measurable function g on $[-1,1]$,

$$
\begin{aligned}
\iint_{[-1,1]^2} \rho(x+hu, x+hv)g(u)g(v)\mathrm{d}u\mathrm{d}v = \rho(x,x)\left(\int_{-1}^{1} g(u)\mathrm{d}u\right)^2 &+ 2ah\int_{-1}^{1} g(u)\mathrm{d}u\int_{-1}^{1} vg(v)\mathrm{d}v \\
&+ bh\iint_{[-1,1]^2} g(u)g(v)|u-v|\mathrm{d}u\mathrm{d}v + o(h).
\end{aligned}
\tag{A.17}
$$

The left-hand side of (A.17) is non-negative since the covariance ρ is a non-negative definite function. By taking $g = \mathrm{Id}_{[-1,1]}$, we have $\int_{-1}^{1} g(u)\mathrm{d}u = 0$ so that the remaining term $bh\iint_{[-1,1]^2} g(u)g(v)|u-v|\mathrm{d}u\mathrm{d}v$ in the right-hand side of (A.17) is also non-negative. Since $\iint_{[-1,1]^2} uv|u-v|\mathrm{d}u\mathrm{d}v = -8/15 < 0$, this means that $b \le 0$ and hence $\alpha(x) = \rho^{(0,1)}(x,x^-) - \rho^{(0,1)}(x,x^+) \ge 0$. □

A.5. Proof of Theorem 4.1.

Applying Theorem 3.1 with $\nu = 2$ and skipping the details, we obtain the bias expression

$$\mathbb{E}\big(\hat\theta_{2,2}(g)\big) - \theta_{2,2} \approx \frac{4}{(p_\theta+1)!}\left(\mathbf{e}_2'\mathbf{S}^{-1}\mathbf{c}\right)\theta_{2,p_\theta+1}\, g^{p_\theta-1} + 4\left(\mathbf{e}_2'\mathbf{S}^{-1}\mathbf{S}^*\mathbf{S}^{-1}\mathbf{e}_2\right) n^{-1}g^{-4}\int_0^1 \rho(x,x)\mathrm{d}x. \tag{A.18}$$

The variance of $\hat\theta_{2,2}$ can be studied along the same lines as [26] and [12] which respectively handle the case of independent errors and the case of correlated stationary errors. It is approximately decomposed as

$$\operatorname{Var}\big(\hat\theta_{2,2}(g)\big) \approx \sum_{l=1}^{7} A_l$$

where the A_l are defined in equation (44) of [12]. It can be shown that

$$\begin{aligned}
&A_1 = O(1/nN^3g^{10}),\ A_2 = O(1/nN^2g^9),\ A_3 = O(1/nN^2g^{10}),\\
&A_4 = O(1/nN^2g^8) + O(1/nN^3g^{10}) + o(1/n^2N^2g^9),\\
&A_5 = O(1/nNg^5) + O(1/nN^2g^{10}) + o(1/n^2Ng^8),\\
&A_6 = O(1/nNg^7) + O(1/nN^2g^9),\ A_7 = O(1/nNg^9) + O(1/n) + o(1/n^2g^7).
\end{aligned}$$

Given that $1/N = O(g)$ (otherwise the estimator $\hat m_{2,p_\theta}$ is not well-defined), it follows that leading rates for the variance are

$$\operatorname{Var}\big(\hat\theta_{2,2}(g)\big) = O(1/nNg^9) + o(1/n^2g^7) + O\left(1/n\right). \tag{A.19}$$

For instance, the covariance term in A_7 decomposes as

$$\begin{aligned}
\operatorname{Cov}\big(\bar Y_k\,\bar Y_u\,,\bar Y_l\,\bar Y_v\big) &= m(x_k)\,m(x_l)\operatorname{Cov}(\bar\varepsilon(x_u),\bar\varepsilon(x_v)) + m(x_k)\,m(x_v)\operatorname{Cov}(\bar\varepsilon(x_u),\bar\varepsilon(x_l))\\
&\quad + m(x_u)\,m(x_l)\operatorname{Cov}(\bar\varepsilon(x_k),\bar\varepsilon(x_v)) + m(x_u)\,m(x_v)\operatorname{Cov}(\bar\varepsilon(x_k),\bar\varepsilon(x_l))\\
&\quad + \operatorname{Cov}(\bar\varepsilon(x_k),\bar\varepsilon(x_l))\operatorname{Cov}(\bar\varepsilon_u,\bar\varepsilon_v) + \operatorname{Cov}(\bar\varepsilon(x_k),\bar\varepsilon(x_v))\operatorname{Cov}(\bar\varepsilon(x_l),\bar\varepsilon(x_u)).
\end{aligned}$$

Noticing the symmetries of the problem, we obtain $A_7 \approx A_{7,1} + A_{7,2}$ with

$$\begin{aligned}
A_{7,1} = \frac{4}{N^6g^{12}}\sum_{i,j}\sum_{k}\sum_{l\neq k}\sum_{u\neq k}\sum_{v\neq k} K_{2,p_\theta}\left(\frac{x_i-x_k}{g}\right)K_{2,p_\theta}\left(\frac{x_i-x_u}{g}\right)K_{2,p_\theta}\left(\frac{x_j-x_l}{g}\right)K_{2,p_\theta}\left(\frac{x_j-x_v}{g}\right)\\
\times m(x_k)\,m(x_l)\operatorname{Cov}\big(\bar\varepsilon(x_u),\bar\varepsilon(x_v)\big)
\end{aligned}$$

and

$$\begin{aligned}
A_{7,2} = \frac{2}{N^6g^{12}}\sum_{i,j}\sum_{k}\sum_{l\neq k}\sum_{u\neq k}\sum_{v\neq k} K_{2,p_\theta}\left(\frac{x_i-x_k}{g}\right)K_{2,p_\theta}\left(\frac{x_i-x_u}{g}\right)K_{2,p_\theta}\left(\frac{x_j-x_l}{g}\right)K_{2,p_\theta}\left(\frac{x_j-x_v}{g}\right)\\
\times \operatorname{Cov}\big(\bar\varepsilon(x_k),\bar\varepsilon(x_l)\big)\operatorname{Cov}\big(\bar\varepsilon(x_u),\bar\varepsilon(x_v)\big).
\end{aligned}$$

The kernel K_{2,p_θ} of order $(2,p_\theta)$ is defined *e.g.*, in equation (16) of [12]. Given that $p_\theta \in \{3,5\}$ is odd, this kernel satisfies

$$\int u^r K_{2,p_\theta}(u)\mathrm{d}u = \begin{cases} 0, & 0\le r\le p_\theta,\ r\neq 2,\\ 2!, & r=2,\\ c_{2,p_\theta}, & r = p_\theta+1,\end{cases}$$

where c_{2,p_θ} is a non-zero constant.

Denoting by $(L_1 * L_2)(x) = \int L_1(u)L_2(x-u)\mathrm{d}u$ the convolution of two real-valued functions L_1 and L_2, we first derive

$$\begin{aligned}
A_{7,1} &\approx \frac{4}{ng^{10}} \iint \left(\int \left(K_{2,p_\theta} * K_{2,p_\theta}\right)\left(\frac{w-x}{g}\right) m(x)\mathrm{d}x \right) \left(\int \left(K_{2,p_\theta} * K_{2,p_\theta}\right)\left(\frac{z-y}{g}\right) m(y)\mathrm{d}y \right) \rho(w,z)\mathrm{d}w\mathrm{d}z \\
&= \frac{4}{ng^{8}} \iint \left(\int \left(K_{2,p_\theta} * K_{2,p_\theta}\right)(s)\, m(w-gs)\mathrm{d}s \right) \left(\int \left(K_{2,p_\theta} * K_{2,p_\theta}\right)(t)\, m(z-gt)dt \right) \rho(w,z)\mathrm{d}w\mathrm{d}z \\
&\approx \frac{4}{ng^{8}} \iint \left(\sum_{k=1}^{4} (-g)^k m^{(k)}(w) \int_s s^k \left(K_{2,p_\theta} * K_{2,p_\theta}\right)(s)\mathrm{d}s \right) \\
&\quad \times \left(\sum_{k=1}^{4} (-g)^k m^{(k)}(z)\mathrm{d}z \int_t t^k \left(K_{2,p_\theta} * K_{2,p_\theta}\right)(t)\mathrm{d}t \right) \rho(w,z)\mathrm{d}w\mathrm{d}z \\
&= \frac{4\,(4!)^2}{n} \int\int m^{(4)}(w)\, m^{(4)}(z)\, \rho(w,z)\mathrm{d}w\mathrm{d}z \\
&= O\left(\frac{1}{n}\right) + O\left(\frac{1}{nNg^9}\right).
\end{aligned}$$

In the above calculation, we have used the fact that K_{2,p_θ} is symmetric, that the convolution $\left(K_{2,p_\theta} * K_{2,p_\theta}\right)$ has null moments up to order 3 and fourth moment equal to 4! (see *e.g.*, [12, 26]), and that m is at least four times continuously differentiable by (H3). The term $O(1/nNg^9)$, which is specified only in the last line, is due to the cost of approximating the sum $A_{7,1}$ by its Riemann integral.

By the regularity assumption (H5) on ρ, we have

$$\begin{aligned}
A_{7,2} &\approx \frac{2}{n^2g^{10}} \int\int\int\int \left(K_{2,p_\theta} * K_{2,p_\theta}\right)\left(\frac{x-w}{g}\right) \left(K_{2,p_\theta} * K_{2,p_\theta}\right)\left(\frac{y-z}{g}\right) \rho(x,y)\, \rho(u,v)\, \mathrm{d}x\mathrm{d}y\mathrm{d}w\mathrm{d}z \\
&= \frac{1}{n^2g^{8}} \int\int\int\int \left(K_{2,p_\theta} * K_{2,p_\theta}\right)(s) \left(K_{2,p_\theta} * K_{2,p_\theta}\right)(t)\, \rho(x-gs, y-gt)\, \rho(x,y)\, \mathrm{d}x\mathrm{d}y\mathrm{d}s\mathrm{d}t \\
&= o\left(\frac{1}{n^2g^7}\right) + O\left(\frac{1}{nNg^9}\right).
\end{aligned}$$

The last equality derives from the Taylor expansion $\rho(x-gs, y-gt) \approx \rho(x,y) - gs\rho^{(1,0)}(x,y) - gt\rho^{(0,1)}(x,y)$ and from the nullity of the first moments of $(K_{2,p_\theta} * K_{2,p_\theta})$. After integrating out s and t, the remaining function of (x,y) to integrate is thus of order $o(g)$. Note that the Taylor expansion is well defined only if $x \neq y$. However, the diagonal $\{x = y\}$ has Lebesgue measure 0 and can be ignored in the integration. The term $O(1/nNg^9)$ in the last line arises from the integral approximation of the sum $A_{7,2}$.

Combining the rates (A.18)−(A.19) for the bias and variance of $\hat{\theta}_{2,2}(g)$, we obtain a MSE of order $O\left(g^{2p_\theta-2}\right) + O\left(1/n^2g^8\right) + O(1/n) + O(1/nNg^9)$. If we assume that $1/n^2g^8 = o(1/nNg^9)$, then the optimization of the MSE yields a bandwidth g of order $(1/nN)^{1/(2p_\theta+7)}$. In this case $1/n^2g^8 = o(1/nNg^9)$ for the optimal g implies that $N = O(n^{(p_\theta+4)/(p_\theta+3)})$, which violates the theorem assumption $n = O(N^{(p_\theta+3)/(p_\theta+4)})$. On the hand, by assuming that $1/nNg^9 = O(1/n^2g^8)$, the optimal bandwidth is of order $n^{-1/(p_\theta+3)}$, a rate that satisfies $1/n^2g^8 = o(1/nNg^9)$. The corresponding MSE is of order $O(n^{-(2p_\theta-2)/(p_\theta+3)}) + O(n^{-1}) = O(n^{-(2p_\theta-2)/(p_\theta+3)})$ for $p_\theta \leq 5$.

For the estimation of I_α, it follows directly from Lemmas 4.2-4.3 of [21] that the bias and variance of $\hat{I}_\alpha$ are respectively of order $O(1/N)$ and $O(1/nN)$. Since $n = O(N^{(p_\theta+3)/(p_\theta+4)}) = o(N)$ by assumption, the corresponding MSE converges at the rate $O(1/nN)$, which is faster than the rate $O(n^{-(2p_\theta-2)/(p_\theta+3)})$ of the MSE of $\hat{\theta}_{2,2}$ for $p_\theta \leq 5$.

Now, from the Taylor approximation, we write

$$\hat{h}_{\text{opt}} \approx h_{\text{opt}} + \frac{\partial h_{\text{opt}}}{\partial I_\alpha}\left(\hat{I}_\alpha - I_\alpha\right) + \frac{\partial h_{\text{opt}}}{\partial \theta_{2,2}}\left(\hat{\theta}_{2,2}(g) - \theta_{2,2}\right)$$

so that

$$\frac{\hat{h}_{\text{opt}} - h_{\text{opt}}}{h_{\text{opt}}} \approx \frac{1}{h_{\text{opt}}}\frac{\partial h_{\text{opt}}}{\partial I_\alpha}\left(\hat{I}_\alpha - I_\alpha\right) + \frac{1}{h_{\text{opt}}}\frac{\partial h_{\text{opt}}}{\partial \theta_{2,2}}\left(\hat{\theta}_{2,2}(g) - \theta_{2,2}\right). \tag{A.20}$$

The partial derivatives of $h_{\text{opt}} = (C_1(K)I_\alpha)^{1/3}\left(2\mu_2^2\theta_{2,2}n\right)^{-1/3}$ with respect to I_α and $\theta_{2,2}$ are

$$\frac{\partial h_{\text{opt}}}{\partial I_\alpha} = \frac{1}{3}\,I_\alpha^{-2/3}\theta_{2,2}^{-1/3}\left(\frac{C_1(K)}{2\mu_2^2\,n}\right)^{1/3}, \quad \frac{\partial h_{\text{opt}}}{\partial \theta_{2,2}} = -\frac{1}{3}\,I_\alpha^{1/3}\theta_{2,2}^{-4/3}\left(\frac{C_1(K)}{2\mu_2^2\,n}\right)^{1/3}.$$

Given the convergence rates of the estimators $\hat{\theta}_{2,2}$ and $\hat{I}_\alpha$, we conclude that

$$\begin{aligned}\frac{\hat{h}_{\text{opt}} - h_{\text{opt}}}{h_{\text{opt}}} &\approx -\frac{1}{3}\theta_{2,2}^{-1}\left(\hat{\theta}_{2,2} - \theta_{2,2}\right)\\ &\approx \frac{D}{n^{(p_\theta-1)/(p_\theta+3)}}\,.\end{aligned}$$

□

Acknowledgements. The authors thank Giles Hooker (Cornell University) and an anonymous reviewer for their comments which led to substantial improvements in the paper.

References

[1] K. Benhenni and S. Cambanis, Sampling designs for estimating integrals of stochastic processes. *Ann. Statist.* **20** (1992) 161–194.

[2] K. Benhenni and M. Rachdi, Nonparametric estimation of the regression function from quantized observations. *Comput. Statist. Data Anal.* **50** (2006) 3067–3085.

[3] K. Benhenni and M. Rachdi, Nonparametric estimation of average growth curve with general nonstationary error process. *Comm. Statist. Theory Methods* **36** (2007) 1173–1186.

[4] S. Cambanis, Sampling designs for time series, in Time Series in the Time Domain. Edited by P.R. Krishnaiah E.J. Hannan and M.M. Rao, vol. 5 of *Handbook of Statistics*. Elsevier (1985) 337–362

[5] H. Cardot, Nonparametric estimation of smoothed principal components analysis of sampled noisy functions. *J. Nonparametr. Statist.* **12** (2000) 503–538.

[6] D. Degras, Asymptotics for the nonparametric estimation of the mean function of a random process. *Statist. Probab. Lett.* **78** (2008) 2976–2980.

[7] D. Degras, Simultaneous confidence bands for nonparametric regression with functional data. *Statist. Sinica* **21** (2011) 1735–1765.

[8] J. Fan and I. Gijbels, Local polynomial modelling and its applications. Vol. 66 of *Monogr. Stat. Appl. Probab.* Chapman & Hall, London (1996).

[9] J. Fan, I. Gijbels, T.C. Hu and L.S. Huang, A study of variable bandwidth selection for local polynomial regression. *Statist. Sinica* **6** (1996) 113–127.

[10] J. Fan and J.S. Marron, Fast implementations of nonparametric curve estimators. *J. Comput. Graph. Statist.* **3** (1994) 35–56.

[11] E. Ferreira, V. Núñez–Antón and J. Rodríguez–Póo, Kernel regression estimates of growth curves using nonstationary correlated errors. *Statist. Probab. Lett.* **34** (1997) 413–423.

[12] M. Francisco–Fernández, J. Opsomer and J. M. Vilar–Fernández, Plug-in bandwidth selector for local polynomial regression estimator with correlated errors. *J. Nonparametr. Stat.* **16** (2004) 127–151.

[13] M. Francisco–Fernández and J.M. Vilar–Fernández, Local polynomial regression estimation with correlated errors. *Comm. Statist. Theory Methods* **30** (2001) 1271–1293.

[14] P. Hall, S. Nath Lahiri and J. Polzehl, On bandwidth choice in nonparametric regression with both short- and long-range dependent errors. *Ann. Statist.* **23** (1995) 1921–1936.

[15] J.D. Hart and T.E. Wehrly, Kernel regression estimation using repeated measurements data. *J. Amer. Statist. Assoc.* **81** (1986) 1080–1088.

[16] J.D. Hart and T.E. Wehrly, Consistency of cross-validation when the data are curves. *Stoch. Process. Appl.* **45** (1993) 351–361.

[17] E. Masry, Local polynomial fitting under association. *J. Multivariate Anal.* **86** (2003) 330–359.
[18] E. Masry and J. Fan, Local polynomial estimation of regression functions for mixing processes. *Scand. J. Statist.* **24** (1997) 165–179.
[19] J. Opsomer, Y. Wang and Y. Yang, Nonparametric regression with correlated errors. *Statist. Sci.* **16** (2001) 134–153.
[20] A. Pérez–González, J.M. Vilar–Fernández and W. González–Manteiga, Asymptotic properties of local polynomial regression with missing data and correlated errors. *Ann. Inst. Statist. Math.* **61** (2009) 85–109.
[21] O. Perrin, Quadratic variation for Gaussian processes and application to time deformation. *Stoch. Process. Appl.* **82** (1999) 293–305.
[22] R. Core Team, *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria (2013).
[23] J.O. Ramsay and B.W. Silverman, Functional data analysis. *Springer Ser. Statist.*, 2nd edition. Springer, New York (2005).
[24] J.A. Rice and B.W. Silverman, Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B* **53** (1991) 233–243.
[25] D. Ruppert, Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation. *J. Amer. Statist. Assoc.* **92** (1997) 1049–1062.
[26] D. Ruppert, S.J. Sheather and M.P. Wand, An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.* **90** (1995) 1257–1270.
[27] M.P. Wand and M.C. Jones, Kernel smoothing. Vol. 60 of *Monogr. Statist. Appl. Probab.* Chapman and Hall Ltd., London (1995).
[28] F. Yao, Asymptotic distributions of nonparametric regression estimators for longitudinal or functional data. *J. Multivariate Anal.* **98** (2007) 40–56.