The reproducing kernel Hilbert space approach in nonparametric regression problems with correlated observations

D. Benelmadani[1] · K. Benhenni[1] · S. Louhichi[1]

Abstract

In this paper, we investigate the problem of estimating the regression function in models with correlated observations. The data are obtained from several experimental units, each of them forms a time series. Using the properties of the reproducing kernel Hilbert spaces, we construct a new estimator based on the inverse of the autocovariance matrix of the observations. We give the asymptotic expressions of its bias and its variance. In addition, we give a theoretical comparison between this new estimator and the popular one proposed by Gasser and Müller, we show that the proposed estimator has an asymptotically smaller variance than the classical one. Finally, we conduct a simulation study to investigate the performance and the robustness of the proposed estimator and to compare it to the Gasser and Müller's estimator in a finite sample set.

Keywords Nonparametric regression · Correlated observations · Growth curve · Reproducing kernel Hilbert space · Projection estimator · Asymptotic normality

1 Introduction

One of the situations that statisticians encounter in their studies is the estimation of a whole function based on partial observations of this function. For instance, in

✉ K. Benhenni
  karim.benhenni@univ-grenoble-alpes.fr

  D. Benelmadani
  djihad.benelmadani@univ-grenoble-alpes.fr

  S. Louhichi
  sana.louhichi@univ-grenoble-alpes.fr

1 Laboratoire Jean Kuntzmann (CNRS 5224), Université Grenoble Alpes, 700 Avenue Centrale, 38401 Saint-Martin-d'Hères, France

pharmacokinetics one wishes to estimate the concentration time of some injected medicine in the organism, based on the observations of the concentration from blood tests over a period of time. In statistical terms, one wants to estimate a function, say g, relating two random variables: the explanatory variable X and the response variable Y, without any parametric restrictions on the function g. The statistical model often used is the following: $Y_i = g(X_i) + \varepsilon_i$ where $(X_i, Y_i)_{1 \leq i \leq n}$ are n independent replicates of (X, Y) and $\{\varepsilon_i, i = 1, \ldots, n\}$ are centered random variables (called errors).

The most intensively treated model has been the one in which $(\varepsilon_i)_{1 \leq i \leq n}$ are independent errors and $(X_i)_{1 \leq i \leq n}$ are fixed within some domain. We mention the works of Priestly and Chao (1972), Benedetti (1977) and Gasser and Müller (1979) among others. However, the independence of the observations is not always a realistic assumption. For instance, the growth curve models are usually used in the case of longitudinal data, where the same experimental unit is being observed on multiple points of time. As a real life example, the heights observed on the same child are correlated. The temperature observations measured along the day are also correlated. For this, we focus, in this paper, on the nonparametric kernel estimation problem where the observations are correlated.

In the current paper, we consider a situation where the data are generated from m experimental units each of them having n measurements of the response. For this data, we consider the so-called fixed design regression model with repeated measurements given by,

$$Y_j(t_i) = g(t_i) + \varepsilon_j(t_i) \quad \text{for } i = 1, \ldots, n \text{ and } j = 1, \ldots, m, \tag{1}$$

where $\{\varepsilon_j, j = 1, \ldots, m\}$ is a sequence of i.i.d. centered error processes with the same distribution as a process ε. The non-correlation of the errors $\{\varepsilon_j, j = 1, \ldots, m\}$ is a natural assumption since it is equivalent to assuming that the experimental units (in general individuals) are independent.

This model is usually used in the growth curve analysis and dose response problems, see for instance, the work of Azzalini (1984). It has also been considered by Müller (1984) with $m = 1$, where he supposed that the observations are asymptotically uncorrelated when the number of observations tends to infinity, i.e., $\text{Cov}(\varepsilon(s), \varepsilon(t)) = O(1/n)$ for $s \neq t$, which is not a realistic assumption, for instance, in the growth curve analysis and temperature.

The correlated observations case was considered by Hart and Wherly (1986), who investigated the estimation of g in Model (1) where ε is a stationary error process. Using the kernel estimator proposed by Gasser and Müller (1979), they proved the consistency in $\mathbb{L}^2$ space of this estimator, when the number of experimental units m tends to infinity, but not when n tends to infinity as in the case of independent observations.

The assumption of stationarity made on the observations is however restrictive. In the previous pharmacokinetics example for instance, it is clear that the concentration of the medicine will be high at the beginning then decreases with time. For this, we shall investigate the estimation of g in Model (1) where ε is not necessarily a stationary error process. This case was partially investigated by Ferreira et al. (1997)

and by Benhenni and Rachdi (2007), where the Gasser and Müller's estimator was used.

In this paper, we propose a new estimator for the regression function g in Model (1). This estimator, which is also a linear kernel estimator, is based on the inverse of the autocovariance matrix of the observations that we assume known and invertible.

The proposed estimator was inspired by the work of Sacks and Ylvisaker (1966, 1968, 1970) but in a different context than ours. They considered the parametric model: $Y(t) = \beta f(t) + \varepsilon(t)$ where β is an unknown real parameter and f is a known function belonging to the reproducing kernel Hilbert space associated with the autocovariance function of the error process ε, denoted by RKHS(R). They also assumed that the autocovariance matrix is known and invertible. It is worth noting that the reproducing kernel Hilbert spaces have been used in several domains, for instance, in statistics by Sacks and Ylvisaker (1966) and more recently by Dette et al. (2016), in mathematical analysis by Schwartz (1964) and in signal processing by Ramsay and Silverman (2005).

We also give the asymptotic statistical performance of the proposed estimator, and we compare it to the classical Gasser and Müller's estimator (GM estimator), proving, in particular, that the proposed estimator outperforms the GM estimator, in the sense that it has an asymptotically smaller variance, whereas they both are asymptotically unbiased. This can be argued by the fact that, in statistics in general, the best linear estimator (or optimal predictor) is based on the inverse of the autocovariance matrix, see for instance, Benhenni and Cambanis (1992), whereas the GM estimator does not take into account this correlation requirement. In addition, the GM estimator is an approximation of an integral, and as known in statistics, the best linear approximation of an integral is based on some projection property.

This paper is organized as follows. In Sect. 2, we construct our proposed estimator for the function g in Model (1) where ε is a centered, second-order error process with a continuous autocovariance function R. It is constructed through the following function defined, for $x \in [0, 1]$, by,

$$f_{x,h}(t) = \int_0^1 R(s,t)\varphi_{x,h}(t)\mathrm{d}s \text{ where } \varphi_{x,h}(t) = \frac{1}{h}K\left(\frac{x-t}{h}\right) \quad \text{for } t \in [0, 1], \quad (2)$$

where K is a Kernel and $h = h(n, m)$ is a bandwidth.

We shall see that this function belongs to the RKHS(R). This allows us to use the properties of this space to control the variance of the proposed estimator. These properties were introduced by Parzen (1959) to solve various problems in statistical inference on time series. We also give, in this section, the analytical expressions of this estimator for the generalized Wiener process and the Ornstein–Uhlenbeck process, since the analytical expression of the inverse of the autocovariance matrix can be derived for this class of processes.

In Sect. 3, we derive the asymptotic performances of this estimator. We give an asymptotic expression of the weights of this linear estimator, which is used to derive the asymptotic expression of its bias. The properties of the RKHS(R) not only allow us to obtain the asymptotic expression of the variance, but also to find the optimal rate of convergence of the residual variance. After obtaining the asymptotic expression of the

integrated mean squared error (IMSE), we derive the asymptotic optimal bandwidth with respect to the IMSE criterion. Moreover, we prove the asymptotic normality of the proposed estimator.

In Sect. 4, we give a theoretical comparison between the new estimator and the Gasser and Müller's estimator. We prove that the proposed estimator has, asymptotically, a smaller variance than that of Gasser and Müller. Moreover, the proposed estimator has an asymptotically smaller IMSE, for instance, in the case of a Wiener process ε.

In Sect. 5, we conduct a simulation study in order to investigate the performance of the proposed estimator in a finite sample set; then we compare it with the Gasser and Müller's estimator for different values of the number of experimental units and different values of the sample size. We also study the robustness of the projection estimator, with respect to the misspecification of the autocovariance function. Since the classical cross-validation criterion is shown to be inefficient in the presence of correlation (see for instance, Altman 1990; Chiu 1989; Hart 1991, 1994), we use the optimal bandwidth that minimizes the exact IMSE, or the estimated mean average squared error (MASE) when the autocovariance function is unknown, obtained using the conjugated gradient algorithm. The results of this simulation study confirm our theoretical statements given in Sects. 3 and 4.

Finally, the supplementary materials section is dedicated to the proofs of the theoretical results, in addition to an Appendix about the RKHS(R) and some technical details.

2 Construction of the estimator using the RKHS approach

We consider Model (1) where g is the unknown regression function on $[0, 1]$ and $\{\epsilon_j(t), t \in [0, 1]\}_{j=1,\dots,m}$ is a sequence of error processes. We assume that $g \in C^2([0, 1])$ and that $(\varepsilon_j)_{j=1,\dots,m}$ are i.i.d. processes with the same distribution as a centered second-order process ε. We denote by R its autocovariance function, assumed to be known, continuous and forms a non-singular matrix when restricted to $T \times T$ for any finite set $T \subset [0, 1]$.

2.1 Projection estimator

In this section, we shall give the definition of the new proposed estimator for the regression function g in Model (1). This estimator (see Definition 1) is constructed using the function $f_{x,h}$ given by (2) for $x \in [0, 1]$, $h \in]0, 1[$ and K is a first-order kernel[1] of support $[-1, 1]$ belonging to C^1. This function is well known in time series analysis and has been used by several authors. We mention, among others, the works of Belouni and Benhenni (2015) and Sacks and Ylvisaker (1966) for linear regression models with correlated errors. It is mainly used due to its belonging to the reproducing kernel Hilbert space associated with the autocovariance function R (RKHS(R)) (see Appendix 1 for more details). This space is spanned by the functions $\{R(\cdot, t_i)_{1 \le i \le n}\}$

---

[1] The kernel K satisfies: $\int_{-1}^{1} K(t)dt = 1$, $\int_{-1}^{1} tK(t)dt = 0$ and $\int_{-1}^{1} t^2 K(t)dt < +\infty$.

forming a closed subspace on which an orthogonal projection of the function $f_{x,h}$ is feasible. We shall call the estimator obtained by this approach, the projection estimator.

The proposed estimator, which is a kernel estimator, is linear in the observations $\overline{Y}(t_i)$ and is given by the following definition.

Definition 1 The projection estimator of the regression function g in Model (1) based on the observations $(t_i, Y_j(t_i))_{\substack{1 \le i \le n \\ 1 \le j \le m}}$ is given for any $x \in [0, 1]$ by,

$$\hat{g}_n^{\text{pro}}(x) = \sum_{i=1}^{n} m_{x,h}(t_i)\overline{Y}(t_i), \tag{3}$$

where $\overline{Y}(t_i) = \frac{1}{m} \sum_{j=1}^{m} Y_j(t_i)$ and the weights $(m_{x,h}(t_i))_{1 \le i \le n}$ are being determined, letting $T_n = (t_i)_{1 \le i \le n}$, by,

$$m'_{x,h_{|T_n}} = f_{x,h_{|T_n}}' R_{|T_n}^{-1}, \tag{4}$$

with $f_{x,h_{|T_n}} := (f_{x,h}(t_1), \ldots, f_{x,h}(t_n))'$, $R_{|T_n} := (R(t_i, t_j))_{1 \le i, j \le n}$, $R_{|T_n}^{-1}$ the inverse of $R_{|T_n}$ and $m_{x,h_{|T_n}} := (m_{x,h}(t_1), \ldots, m_{x,h}(t_n))'$, where v' denotes the transpose of a vector v.

Remark 1 The assumption that the autocovariance function is known and that the autocovariance matrix is invertible, is well known in the literature. For instance, it has been used in the linear regression models, see for example Sacks and Ylvisaker (1966, 1968, 1970), in order to construct the best linear unbiased estimator (BLUE) of the model's parameter. Likewise, this assumption was used to obtain the best linear predictor of a process at a given value, and the best linear predictor of the integral stochastic process, see for instance Benhenni and Cambanis (1992) and Su and Cambanis (1993).

Remark 2 In order to motivate the proposed estimator, consider the regression model using m continuous experimental units, i.e.,

$$Y_j(t) = g(t) + \varepsilon_j(t) \text{ for } t \in [0, 1] \text{ and } j = 1, \ldots, m. \tag{5}$$

A continuous kernel estimator of g in Model (5) is given for any $x \in [0, 1]$ by,

$$\hat{g}_{[0,1]}(x) = \int_0^1 \varphi_{x,h}(t)\overline{Y}(t)\mathrm{d}t \quad \text{with} \quad \overline{Y}(t) = \frac{1}{m} \sum_{j=1}^{m} Y_j(t), \tag{6}$$

where $\varphi_{x,h}(t) = \frac{1}{h} K\left(\frac{x-t}{h}\right)$ for a kernel K and a bandwidth h. We refer the reader to the works of Blanke and Bosq (2008) and Didi and Louani (2013) for more details on the Kernel estimation of the regression function based on continuous observations.

Since in practice we only have access to discrete observations, then a linear approximation of the continuous kernel estimator should be of the form:

$$\hat{g}_n(x) = \sum_{i=1}^{n} W_{x,h}(t_i)\overline{Y}(t_i).$$

Now let,

$$f_{n,x}(t) = \sum_{i=1}^{n} W_{x,h}(t_i) R(t_i, t) \quad \text{for } t \in [0, 1].$$

Then the mean squared error (MSE) of approximation can be written as:

$$\mathbb{E}\left(\hat{g}_{[0,1]}(x) - \hat{g}_n(x)\right)^2 = ||f_{x,h} - f_{n,x}||^2,$$

where $f_{x,h}$ is given by (2) and $|| \cdot ||$ is the norm of the RKHS(R) (see Appendix for more details). Then the best linear predictor $\hat{g}_n^{\text{pro}}(x)$ of $\hat{g}_{[0,1]}(x)$ satisfies:

$$\inf_{W_{x,h|T_n}} \mathbb{E}\left(\hat{g}_{[0,1]}(x) - \hat{g}_n(x)\right)^2 = ||f_{x,h} - P_{|T_n} f_{x,h}||^2,$$

where $P_{|T_n} f_{x,h}$ is the orthogonal projection of $f_{x,h}$ on the subspace of RKHS spanned by the function $\{R(\cdot, t_i), i = 1, \ldots, n\}$. The optimal coefficients $(W_{x,h}^*(t_i))_{1 \le i \le n}$ can then be derived by using the fact that $P_{|T_n} f_{x,h}(t_i) = f_{x,h}(t_i)$ for $i = 1, \ldots, n$ (see Equation (88) in Appendix) and this yields $W_{x,h|T_n}^{*}{}' = f_{x,h|T_n}{}' R_{|T_n}^{-1}$.

For some classical error processes, such as the Wiener and the Ornstein–Uhlenbeck processes, the estimator (3) has a simplified expression as shown in the following proposition.

Proposition 1 *Consider the regression model* (1) *where ε is of autocovariance function $R(s, t) = \int_0^{\min(s,t)} u^\beta du$ for a positive constant β. Let $t_0 = 0$, $t_{n+1} = 1$. Set $\overline{Y}(t_0) = 0$ and $\overline{Y}(t_{n+1}) = \overline{Y}(t_n)$. For any $x \in [0, 1]$, the projection estimator* (3) *can be written as follows:*

$$\hat{g}_n^{\text{pro}}(x) = \frac{1}{\beta + 1} \left(\sum_{i=1}^{n+1} \overline{Y}(t_i) \int_{t_{i-1}}^{t_i} \varphi_{x,h}(s) ds \right.$$

$$\left. + \sum_{i=0}^{n-1} \frac{\overline{Y}(t_{i+1}) - \overline{Y}(t_i)}{t_{i+1}^{\beta+1} - t_i^{\beta+1}} \int_{t_i}^{t_{i+1}} \left(s^{\beta+1} - t_{i+1}^{\beta+1} \right) \varphi_{x,h}(s) ds \right). \qquad (7)$$

Remark 3 Taking $\beta = 0$ in the previous proposition gives the expression of the projection estimator (3) in the case where ε is the classical standard Wiener error process.

Proposition 2 *If the error process ε in Model* (1) *is the Ornstein–Uhlenbeck process with $R(s,t) = e^{-|t-s|}$, then for any $x \in [0,1]$,*

$$
\hat{g}_n^{\mathrm{pro}}(x) = \sum_{i=2}^{n-1} \overline{Y}(t_i) \int_{t_{i-1}}^{t_{i+1}} e^{|s-t_i|} \varphi_{x,h}(s)ds + \overline{Y}(t_1) \int_0^{t_2} e^{s-t_1} \varphi_{x,h}(s)ds
$$

$$
+ \overline{Y}(t_n) \int_{t_{n-1}}^1 e^{t_n-s} \varphi_{x,h}(s)ds - \sum_{i=1}^{n-1} \frac{e^{t_{i+1}}\overline{Y}(t_{i+1}) - e^{t_i}\overline{Y}(t_i)}{1 - e^{-2(t_{i+1}-t_i)}} \int_{t_i}^{t_{i+1}} e^{-s} \varphi_{x,h}(s)ds
$$

$$
+ \sum_{i=1}^{n-1} \frac{e^{-t_{i+1}}\overline{Y}(t_{i+1}) - e^{-t_i}\overline{Y}(t_i)}{1 - e^{-2(t_{i+1}-t_i)}} \int_{t_i}^{t_{i+1}} e^s \varphi_{x,h}(s)ds,
$$

where $\varphi_{x,h}$ is defined in the previous proposition.

Remark 4 As the previous propositions show, the expression of $m_{x,h|T_n}$ is known analytically for error processes of practical interest. For more complicated error processes, numerical methods can be used. For more general error processes, we will give an asymptotic simplified expression of the weights of the projection estimator (see Lemma 3).

2.2 Assumptions and comments

In order to derive our asymptotic results, the following assumptions on the autocovariance function R and the Kernel K are required.

(A) R is continuous on the entire unit square and has left and right derivatives up to order two at the diagonal (i.e., when $s = t$), i.e.,

$$
R^{(0,1)}(t, t^-) = \lim_{s \uparrow t} \frac{\partial R(t,s)}{\partial s} \quad \text{and} \quad R^{(0,1)}(t, t^+) = \lim_{s \downarrow t} \frac{\partial R(t,s)}{\partial s},
$$

exist and are continuous. In a similar way we define $R^{(0,2)}(t, t^-)$ and $R^{(0,2)}(t, t^+)$. Off the diagonal (i.e., when $s \neq t$ in the unit square), R has continuous derivatives up to order two.

For $t \in]0, 1[$, let $\alpha(t) = R^{(0,1)}(t, t^-) - R^{(0,1)}(t, t^+)$. Assumption (A) gives the following lemma concerning the jump function α.

Lemma 1 *If Assumption* (A) *is satisfied, then the jump function α is a positive function.*

To obtain our asymptotic results, we shall give next a stronger assumption on the jump function α.

(B) We assume that α is Lipschitz on $]0, 1[$, $\inf_{0 < t < 1} \alpha(t) = \alpha_0 > 0$ and $\sup_{0 < t < 1} \alpha(t) = \alpha_1 < \infty$.

Assumptions (A) and (B) are classical regularity conditions and were used in several works, see for instance Belouni and Benhenni (2015), Sacks and Ylvisaker (1966) and Su and Cambanis (1993).

(C) For each $t \in [0, 1]$, $R^{(0,2)}(., t^+)$ is in the reproducing kernel Hilbert space associated with R, denoted by RKHS(R), equipped with the norm $|| \cdot ||$. In addition, $\sup\limits_{0 \leq t \leq 1} ||R^{(0,2)}(., t^+)|| < \infty$ (see Appendix for more details).

Assumption (C), which is more restrictive than (B) as indicated by Sacks and Ylvisaker (1966), is necessary to evaluate the weights of the projection estimator (see Lemma 3).

(D) K is an even function and K' is a Lipschitz function on $[-1, 1]$.

Examples of autocovariance functions which satisfy Assumptions (A), (B) and (C) are given below.

Example 1 1. The autocovariance function $R(s, t) = \sigma^2 min(s, t)$ of the Wiener process has a constant jump function $\alpha(t) = \sigma^2$ and $R^{(i,j)}(s, t) = 0$ for all integers i, j such that $i + j = 2$ and $s \neq t$.
2. The autocovariance function $R(s, t) = \sigma^2 e^{-\lambda|s-t|}$ of the stationary Ornstein–Uhlenbeck process with $\sigma > 0$ and $\lambda > 0$. For this process the jump function is $\alpha(t) = 2\sigma^2\lambda$ and $R^{(0,2)}(s, t) = \sigma^2\lambda^2 e^{-\lambda|s-t|}$.
3. Another general class of autocovariance functions was given by Sacks and Ylvisaker (1966) and has the form,

$$R(s, t) = \int_0^{1/|t-s|} (1 - \mu|t - s|)p(\mu)d\mu,$$

where p is a probability density and p' its derivative are such that,

$$\lim_{\mu \to \infty} \mu^3 p(\mu) < \infty, \quad \text{and} \quad \int_a^\infty (\mu p'(\mu) + 3p(\mu))^2)\mu^6 d\mu < \infty,$$

for some a. We have $\alpha(t) = 2 \int_0^\infty u p(u) du$.

3 Local asymptotic results

Let $T_n = (t_{i,n})_{1 \leq i \leq n}$ for $n \geq 1$, be a fixed sequence of designs with $T_n \in D_n$, where

$$D_n = \{(s_1, s_2, \ldots, s_n) \ : \ 0 \leq s_1 < s_2 < \cdots < s_n \leq 1\}.$$

Set $t_{0,n} = 0$, $t_{n+1,n} = 1$, $d_{j,n} = t_{j+1,n} - t_{j,n}$ and let for $x \in [0, 1]$, $h = h(n, m)$,

$$I_{x,h} = \{i = 1, \ldots, n \ : \ [t_{i-1,n}, t_{i+1,n}]\cap]x - h, x + h[\neq \emptyset\}.$$

Denote by $N_{T_n} = \text{Card}(I_{x,h})$. Recall that $[x - h, x + h]$ is the support of the function $\varphi_{x,h}$. To obtain the asymptotic results, we require that the sequence $(T_n)_{n \geq 1}$ satisfies the next assumption.

(E) $\lim_{n\to\infty} \sup_{0\le j\le n} d_{j,n} = 0$, $\lim_{n,m\to\infty} \left(\frac{1}{h} \sup_{0\le j\le n} d_{j,n}\right) = 0$, $\lim_{n,m\to\infty}$
$\left(N_{T_n} \frac{1}{h^2} \sup_{0\le j\le n} d_{j,n}^2\right) = 0$ and $\lim\sup_{n,m\to\infty} \left(N_{T_n}^2 \frac{1}{h^2} \sup_{0\le j\le n} d_{j,n}^2\right) < \infty$.

A simple sequence of designs that verifies Assumption (E) was presented by Sacks and Ylvisaker (1970) as follows.

Definition 2 Let F be a distribution function of some density function f such that $\sup_{0<t<1} f(t) < \infty$ and $\inf_{0<t<1} f(t) > 0$. The so-called regular sequence of designs generated by f is defined by,

$$T_n = \left\{ t_{i,n} = F^{-1}\left(\frac{i}{n}\right), i = 1, \ldots, n \right\}.$$

In the sequel, the density f is assumed to be at least in $C^2([0, 1])$. This sequence of designs verifies the following Lemma (see for instance Benelmadani et al. 2019b for its proof).

Lemma 2 *Let $(T_n)_{n\ge 1}$ be a regular sequence of designs generated by some density function. For $x \in]0, 1[$ and $h > 0$, suppose that $T_n \cap [x - h, x + h] \ne \emptyset$ and that $nh \ge 1$. Then,*

$$\sup_{0\le j\le n} d_{j,n} = O\left(\frac{1}{n}\right) \quad and \quad N_{T_n} = O(nh), \tag{8}$$

where N_{T_n} and $d_{j,n}$ are defined as above. In addition, if $\lim_{n,m\to\infty} nh = \infty$, then the regular sequence of designs verifies Assumption (E).

3.1 Evaluation of the bias

In order to derive the asymptotic expression of the bias term of the projection estimator, we shall first give the asymptotic approximation of the weights $m_{x,h|T_n}$ [defined by (4)] in the following lemma.

Lemma 3 *Suppose that Assumptions (A), (B) and (C) are satisfied. Then for any $x \in]0, 1[$,*

$m_{x,h}(t_{i,n})$
$$= \begin{cases} \frac{1}{2}\varphi_{x,h}(t_{i,n})(t_{i+1,n} - t_{i-1,n}) + O\left(\alpha_{n,h} + \beta_{n,h}\right) & if \ \ i \notin \{1, n\} \ and \\ [t_{i-1,n}, t_{i+1,n}] \cap [x - h, x + h] \ne \emptyset, \\ O\left(N_{T_n}\alpha_{n,h} + n\beta_{n,h}\right) & if \ \ i \in \{1, n\}, \\ O\left(\beta_{n,h}\right) & otherwise, \end{cases}$$

where

$$\alpha_{n,h} = \sup_{0 \le i \le n} \sup_{t_{i,n} \le s, t \le t_{i+1,n}} d_{i,n} \, |\alpha(s)\varphi_{x,h}(s) - \alpha(t)\varphi_{x,h}(t)| = O\left(\frac{1}{h^2} \sup_{0 \le j \le n} d_j^2\right),$$

$$\beta_{n,h} = \sup_{0 \le t \le 1} \frac{1}{\alpha(t)} \|R^{(0,2)}(.,t)\| \frac{\sqrt{C}}{\sqrt{h}} \sup_{0 \le j \le n} d_j^2 = O\left(\frac{1}{\sqrt{h}} \sup_{0 \le j \le n} d_j^2\right),$$

and C is a positive constant defined in Proposition 5.

Remark 5 This Lemma shows that the weights of the projection estimator are asymptotically equivalent to those of some well-known linear estimators of the regression function g. For instance,

- Priestly and Chao (1972) and Benedetti (1977) used the following weights:

$$W_{x,h}(t_i) = (t_{i+1,n} - t_{i,n})\varphi_{x,h}(t_i) \quad \text{for } i = 1, \ldots, n.$$

- Gasser and Müller (1979) used the following weights:

$$W_{x,h}(t_i) = \int_{s_{i-1,n}}^{s_{i,n}} \varphi_{x,h}(s)\mathrm{d}s \quad \text{for } i = 1, \ldots, n,$$

  where $s_0 = 0$, $s_n = 1$ and $s_{i,n} = (t_{i+1,n} + t_{i,n})/2$ for $i = 1, \ldots, n-1$.
- Cheng and Lin (1981) replaced $s_{i,n}$ by $t_{i,n}$, in the weights of the Gasser and Müller estimator.

Using the asymptotic approximation of the weights given in Lemma 3, we can obtain the asymptotic expression of the bias of the projection estimator as shown in the following proposition.

Proposition 3 *Suppose that Assumptions (A)–(D) are satisfied. If $T_n \cap]x - h, x + h[\ne \emptyset$ and $nh \ge 1$, then for any $x \in]0, 1[$,*

$$\mathbb{E}(\hat{g}_n^{\mathrm{pro}}(x)) - g(x) = \frac{1}{2}h^2 g''(x)B + o(h^2) + O\left(\frac{N_{T_n}}{h^3} \sup_{0 \le j \le 1} d_{j,n}^3 + N_{T_n}\alpha_{n,h} + n\beta_{n,h}\right),$$

where $\alpha_{n,h}$ and $\beta_{n,h}$ are given in Lemma 3 and $B = \int_{-1}^{1} t^2 K(t)dt$.

Remark 6 Under the assumption of Lemma 2 we have,

$$\mathbb{E}(\hat{g}_n^{\mathrm{pro}}(x)) - g(x) = \frac{1}{2}h^2 g''(x)B + o(h^2) + O\left(\frac{1}{nh}\right).$$

In the case of a Wiener error process, a direct computation of the bias term of the projection estimator (7), with $\beta = 0$, shows that the order term $O\left(\frac{1}{nh}\right)$ can be improved. The result is given by the following proposition.

Proposition 4 *Consider Model* (1) *with a Wiener error process of autocovariance function* $R(s, t) = \min(s, t)$. *Let* $(T_n)_{n \geq 1}$ *be a regular sequence of designs generated by a density function* f *(cf. Definition* 2*) and let* K *be a kernel satisfying Assumption* (D). *If* $T_n \cap]x - h, x + h[\neq \emptyset$ *and* $nh \geq 1$, *then*

$$\mathbb{E}\,(\hat{g}_n^{\text{pro}}(x)) - g(x) = \frac{1}{2}h^2 g''(x)B + o(h^2) + O\left(\frac{1}{n^2 h}\right),$$

where B *is given in Proposition* 3 *above.*

3.2 Evaluation of the variance

It is shown in Lemma 5 of Appendix that $f_{x,h}$ defined by (2) belongs to the RKHS(R) equipped with its norm $|| \; ||$, and

$$||f_{x,h}||^2 = \int_0^1 \int_0^1 \varphi_{x,h}(s)R(s, t)\varphi_{x,h}(t)\mathrm{d}s\;\mathrm{d}t \overset{\Delta}{=} \sigma_{x,h}^2. \tag{9}$$

In addition if $P_{|T_n} f_{x,h}$ is the projection of $f_{x,h}$ on the subspace of $\mathcal{F}$ spanned by $\{R(., t), t \in T_n\}$ then it is shown by (F2) in the supplementary facts of Appendix that

$$||P_{|T_n} f_{x,h}||^2 = m\,\text{Var}\,\hat{g}_n^{\text{pro}}(x). \tag{10}$$

The following proposition controls the residual variance $\frac{\sigma_{x,h}^2}{m} - \text{Var}\,\hat{g}_n^{\text{pro}}(x)$.

Proposition 5 *Suppose that Assumptions* (A) *and* (B) *are satisfied. Moreover, assume that* $\frac{1}{h}\sup_{1 \leq i \leq n} d_i \leq 1$ *and let,*

$$K_\infty = \sup_{t \in [-1,1]} |K(t)|, \quad R_1 = \sup_{t,s \in [0,1]} |R^{(1,1)}(s-, t+)| \; and \; R_2 = \sup_{t,s \in [0,1]} |R^{(0,2)}(s, t+)|.$$

Then we have, for any $x \in]0, 1[$,

$$0 \leq \frac{\sigma_{x,h}^2}{m} - \text{Var}\,\hat{g}_n^{\text{pro}}(x) \leq \frac{C}{mh} \sup_{0 \leq j \leq n} d_{j,n}^2,$$

where $C = \begin{cases} K_\infty^2(\frac{4}{3}\alpha_1 + R_1 + \frac{4}{3}R_2) & \text{if } (x - h) \text{ and } (x + h) \in T_n, \\ K_\infty^2(\frac{8}{3}\alpha_1 + \frac{5}{3}R_1 + \frac{8}{3}R_2) & \text{otherwise.} \end{cases}$

If moreover $\{T_n, n \geq 1\}$ satisfies Assumption (E) then Proposition 5 gives,

$$\lim_{n,m \to \infty} \left(\text{Var}\,\hat{g}_n^{\text{pro}}(x) - \frac{\sigma_{x,h}^2}{m}\right) = 0.$$

The next proposition gives the rate of convergence of this residual variance.

Proposition 6 *Suppose that Assumptions (A), (B) and (C) are satisfied. Moreover, assume that $(T_n)_{n\geq 1}$ is a sequence of designs verifying Assumption (E). Then for any $x \in]0, 1[$ and for any positive integer m,*

$$\varliminf_{n\to\infty} \frac{mN_{T_n}^2}{h}\left(\frac{\sigma_{x,h}^2}{m} - \text{Var } \hat{g}_n^{\text{pro}}(x)\right) \geq \frac{1}{12}\alpha(x)\left\{\int_{-1}^{1} K^{2/3}(t)dt\right\}^3, \quad (11)$$

where $\sigma_{x,h}^2$ is given by (9).

Using Propositions 5 and 6 we can obtain the optimal convergence rate $1/(mn^2h)$ of the residual variance. The result is given by the following proposition.

Proposition 7 *Suppose that all the assumptions of Lemma 2, Propositions 5 and 6 are satisfied. Then there exist some positive constants C and C' such that for any $x \in]0, 1[$ and for any positive integer m,*

$$\varlimsup_{n\to\infty} mn^2h\left(\frac{\sigma_{x,h}^2}{m} - \text{Var}(\hat{g}_n^{\text{pro}}(x))\right) \leq C, \quad (12)$$

and

$$\varliminf_{n\to\infty} mn^2h\left(\frac{\sigma_{x,h}^2}{m} - \text{Var } \hat{g}_n^{\text{pro}}(x)\right) \geq C'. \quad (13)$$

Under the stronger assumption (D) on the kernel K and using a regular sequence of designs (see Definition 2), we obtain the asymptotic expression of the variance as shown by the following proposition.

Proposition 8 *Suppose that Assumptions $(A)-(D)$ are satisfied. Moreover assume that $(T_n)_{n\geq 1}$ is a regular sequence of designs generated by a density function f (see Definition 2). If $\lim_{n,m\to\infty}h = 0$ and $\lim_{n,m\to\infty}nh = \infty$, then for any $x \in]0, 1[$,*

$$\text{Var}(\hat{g}_n^{\text{pro}}(x)) = \frac{\sigma_{x,h}^2}{m} - \frac{1}{12mn^2}\int_{x-h}^{x+h}\frac{\alpha(t)}{f^2(t)}\varphi_{x,h}^2(t)dt + O\left(\frac{1}{mn^3h^2}\right), \quad (14)$$

where $\sigma_{x,h}^2$ is given by (9).

The following lemma (proved in Benhenni and Rachdi 2007) gives the expression of the main term of the asymptotic variance $\sigma_{x,h}^2/m$ in terms of h.

Lemma 4 *Suppose that Assumptions (A), (B) and (D) are satisfied. If $\lim_{n,m\to\infty}h = 0$, then for any $x \in]0, 1[$, $\sigma_{x,h}^2$ (as given by* (9)*) has the following asymptotic expression*

$$\sigma_{x,h}^2 = \left(R(x, x) - \frac{1}{2}\alpha(x)C_Kh\right) + o(h), \quad (15)$$

where $C_K = \int_{-1}^{1}\int_{-1}^{1}|u - v|K(u)K(v)dudv$.

3.3 IMSE and optimal bandwidth

Proposition 8 and Remark 6 allow to derive the asymptotic expression of the mean squared error (MSE) of the projection estimator (3). The integrated mean squared error (IMSE) is then obtained by integrating the MSE with respect to a positive density function w. The MSE and IMSE are given, without proof, in the next theorem.

Theorem 1 *If all the assumptions of Propositions 3 and 8 are satisfied and if $(T_n)_{n \geq 1}$ is a regular sequence of designs generated by some density function (see Definition 2), then for any $x \in]0, 1[$,*

$$
\mathrm{MSE}(\hat{g}_n^{\mathrm{pro}}(x)) = \frac{1}{m} \left(R(x, x) - \frac{1}{2}\alpha(x)C_K h \right) + \frac{1}{4}h^4(g''(x))^2 B^2 + o\left(h^4 + \frac{h}{m} \right)
$$
$$
+ O\left(\frac{1}{mn^2h} + \frac{h}{n} + \frac{1}{n^2h^2} \right),
$$
$$
\mathrm{IMSE}(\hat{g}_n^{\mathrm{pro}}) = \frac{1}{m} \int_0^1 R(x, x)w(x)dx - \frac{C_K h}{2m} \int_0^1 \alpha(x)w(x)dx
$$
$$
+ \frac{B^2}{4}h^4 \int_0^1 [g''(x)]^2 w(x)dx + o\left(h^4 + \frac{h}{m} \right)
$$
$$
+ O\left(\frac{1}{mn^2h} + \frac{h}{n} + \frac{1}{n^2h^2} \right),
$$

where w is a positive density function, B is given in Proposition 3 and C_K is given in Lemma 4.

Remark 7 We note here that the term $\frac{1}{12mn^2} \int_{x-h}^{x+h} \frac{\alpha(t)}{f^2(t)} \varphi_{x,h}^2(t)dt$ appearing in the asymptotic variance does not appear in the asymptotic MSE and IMSE, because it is negligible comparing to the squared bias, precisely due to the term $O\left(\frac{1}{nh} \right)$.

However in the case of a Wiener error process, we have proven (see Proposition 4) that the previous term can be replaced by $O\left(\frac{1}{n^2h} \right)$ when using exact weights of the projection estimator (and not their asymptotic expression). Therefore, when ε is a Wiener process, the asymptotic expressions of the MSE and IMSE of the projection estimator (7) (with $\beta = 0$) are given by the following theorem.

Theorem 2 *Consider Model (1) with a Wiener error process and suppose that the kernel K verifies Assumption (D). Moreover, assume that $(T_n)_{n \geq 1}$ is a regular sequence of designs generated by a function f (see Definition 2). If $\lim_{n,m \to \infty} h = 0$ and $\lim_{n,m \to \infty} nh = \infty$, then for any $x \in]0, 1[$,*

$$
\mathrm{MSE}(\hat{g}_n^{\mathrm{pro}}(x)) = \frac{1}{m} \left(R(x, x) - \frac{1}{2}\alpha(x)C_K h \right) - \frac{1}{mn^2h} \frac{\alpha(x)}{f^2(x)} \int_{-1}^1 K^2(t)dt
$$
$$
+ \frac{1}{4}h^4[g''(x)]^2 B^2 + o\left(\frac{h}{m} + h^4 \right) + O\left(\frac{h}{n^2} + \frac{1}{mn^3h^2} + \frac{1}{mn^2} + \frac{1}{n^4h^2} \right),
$$

and

$$\text{IMSE}(\hat{g}_n^{\text{pro}}) = \frac{1}{m} \int_0^1 R(x,x)w(x)\mathrm{d}x - \frac{C_K h}{2m} \int_0^1 \alpha(x)w(x)\mathrm{d}x$$

$$- \frac{A}{12mn^2 h} \int_0^1 \frac{\alpha(x)}{f^2(x)} w(x)\mathrm{d}x + \frac{B^2}{4} h^4 \int_0^1 [g''(x)]^2 w(x)\mathrm{d}x + o\left(\frac{h}{m} + h^4\right)$$

$$+ O\left(\frac{h}{n^2} + \frac{1}{mn^3 h^2} + \frac{1}{mn^2} + \frac{1}{n^4 h^2}\right),$$

where $A = \int_{-1}^1 K^2(t)\mathrm{d}t$, w, B and C_K are given in Theorem 1.

The asymptotic optimal bandwidth is obtained by minimizing the asymptotic IMSE and is given in the following corollary.

Corollary 1 (Optimal bandwidth) *Suppose that the assumptions of Theorem* 1 *are satisfied. Moreover assume that $\frac{n}{m} = O(1)$ as $n, m \to \infty$. Denote by IMSE(h) the IMSE of the projection estimator when the bandwidth h is used. Then the bandwidth,*

$$h^* = \left(\frac{C_K \int_0^1 \alpha(x)w(x)\mathrm{d}x}{2B \int_0^1 [g''(x)]^2 w(x)\mathrm{d}x}\right)^{1/3} m^{-1/3}, \tag{16}$$

is optimal in the sense that

$$\varlimsup_{n,m \to \infty} \frac{\text{IMSE}(h^*)}{\text{IMSE}(h_{n,m})} \le 1,$$

for any sequence of bandwidths $h_{n,m}$ verifying:

$$\lim_{n,m \to \infty} h_{n,m} = 0 \quad and \quad \varlimsup_{n,m \to \infty} mh_{n,m}^3 < +\infty.$$

3.4 Asymptotic normality

The next theorem presents the asymptotic normality of the projection estimator (3) for any error process ε.

Theorem 3 *Suppose that the assumptions of Theorem* 1 *are satisfied. Moreover assume that $\frac{n}{m} = O(1)$ as $n, m \to \infty$, that $\lim_{n,m \to \infty} nh^2 = \infty$ and that $\lim_{n,m \to \infty} \sqrt{m} h^2 = 0$. Then for any $x \in]0, 1[$,*

$$\sqrt{m}\left(\hat{g}_n^{\text{pro}}(x) - g(x)\right) \xrightarrow{\mathscr{D}} Z \quad with \ Z \sim \mathcal{N}(0, R(x,x)) \quad as \ n, m \to \infty,$$

where $\mathscr{D}$ denotes the convergence in distribution and $\mathcal{N}$ is the normal distribution.

4 Comparison with the Gasser and Müller's estimator

In this section, we shall perform a theoretical comparison between the projection estimator given in (3) and the classical estimator proposed by Gasser and Müller (1979), for i.i.d. observations, and used by Hart and Wherly (1986) for correlated observations in a time series setting. We recall its definition below.

Definition 3 The Gasser and Müller's estimator of the regression function g based on the observations $(t_i, Y_j(t_i))_{\substack{1 \leq i \leq n \\ 1 \leq j \leq m}}$ is given for any $x \in [0, 1]$ by,

$$\hat{g}_n^{\mathrm{GM}}(x) = \sum_{i=1}^{n} \overline{Y}(t_i) \int_{s_{i-1}}^{s_i} \varphi_{x,h}(s)\mathrm{d}s , \tag{17}$$

where $\overline{Y}$, $\varphi_{x,h}$ and h are given in Definition 1. The midpoints $(s_i)_{1 \leq i \leq n}$ are such that: $s_0 = 0$, $s_n = 1$ and for $i = 1, \ldots, n-1$, $s_i = (t_i + t_{i+1})/2$.

In order to compare this estimator to the projection estimator with respect to the IMSE, we recall in the next theorem the asymptotic expression of the IMSE of the Gasser and Müller's estimator (for the proof see Benhenni and Rachdi 2007; Benelmadani 2019a).

Theorem 4 *Suppose that Assumptions (A), (B) and (D) are satisfied. Moreover assume that $(T_n)_{n \geq 1}$ is a regular sequence of designs generated by a density function f (see Definition 2). If $\lim_{n,m \to \infty} h = 0$ and $\lim_{n,m \to \infty} nh = \infty$, then for any $x \in]0, 1[$,*

$$\mathrm{MSE}(\hat{g}_n^{\mathrm{GM}}(x)) = \frac{1}{m}\left(R(x,x) - \frac{1}{2}\alpha(x)C_K h\right) + \frac{1}{4}h^4(g''(x))^2 B^2 + o\left(h^4 + \frac{h}{m}\right)$$
$$+ O\left(\frac{h}{n^2} + \frac{1}{n^4h^2} + \frac{1}{mn^3h^2} + \frac{1}{mn^2}\right),$$

and

$$\mathrm{IMSE}(\hat{g}_n^{\mathrm{GM}}) = \frac{1}{m}\int_0^1 R(x,x)w(x)\mathrm{d}x - \frac{C_K h}{2m}\int_0^1 \alpha(x)w(x)\mathrm{d}x$$
$$+ \frac{B^2}{4}h^4\int_0^1 [g''(x)]^2 w(x)\mathrm{d}x$$
$$+ o\left(h^4 + \frac{h}{m}\right) + O\left(\frac{h}{n^2} + \frac{1}{n^4h^2} + \frac{1}{mn^3h^2} + \frac{1}{mn^2}\right),$$

where B and C_K are given in Propositions 3 and 8 and w is a continuous positive density.

The following theorem gives an asymptotic comparison in term of the variance of the projection estimator (3) and the Gasser and Müller's estimator (17).

Theorem 5 *Suppose that Assumptions* $(A), (B)$ *and* (D) *are satisfied. Moreover assume that* $(T_n)_{n\geq 1}$ *is a regular sequence of designs generated by a density function* f *(see Definition* 2*). If* $\lim_{n,m\to\infty} h = 0$ *and* $\lim_{n,m\to\infty} nh = \infty$, *then for any* $x \in]0, 1[$,

$$\lim_{n,m\to\infty} mn^2h \left(\text{Var } \hat{g}_n^{\text{GM}}(x) - \text{Var } \hat{g}_n^{\text{pro}}(x)\right) = \frac{1}{12}\frac{\alpha(x)}{f^2(x)} > 0.$$

For a comparison of the bias of these estimators, we mention that the Gasser and Müller's estimator converges to zero slightly faster than the bias of the projection estimator, i.e., the term $O(\frac{1}{nh})$ in the bias of the projection estimator (see Remark 6) is replaced by $O(\frac{1}{n^2h})$ in the bias of the Gasser and Müller's estimator (see Benelmadani 2019a). However, for the Wiener error process both estimators have the same bias convergence rates; thus, we can compare the asymptotic IMSE of both estimators in the following theorem.

Theorem 6 *Consider Model* (1) *where* ε *is a Wiener error process. Suppose that the assumptions of Theorem* 2 *are satisfied. Moreover, assume that* $\lim_{n,m\to\infty} nh^2 = 0$ *and that* $\frac{m}{n} = O(1)$ *then,*

$$\lim_{n,m\to\infty} mn^2h \, (\text{IMSE } (\hat{g}_n^{\text{GM}}) - \text{IMSE } (\hat{g}_n^{\text{pro}})) = \frac{\sigma^2}{12} \int_0^1 \frac{w(x)}{f^2(x)} \, dx > 0.$$

Remark 8 Theorems 5 and 6 show that the projection estimator has an asymptotically smaller variance than the Gasser and Müller's estimator for any error process, it also has an asymptotically smaller IMSE when ε is a Wiener error process. However the Gasser and Müller's estimator doesn't require the knowledge of the autocovariance function whereas the projection estimator does.

5 Simulation study

In this section, we investigate the performance of the proposed estimator (3) using finite values of experimental units m and sampling points n. The following growth curves are considered:

$$\text{(M1)} \quad g(x) = 10x^3 - 15x^4 + 6x^5 \quad \text{for } 0 < x < 1.$$

$$\text{(M2)} \quad g(x) = x + 0.5 \, e^{-80(x-0.5)^2} \quad \text{for } 0 < x < 1.$$

This growth curves were used by Hart and Wherly (1986) and Benhenni and Rachdi (2007), due to its similarity in shape to that of the logistic function, which is frequently found in growth curve analysis as noted by Hart and Wherly (1986). The sampling points are taken to be:

$$t_i = (i - 0.5)/n \quad \text{for } i = 1, \dots, n. \tag{18}$$
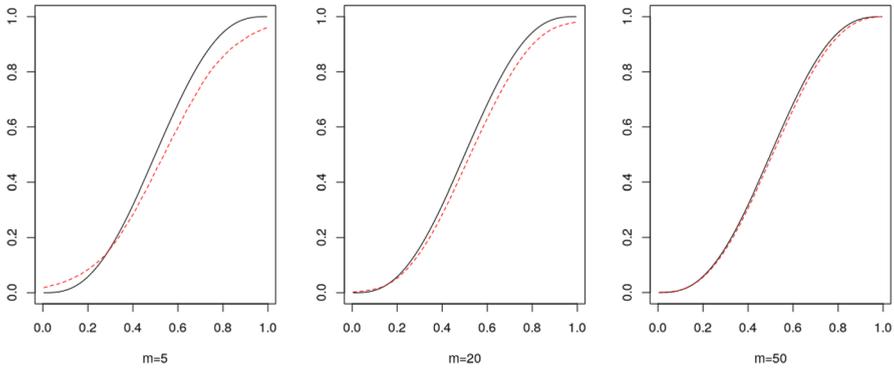
Fig. 1 The regression function of model (M1) is in solid line and the projection estimator is in dashed line
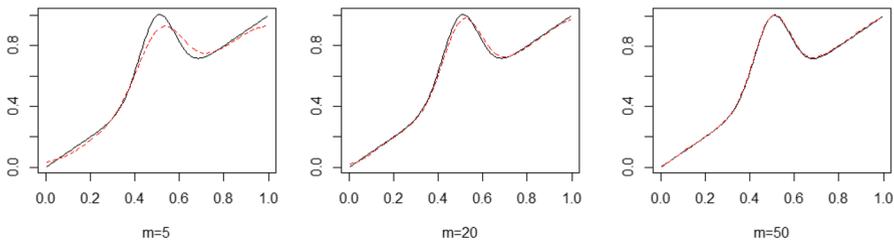
Fig. 2 The regression function of model (M2) is in solid line and the projection estimator is in dashed line

The error process ε is taken to be the Wiener error process with autocovariance function $R(s, t) = \sigma^2 \min(s, t)$. The Kernel used here is the quartic kernel given by $K(u) = \frac{15}{16}(1-u^2)^2 I_{[-1,1]}(u)$, and the bandwidth is the optimal one with respect to the exact IMSE, obtained using the conjugated gradient algorithm (CGA). We consider the mean of all estimators obtained from 100 simulations. We take $\sigma^2 = 0.5$, simulations for other values of σ^2 gave similar results. The results are given in Figs. 1 and 2 for Models (M1) and (M2), respectively, for a fixed number of observations $n = 100$ and three different values of experimental units $m = 5, 20, 50$.

We can see for Model (M1), from Fig. 1, that the projection estimator gets closer to the regression function when m gets bigger, which proves its good performance and consistency when m increases. These results are confirmed for the growth curve Model (M2) given in Fig. 2.

In this simulation study, we consider the comparison of the proposed estimator (3) to the Gasser and Müller (17) (referred by GM estimator) with respect to the exact IMSE in a finite sample set. For this, we consider the cubic growth curve of model (M1). We consider also the uniform design given by (18) and the quartic kernel $K(u) = \frac{15}{16}(1-u^2)^2 I_{[-1,1]}(u)$. For the error process, we shall consider both the Wiener of autocovariance function $R(s, t) = \min(s, t)$, and the Ornstein–Uhlenbeck process with autocovariance $R(s, t) = e^{-|s-t|}$.

The weight w, chosen here, is the uniform density on $[0, 1]$, i.e., $w \equiv 1$ on $[0, 1]$, we consider the optimal bandwidth with respect to the exact IMSE of the two estimators,

Table 1 The integrated squared bias, integrated variance, IMSE and the optimal bandwidth for $n = 10$ and different values of m under the Wiener error process, for the GM and the projection estimators

$n = 10$	m	Ibias^2	Ivar	IMSE	h_{opt}
GM	10	1.508×10^{-3}	4.507×10^{-2}	4.658×10^{-2}	0.335
Pro		1.304×10^{-3}	4.399×10^{-2}	4.530×10^{-2}	0.321
GM	50	2.662×10^{-4}	9.494×10^{-3}	9.760×10^{-3}	0.198
Pro		1.981×10^{-4}	9.228×10^{-3}	9.426×10^{-3}	0.187
GM	100	1.505×10^{-4}	4.826×10^{-3}	4.977×10^{-3}	0.154
Pro		0.897×10^{-4}	4.689×10^{-3}	4.778×10^{-3}	0.142

Table 2 The integrated squared bias, integrated variance, IMSE and the optimal bandwidth for $n = 10$ and different values of m under the Ornstein–Uhlenbeck error process, for the GM and the projection estimators

$n = 10$	m	Ibias^2	Ivar	IMSE	h_{opt}
GM	10	2.596×10^{-3}	8.821×10^{-2}	9.080×10^{-2}	0.387
Pro		2.494×10^{-3}	8.703×10^{-2}	8.952×10^{-2}	0.386
GM	50	4.481×10^{-4}	1.848×10^{-2}	1.893×10^{-2}	0.236
Pro		4.097×10^{-4}	1.822×10^{-2}	1.863×10^{-2}	0.237
GM	100	2.299×10^{-4}	9.390×10^{-3}	9.620×10^{-3}	0.186
Pro		1.885×10^{-4}	9.265×10^{-3}	9.453×10^{-3}	0.187

Table 3 The integrated squared bias, integrated variance, IMSE and the optimal bandwidth for $m = 20$ and different values of n under the Wiener error process, for the GM and the projection estimators

$m = 20$	n	Ibias^2	Ivar	IMSE	h_{opt}
GM	10	3.293×10^{-4}	1.180×10^{-2}	1.213×10^{-2}	0.213
Pro		2.571×10^{-4}	1.147×10^{-2}	1.173×10^{-2}	0.203
GM	50	2.579×10^{-4}	1.136×10^{-2}	1.162×10^{-2}	0.230
Pro		2.532×10^{-4}	1.137×10^{-2}	1.162×10^{-2}	0.228
GM	100	2.573×10^{-4}	1.135×10^{-2}	1.161×10^{-2}	0.235
Pro		2.549×10^{-4}	1.136×10^{-2}	1.161×10^{-2}	0.229

i.e., $\inf_{0<h<1} \text{IMSE}(h)$. The bandwidth h is chosen through the algorithm CGA. The results are given in Tables 1 and 2 for $n = 10$ and for different values of m, and in Tables 3 and 4 for a fixed $m = 20$ and for different values of n. These tables present the integrated bias squared denoted by Ibias^2, integrated variance denoted by Ivar and the IMSE together with the optimal bandwidth associated with each estimator.

First, we can see from Tables 1, 2, 3 and 4 that the optimal bandwidth decreases when m increases, as shown in Corollary 1. In addition, the optimal bandwidth of the projection estimator is slightly smaller than that of the GM estimator.

Table 4 The integrated squared bias, integrated variance, IMSE and the optimal bandwidth for $m = 20$ and different values of n under the Ornstein–Uhlenbeck error process, for the GM and the projection estimators

$m = 20$	n	Ibias^2	Ivar	IMSE	h_{opt}
GM	10	1.199×10^{-3}	4.507×10^{-2}	4.627×10^{-2}	0.315
Pro		1.145×10^{-3}	4.445×10^{-2}	4.559×10^{-2}	0.315
GM	50	1.092×10^{-3}	4.431×10^{-2}	4.540×10^{-2}	0.326
Pro		1.091×10^{-3}	4.428×10^{-2}	4.537×10^{-2}	0.326
GM	100	1.090×10^{-3}	4.428×10^{-2}	4.537×10^{-3}	0.326
Pro		1.089×10^{-3}	4.428×10^{-2}	4.537×10^{-3}	0.326

It is also seen that both the Ivar and the Ibias^2 of the two estimators decrease when m increases. In addition, the projection estimator has a smaller Ibias^2 and Ivar than that of the GM estimator, which leads to a smaller IMSE.

Another way to look at these results is as follows: for a fixed number of experimental units $m = 10$ and when the error process is a Wiener process (similar results for the Ornstein–Uhlenbeck error process), the projection estimator would only need $n = 10$ observations on each experimental unit to obtain the performance IMSE $= 4.53 \times 10^{-2}$ (see Table 1), whereas the GM estimator would need to have $n = 18$ observations to obtain the same performance, and thus requires 80% more samples in order to achieve the same performance.

The results of this simulation study show that, even for small number of observations, the projection estimator outperforms the GM estimator with respect to IMSE.

It should be noted here that, in order to solve the problem at the edges $[0, h] \cap [1 - h, 1]$, it was necessary to adjust the kernel as suggested by Hart and Wherly (1986).

Robustness of the projection estimator

Since the projection estimator depends on the autocovariance function of the errors, which is not always known in practical cases, we perform here a simulation study to test its performance with an estimated autocovariance matrix. Suppose that a reasonable parametric model for the error autocovariance function is known, consider for instance the Ornstein–Uhlenbeck error process with an unknown parameter λ.

To estimate the autocovariance parameter, we use the following criterion which gives consistent estimator of the autocovariance parameter, as done for instance in Ferreira et al. (1997):

$$Q_{n,m}(\lambda) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{i=1}^{n} \left(\hat{R}(t_i, t_j) - R(t_i, t_j) \right)^2,$$

where the empirical correlation estimation is given by

$$\hat{R}(t_i, t_j) = \frac{1}{m-1} \sum_{k=1}^{m} \left(Y_k(t_i) - \overline{Y}(t_i) \right) \left(Y_k(t_j) - \overline{Y}(t_j) \right). \tag{19}$$

We generate 100 matrices $(Y_j(t_i))_{\substack{1 \le j \le n \\ 1 \le i \le n}}$ of observations of the function g defined by (M1), the Ornstein–Uhlenbeck error process with $\lambda = 1$ and using the uniform design. For every matrix, we estimate the parameter λ using the Generalized Simulated Annealing (GSA) algorithm to minimize $Q_{n,m}$, the estimated parameter noted by $\hat{\lambda}$ is then the median of 100 estimated values. For more details on the use of the software R algorithm function, see Xiang et al. (2013). This algorithm is essentially known for its ability to handle very complex nonlinear objective functions with a very large number of optima.

To compare the projection estimator with the Gasser and Müller's estimator, we used the estimated mean average squared error (MASE) given, for the projection estimator, by:

$$\text{MASE}^{\text{Pro}}(h) = \mathbb{E}(\text{RSS}(h)) - \frac{1}{nm} \sum_{i=1}^{n} R(t_i, t_i) + \frac{2}{nm} \text{tr}(K_h R_{|T_n}), \qquad (20)$$

where

$$\text{RSS}(h) = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{g}_n^{\text{pro}}(t_i) - \overline{Y}(t_i) \right)^2,$$

and

$$K_h = \left(m_h(t_i, t_l) \right)_{1 \le i,l \le n} \quad \text{with} \quad m_h(\cdot, x)_{|T_n} = f'_{x,h|T_n} R_{|T_n}^{-1}.$$

For the Gasser and Müller's estimator, we replace the matrix K_h by:

$$K_h = \left(\frac{1}{h} \int_{m_{i-1,}}^{m_i} K\left(\frac{x - t_l}{h}\right) \mathrm{d}t \right)_{1 \le i,l \le n} \quad \text{with} \quad m_i = \frac{t_i + t_{i+1}}{2}.$$

The estimated MASE is given by

$$\hat{M}(h) = \text{RSS}(h) - \frac{1}{nm} \sum_{i=1}^{n} \hat{R}(t_i, t_i) + \frac{2}{nm} tr(K_h \hat{R}_{|T_n}). \qquad (21)$$

The results are presented in Table 5, which shows the estimated parameter λ, the optimal bandwidth h^* minimizing $\hat{M}(h)$ using the CGA, $\hat{M}(h^*)$ for the estimator of Gasser and Müller, the projection estimator with the true value $\lambda = 1$ and the projection estimator with the estimated values $\hat{\lambda}$.

Table 5 shows that the projection estimator is robust with respect to the misspecification of the autocovariance function, i.e., the MASE of the projection estimator with an estimated autocovariance function is smaller than the Gasser and Müller's one, for a large number of measurement units m, where the estimated values of λ are close to the real value $\lambda = 1$.

Table 5 The estimated parameter λ, the optimal bandwidth h^* and $M(h^*)$ for $n = 10$ and different values of m under the Ornstein–Uhlenbeck error process

$n = 10$	m	$\hat{\lambda}$	h^*	$M(h^*)$
GM	10	1.5	0.563	6.214×10^{-2}
Pro(λ)			0.562	6.084×10^{-2}
Pro ($\hat{\lambda}$)			0.575	6.215×10^{-2}
GM	50	1.08	0.331	1.555×10^{-2}
Pro(λ)			0.333	1.525×10^{-2}
Pro ($\hat{\lambda}$)			0.337	1.545×10^{-2}
GM	100	*0.94	0.266	8.234×10^{-3}
Pro(λ)			0.266	8.076×10^{-3}
Pro ($\hat{\lambda}$)			0.265	8.154×10^{-3}

Remark 9 In our simulations, the used bandwidth is the optimal one selected to minimize the exact IMSE, which is not known in practice. As an alternative, one can use for instance the data driven selection method, such as the cross-validation (leave one observation out) which turned out to be inefficient in the presence of correlations of the errors, see for instance Altman (1990) and Chiu (1989) and Hart (1991, 1994). In the presence of correlations, we use the adaptative criterion based on Rice (1984) (see also Hart and Wherly 1986, which consists of minimizing $\hat{M}(h)$ given by Equation (21) above. If R is unknown, one can estimate it using the function $\hat{R}$ given by Equation (19) above, where the function $f_{x,h}(t_i)$ given in (2) can be estimated by:

$$\hat{f}_{x,h}(t_i) = \sum_{j=1}^{n} \hat{R}(t_i, t_j) \varphi_{x,h}(t_j)(t_{j+1} - t_j)/2.$$

References

Altman, N. S. (1990). Kernel smoothing of data with correlated errors. *American Statistical Association*, *85*, 749–759.

Azzalini, A. (1984). Estimation and hypothesis testing for collections of autoregressive time series. *Biometrika*, *71*, 85–90.

Belouni, M., Benhenni, K. (2015). Optimal and robust designs for estimating the concentration curve and the AUC. *Scandinavian Journal of Statistics*: *Theory and Application*, *42*, 453–470.

Benedetti, J. (1977). On the nonparametric estimation of the regression function. *Journal of the Royal Statistical Society*, *39*, 248–253.

Benelmadani, D. (2019a). *Contribution à la Régression non Paramétrique avec un Processus Erreur d'Autocovariance Générale et Application en Pharmacocinétique*, Ph.D. 2019, Grenoble Alpes University, France.

Benelmadani, D., Benhenni, K., Louhichi, S. (2019b). Trapezoidal rule and sampling designs for the nonparametric estimation of the regression function in models with correlated errors. arXiv:1806.04896.

Benhenni, K., Cambanis, S. (1992). Sampling designs for estimating integrals of stochastic processes. *The Annals of Statistics*, *20*, 161–194.

Benhenni, K., Rachdi, M. (2007). Nonparametric estimation of average growth curve with general nonstationary error process. *Communications in Statistics*: *Theory and Methods*, *36*, 1137–1186.

Blanke, D., Bosq, D. (2008). Regression estimation and prediction in continuous time. *Journal of the Japan Statistical Society (Nihon Tôkei Gakkai Kaihô)*, *38*, 15–26.

Cheng, K. F., Lin, P. E. (1981). Nonparametric estimation of the regression function. *Z. Wahrscheinlichkeit-stheorie verw. Gebiete*, *57*, 223–233.

Chiu, S. T. (1989). Bandwidth selection for kernel estimation with correlated noise. *Statistics and Probability Letters*, *8*, 347–354.

Dette, H., Pepelyshev, A., Zhigljavsky, A. (2016). Best linear unbiased estimators in continuous time regression models. arXiv:1611.09804.

Didi, S., Louani, D. (2013). Asymptotic results for the regression function estimate on continuous time stationary and ergodic data. *Journal of Statistics and Risk Modelling*, *31*(2), 129–150.

Ferreira, E., Núñez-Antón, V., Rodríguez-Póo, J. (1997). Kernel regression estimates of growth curves using nonstationary correlated errors. *Statistics and Probability Letters*, *34*, 413–423.

Gasser, T., Müller, H. G. (1979). Kernel estimation of regression functions. *Lecture Notes in Mathematics*, *757*, 23–68.

Hart, J. D. (1991). Kernel regression estimation with time series errors. *Royal Statistical Society B*, *53*, 173–187.

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

Hart, J. D., Wherly, T. E. (1986). Kernel regression estimation using repeated measurements data. *Journal of the American Statistical Association*, *81*, 1080–1088.

Müller, G. H. (1984). Optimal designs for nonparametric kernel regression. *Statistics and Probability Letters*, *2*, 285–290.

Parzen, E. (1959). *Statistical inference on time series by hilbert space methods*, p. 23. Technical Report, Department of Statistics, Stanford University, Stanford, CA.

Priestly, M. B., Chao, M. T. (1972). Nonparametric function fitting. *Journal of Royal Statistical Society*, *34*, 384–392.

Ramsay, J. O., Silverman, B. W. (2005). *Functional data analysis*, Springer series in statistics, New York: Springer.

Rice, J. (1984). Bandwidth choice for nonparametric regression. *The Annals of Statistics*, *12*, 1215–1230.

Sacks, J., Ylvisaker, D. (1966). Designs for regression problems with correlated errors. *The Annals of Mathematical Statistics*, *37*, 66–89.

Sacks, J., Ylvisaker, D. (1968). Designs for regression problems with correlated errors: Many parameters. *The Annals of Mathematical Statistics*, *39*, 69–79.

Sacks, J., Ylvisaker, D. (1970). Designs for regression problems with correlated errors III. *The Annals of Mathematical Statistics*, *41*, 2057–2074.

Schwartz, L. (1964). Sous Espace Hilbertiens d'Espace Vectoriels Topologiques et Noyaux Associés (Noyaux Reproduisant). *Journal d'Analyse Mathématique*, *13*, 115–256.

Su, Y., Cambanis, S. (1993). Sampling designs for estimation of a random process. *Stochastic Processes and their Applications*, *46*, 47–89.

Xiang, Y., Gubian, S., Suomela, B., Hoeng, J. (2013). Generalized simulated annealing for efficient global optimization: The GenSA package for R. *The R Journal*, *5*, 13–28.