

# OPTIMAL DESIGNS FOR SERIES ESTIMATION IN NONPARAMETRIC REGRESSION WITH CORRELATED DATA

Kirsten Schorning, Maria Konstantinou and Holger Dette

*Technische Universität Dortmund, Tickmill Europe Ltd  
and Ruhr-Universität Bochum*

*Abstract:* We investigate the problem of designing experiments for series estimators in nonparametric regression models with correlated observations. We use projection-based estimators to derive an explicit solution of the best linear oracle estimator in the continuous-time model for all Markovian-type error processes. These solutions are then used to construct estimators, which can be calculated from the available data, along with their corresponding optimal design points. Our results are illustrated by means of a simulation study, which demonstrates that the proposed series estimator outperforms commonly used techniques based on the optimal linear unbiased estimators. Moreover, we show that the performance of the proposed estimators can be further improved by choosing the design points appropriately.

*Key words and phrases:* Integrated mean squared error, nonparametric regression, Optimal design, optimal estimator.

## 1. Introduction

Nonparametric regression is a common tool in statistical inference, with numerous applications (see the monographs of Efromovich (1999) and Tsybakov (2009), among many others). The basic model is formulated as

$$Y_i = f(X_i) + \varepsilon_i, \quad i = 1, \dots, n, \quad (1.1)$$

where one usually distinguishes between random and fixed predictors  $X_i$ . In the latter case, a natural question is how to choose  $X_1, \dots, X_n$  to obtain the most precise estimates of the regression function  $f$ . Several authors have worked on this problem, including Müller (1984) and Zhao and Yao (2012), who derived optimal designs with respect to different criteria for kernel estimates, and Dette and Wiens (2008), who considered the design problem for series estimation in terms

---

Corresponding author: Holger Dette, Fakultät für Mathematik, Ruhr-Universität Bochum, 44801 Bochum, Germany. E-mail: [holger.dette@rub.de](mailto:holger.dette@rub.de).

of spherical harmonics. We also refer to the work of Efromovich (2008), who proposed a sequential allocation scheme in a nonparametric model of the form given in (1.1), with random predictors and heteroscedastic errors. A common feature of the literature in this field is that all authors investigate the design problem in a model (1.1) with independent errors. However, in many situations, this assumption is not satisfied, particularly when the explanatory variable represents time.

The reason for this gap in the literature is that the design problem for models with correlated errors (even parametric models) is substantially harder compared to the uncorrelated case. In the latter case, a well developed and powerful methodology for the construction of optimal designs has been established (see, e.g., the monograph of Pukelsheim (2006)). In contrast, optimal designs for models with correlated observations are only available in rare circumstances involving parametric models (see, e.g., Pázman and Müller (2001), Näther and Simák (2003), Müller and Stehlík (2004), Harman and Stulažter (2010), Amo-Salas, López-Fidalgo and López-Rios (2012), Stehlik et al. (2015), and Rodríguez-Díaz (2017), among others). Some general results on optimal designs for linear models with correlated observations can be found in the seminal work of Sacks and Ylvisaker (1966, 1968). More recently, Dette, Pepelyshev and Zhigljavsky (2013, 2016); Dette, Konstantinou and Zhigljavsky (2017) provided a general approach for the problem of designing experiments in linear models with correlated observations by considering the problem of optimal (unbiased linear) estimation and optimal design simultaneously. Usually, authors use asymptotic arguments to embed the discrete (nonconvex) optimization problem in a continuous (or approximate) one. However, unlike the uncorrelated case, in the context of correlated observations, this approach does not simplify the problem substantially. Furthermore, owing to the lack of convexity, the resulting approximate optimal design problems for regression models with correlated observations are still extremely difficult to solve.

In this study we consider optimal design theory for series estimation in the nonparametric regression model (1.1) with correlated data. The basic notation and general design problem are introduced in Section 2. In order to address the particular difficulties in design problems for series estimation from correlated data, in Section 3, we consider a continuous-time version of the discrete model. We first determine optimal oracle estimators for the coefficients in a Fourier expansion of the regression function  $f$ . These are shrinkage estimators, and are not unbiased.

Section 4 is devoted to the implementation of the results from Section 3 for

the construction of an efficient estimator with a corresponding optimal design. In particular, we determine an optimal approximation of the Fourier coefficients in the continuous model (which requires the full trajectory of the process) using an estimator that can be calculated from the available data  $\{Y_{t_1}, \dots, Y_{t_n}\}$ . Then, we determine the design points  $t_1, \dots, t_n$ , such that the approximation has a minimal mean squared error with respect to the solution in the continuous-time model. The resulting two-stage estimator shrinks the best linear unbiased estimator when the design points are chosen in an optimal way. The superior performance of our approach is demonstrated in Section 5 by means of a small simulation study. The online Supplementary Material contains all technical details and further numerical results.

## 2. Optimal Designs for Series Estimation

Throughout this paper, we consider the following nonparametric regression model with a fixed design:

$$Y_{t_i} = f(t_i) + \varepsilon_{t_i}, \quad i = 1, \dots, n, \quad (2.1)$$

where  $f : [0, 1] \rightarrow \mathbb{R}$  is the regression function,  $0 \leq t_1 < t_2 < \dots < t_n \leq 1$  are  $n$  distinct time points in the interval  $[0, 1]$ ,  $\mathbb{E}[\varepsilon(t_j)] = 0$ , and  $K(t_i, t_j) = \mathbb{E}[\varepsilon_{t_i} \varepsilon_{t_j}]$  denotes the covariance between the observations at points  $t_i$  and  $t_j$  ( $i, j = 1, \dots, n$ ). Let

$$L^2([0, 1]) = \left\{ g : [0, 1] \rightarrow \mathbb{R} : \int_0^1 g^2(t) dt < \infty \right\}$$

denote the space of square-integrable (real-valued) functions with inner product  $\langle g_1, g_2 \rangle = \int_0^1 g_1(t) g_2(t) dt$  and norm  $\|g\|_2 = (\int_0^1 g^2(t) dt)^{1/2}$ . Let  $\{\varphi_j(\cdot) : j \in \mathbb{N}\}$  be an orthonormal basis. Then any function  $f \in L^2([0, 1])$  admits a series expansion of the form

$$f(t) = \sum_{j \in \mathbb{N}} \theta_j \varphi_j(t) \quad (2.2)$$

in  $L^2([0, 1])$  with Fourier coefficients

$$\theta_j = \langle f, \varphi_j \rangle = \int_0^1 f(t) \varphi_j(t) dt, \quad j \in \mathbb{N}. \quad (2.3)$$

Moreover, the coefficients are squared summable; that is,  $\sum_{j \in \mathbb{N}} \theta_j^2 < \infty$ . In order to estimate the unknown function  $f$ , we now follow the idea of projection estimators (see Tsybakov (2009), p.47) and estimate the truncated series

$f^{(J)}(t) = \sum_{j=1}^J \theta_j \varphi_j(t)$  by

$$\hat{f}^{(J)}(t) = \sum_{j=1}^J \hat{\theta}_j \varphi_j(t), \quad (2.4)$$

where  $\hat{\theta}_j$  is an appropriate estimator for the Fourier coefficient  $\theta_j$  ( $j = 1, \dots, J$ ). For example, if  $\max_{i=2}^n (t_i - t_{i-1}) \rightarrow 0$  as  $n \rightarrow \infty$ , an asymptotically unbiased estimator of  $\theta_j$  is given by

$$\sum_{i=2}^n (t_i - t_{i-1}) \varphi_j(t_{i-1}) Y_{t_{i-1}}. \quad (2.5)$$

More general estimators are specified later on. At this point, it is only important to note that the performance of any reasonable estimator will depend on the design points  $t_1, \dots, t_n$ . We are interested in choosing these design points such that the mean integrated squared error,

$$\mathbb{E} \left[ \int_0^1 (\hat{f}^{(J)}(t) - f(t))^2 dt \right] = \sum_{j=1}^J \mathbb{E} [(\hat{\theta}_j - \theta_j)^2] + \sum_{j=J+1}^{\infty} \theta_j^2,$$

is minimal. We also note that any solution of this discrete optimization problem depends on the unknown regression function  $f$ , truncation point  $J$  used in (2.4), and covariance kernel  $K$ , which is assumed to be known throughout this paper. On the other hand, the term  $\sum_{j=J+1}^{\infty} \theta_j^2$  does not depend on the design points. Therefore, it can be determined by minimizing  $\sum_{j=1}^J \mathbb{E} [(\hat{\theta}_j - \theta_j)^2]$  with respect to the choice of  $t_1, \dots, t_n$ . For example, if  $\hat{\theta}_j = \sum_{i=1}^{\ell_j} \alpha_{ji} Y_{t_i}$  is a linear estimator of  $\theta_j$  ( $j = 1, \dots, J$ ), we have that

$$\sum_{j=1}^J \mathbb{E} [(\hat{\theta}_j - \theta_j)^2] = \sum_{j=1}^J \left( \sum_{i=1}^{\ell_j} \alpha_{ji} f(t_i) - \theta_j \right)^2 + \sum_{j=1}^J \sum_{i_1, i_2=1}^{\ell_j} \alpha_{ji_1} \alpha_{ji_2} K(t_{i_1}, t_{i_2}),$$

which has to be minimized with respect to the choice of the time points  $t_1, \dots, t_n$ .

### 3. Optimal Estimation in the Continuous-Time Model

The discrete optimization problem stated at the end of the previous section is extremely difficult to solve. In this section, in order to derive efficient designs, we investigate a simpler problem. Specifically, we consider the continuous-time nonparametric regression model of the form

$$Y_t = f(t) + \varepsilon_t, \quad t \in [0, 1], \quad (3.1)$$

where  $f$  is an unknown square-integrable function, and the error process  $\varepsilon = \{\varepsilon_t : t \in [0, 1]\}$  is a centered Gaussian process with covariance kernel  $K(s, t) = \mathbb{E}[\varepsilon_s \varepsilon_t]$ . Because we assume that the full trajectory of the process is available, there is in fact no optimal design problem, but only the issue of an optimal estimation of the regression function  $f$ . The optimal design question will appear later, when we return to the discrete model (2.1). The main result of this section provides an oracle solution of the optimal estimation problem. In particular, the optimal estimator depends on the unknown function  $f$  in model (3.1) and, therefore, is not implementable (even if the full trajectory of the process  $\{Y_t : t \in [0, 1]\}$  is available). However, our solution serves as a benchmark and suggests how good estimators and corresponding optimal designs can be constructed. This is formulated precisely in Section 4.

Model (3.1) is often written in terms of a stochastic differential equation (provided that the regression function  $f$  is differentiable, with derivative  $\dot{f}$ ); that is,

$$dY_t = \dot{f}(t)dt + d\varepsilon_t, \quad t \in [0, 1]. \quad (3.2)$$

If  $\varepsilon = \{\varepsilon_t : t \in [0, 1]\}$  is a Brownian motion, model (3.2) is called a Gaussian white noise model, and has found garnered attention in the statistical literature (see, e.g., Ibragimov and Hasminskiĭ (1981) or Tsybakov (2009), among many others). In particular, the model is asymptotically equivalent to the nonparametric regression model  $Z_i = \dot{f}(i/n) + \eta_i$  ( $i = 1, \dots, n$ ), where  $\eta_1, \dots, \eta_n$  are independent standard normally distributed random variables (see Brown and Low (1996)). Note that the focus in the aforementioned publications is on the optimal estimation of the function  $\dot{f}$ , whereas in this section, we are interested in estimating the function  $f$  in model (3.1). Nevertheless, under additional assumptions, we can investigate the properties of the derivative of the oracle estimator developed in what follows. A brief discussion of these relations is given in Example 1.

Another important difference between model (3.1) and the Gaussian white noise model is that we consider a general error process  $\{\varepsilon_t : t \in [0, 1]\}$ . In particular, we concentrate on Markovian Gaussian error processes with a covariance kernel of the form

$$\mathbb{E}[\varepsilon_s \varepsilon_t] = K(s, t) = u(s)v(t) \quad \text{for } s \leq t, \quad (3.3)$$

where  $u(\cdot)$  and  $v(\cdot)$  are some (known) functions defined on the interval  $[0, 1]$ , such that  $v(t) \neq 0$ , for  $t \in [0, 1]$ . Kernels of this form generalize the Brownian motion, which is obtained for  $u(t) = t$  and  $v(t) = 1$ , and are called triangular kernels in the literature. Property (3.3) essentially characterizes a Gaussian process as

Markovian (see Doob (1949) for more details). We assume that the process  $\{\varepsilon_t : t \in [0, 1]\}$  is nondegenerate on the open interval  $(0, 1)$ , which implies that the function

$$q(t) = \frac{u(t)}{v(t)}, \quad (3.4)$$

is positive on the interval  $(0, 1)$ , and strictly increasing and continuous on  $[0, 1]$ .

With regard to estimating the unknown function  $f$ , we propose estimating the coefficients  $\theta_j$  in the projection estimator (2.4) using statistics of the form (see Grenander (1950))

$$\hat{\theta}_j = \int_0^1 Y_t \xi_j(dt), \quad j \in \mathbb{N}, \quad (3.5)$$

where  $\xi_j$  is a signed measure on the interval  $[0, 1]$ , such that

$$\begin{aligned} \sum_{j=1}^{\infty} \{(\mathbb{E}[\hat{\theta}_j])^2 + \text{Var}(\hat{\theta}_j)\} &= \sum_{j=1}^{\infty} \left( \int_0^1 f(t) d\xi_j(t) \right)^2 \\ &+ \sum_{j=1}^{\infty} \int_0^1 \int_0^1 K(s, t) d\xi_j(s) d\xi_j(t) < \infty. \end{aligned} \quad (3.6)$$

Obviously, this condition implies that for the sequence of estimators  $(\hat{\theta}_j)_{j \in \mathbb{N}}$ ,  $\sum_{j=1}^{\infty} \mathbb{E}[\hat{\theta}_j^2] < \infty$ ; thus, and thus we can define the random variable

$$\hat{f}(t) = \sum_{j=1}^{\infty} \hat{\theta}_j \varphi_j(t). \quad (3.7)$$

In particular, if  $\hat{f}^{(J)}(t) = \sum_{j=1}^J \hat{\theta}_j \varphi_j(t)$  is the truncated series from (3.7), we have that

$$\begin{aligned} \lim_{J \rightarrow \infty} \mathbb{E} \left[ \int_0^1 (\hat{f}^{(J)}(t) - f(t))^2 dt \right] &= \lim_{J \rightarrow \infty} \sum_{j=1}^J \mathbb{E}[(\hat{\theta}_j - \theta_j)^2] \\ &= \sum_{j=1}^{\infty} \mathbb{E}[(\hat{\theta}_j - \theta_j)^2] < \infty, \end{aligned}$$

and the mean integrated squared error of the estimator  $\hat{f}$  in (3.7) is given by

$$\text{MISE}(\hat{f}) := \mathbb{E} \left[ \int_0^1 (\hat{f}(t) - f(t))^2 dt \right] = \sum_{j=1}^{\infty} \mathbb{E}[(\hat{\theta}_j - \theta_j)^2]. \quad (3.8)$$

We conclude that the optimal linear oracle estimator  $\hat{f}$  of the function  $f$

minimizing (3.8) can be determined by minimizing the individual mean squared errors  $\mathbb{E}[(\hat{\theta}_j - \theta_j)^2]$  separately. From the definition of linear estimators in (3.5), this problem corresponds to one of determining a signed measure  $\xi_j^*$  on the interval  $[0, 1]$ , which minimizes the functional

$$\begin{aligned} \Psi_j(\xi_j) &:= \mathbb{E} \left[ \left( \int_0^1 Y_t \xi_j(dt) - \theta_j \right)^2 \right] \\ &= \int_0^1 \int_0^1 [f(s)f(t) + K(s, t)] \xi_j(ds) \xi_j(dt) - 2\theta_j \int_0^1 f(s) \xi_j(ds) + \theta_j^2 \\ &= \int_0^1 \int_0^1 K(s, t) \xi_j(ds) \xi_j(dt) + \left( \int_0^1 f(s) \xi_j(ds) - \theta_j \right)^2. \end{aligned} \quad (3.9)$$

**Remark 1.**

- (1) Note that, in contrast to most of the literature, we do not assume that  $\hat{\theta}_j$  is an unbiased estimator of the Fourier coefficient  $\theta_j$  ( $j \in \mathbb{N}$ ). A prominent unbiased estimator for  $\theta_j$  is given by

$$\tilde{\theta}_j = \int_0^1 Y_t \varphi_j(t) dt \quad (j \in \mathbb{N}), \quad (3.10)$$

and for general unbiased estimates of the form (3.5), condition (3.6) reduces to

$$\sum_{j=1}^{\infty} \int_0^1 \int_0^1 K(s, t) d\xi_j(s) d\xi_j(t) < \infty. \quad (3.11)$$

Moreover, if the kernel  $K$  is continuous on  $[0, 1] \times [0, 1]$ , and if  $\varphi_1, \varphi_2, \dots$  are the eigenfunctions of the integral operator associated with the covariance kernel  $K$ , with corresponding eigenvalues  $\lambda_1, \lambda_2, \dots$ , then condition (3.6) further reduces to

$$\sum_{j=1}^{\infty} \int_0^1 \int_0^1 K(s, t) \varphi_j(s) \varphi_j(t) ds dt = \sum_{j=1}^{\infty} \lambda_j < \infty.$$

- (2) Under the additional assumption that the estimator (3.5) is unbiased for  $\theta_j$ , the second term in (3.9) vanishes. Thus, the resulting optimization problem corresponds to one of finding the best linear estimator in the location scale model  $Y_t = \theta_j + \varepsilon_t$ , first studied by Grenander (1950). This author showed that under the additional constraint  $\int_0^1 d\xi_j(dt) = 1$ , the optimal solution  $\xi_j^*$  minimizing  $\int_0^1 \int_0^1 K(s, t) \xi_j(ds) \xi_j(dt)$  can be characterized by the property that the function  $t \rightarrow \int_0^1 K(s, t) \xi_j^*(ds)$  is constant on the interval  $[0, 1]$ .

The following theorem provides a complete solution of the optimization problem (3.9). The proof can be found in the Supplementary Material. For a precise statement of the result, we denote by  $\delta_x$  the Dirac measure at point  $x$ , and distinguish the following cases for the triangular kernel (3.3).

- (A)  $u(0) \neq 0$ .
- (B)  $u(0) = 0, f(0) = 0$ .
- (C)  $u(0) = 0, f(0) \neq 0$ .

**Theorem 1.** *Consider the functional  $\Psi_j$  in (3.9) with a twice differentiable regression function  $f$  and a triangular covariance kernel of the form (3.3), where the functions  $u$  and  $v$  are also twice differentiable. For any  $j \in \mathbb{N}$ , the signed measure  $\xi_j^*(dt)$  minimizing the functional  $\Psi_j$  in the class of all signed measures on the interval  $[0, 1]$  is given by*

$$\xi_j^*(dt) = \frac{\theta_j}{1+c} (P_0\delta_0(dt) + P_1\delta_1(dt) + p(t)dt), \tag{3.12}$$

where  $\theta_j$  is the  $j$ th Fourier coefficient in the Fourier expansion (2.2). The values for  $c, P_0, P_1$ , and the function  $p(\cdot)$  do not depend on the index  $j$ , and take different values corresponding to the properties of the functions  $u(\cdot)$  and  $f(\cdot)$ . In particular, we have the following cases:

- (A) *If  $u(0) \neq 0$ , the quantities  $c, P_0, P_1$ , and  $p$  are given by*

$$c = \int_0^1 \left\{ \frac{d}{dt} \left[ \frac{f(t)}{v(t)} \right] \right\}^2 \left( \frac{d}{dt} q(t) \right)^{-1} dt + \frac{f^2(0)}{v^2(0)} (q(0))^{-1}, \tag{3.13}$$

$$P_0 = -\frac{1}{v(0)} \frac{d}{dt} \left[ \frac{f(t)}{u(t)} \right] \Big|_{t=0} \left( \frac{d}{dt} q(t) \Big|_{t=0} \right)^{-1} q(0), \tag{3.14}$$

$$P_1 = \frac{1}{u(1)} \frac{d}{dt} \left[ \frac{f(t)}{v(t)} \right] \Big|_{t=1} \left( \frac{d}{dt} q(t) \Big|_{t=1} \right)^{-1} q(1), \tag{3.15}$$

$$p(t) = -\frac{1}{v(t)} \frac{d}{dt} \left\{ \frac{d}{dt} \left[ \frac{f(t)}{v(t)} \right] \left( \frac{d}{dt} q(t) \right)^{-1} \right\}, \tag{3.16}$$

where the function  $q$  is defined in (3.4).

- (B) *If  $u(0) = 0$  and  $f(0) = 0$ , the quantity  $c$  is given by*

$$c = \int_0^1 \left\{ \frac{d}{dt} \left[ \frac{f(t)}{v(t)} \right] \right\}^2 \left( \frac{d}{dt} q(t) \right)^{-1} dt, \tag{3.17}$$

where  $P_0 = 0$  and  $P_1$  and  $p$  are given by (3.15) and (3.16), respectively.

(C) If  $u(0) = 0$  and  $f(0) \neq 0$ , the quantities  $c$ ,  $p(t)$ , and  $P_1$  are equal to zero, whereas  $P_0$  is given by  $P_0 = 1/f(0)$ .

**Corollary 1.** Consider the regression model (3.1) with a twice differentiable regression function  $f$  and a nondegenerate centered Gaussian error process  $\{\varepsilon_t : t \in [0, 1]\}$ , with a triangular covariance kernel of the form (3.3), where the functions  $u$  and  $v$  are twice differentiable. The best linear oracle estimator minimizing the mean integrated squared error in (3.8) in the class of all linear estimators of the form (3.7) satisfying (3.6) is defined by  $f^*(t) = \sum_{j=1}^{\infty} \hat{\theta}_j^* \varphi_j(t)$ , where the coefficients  $\hat{\theta}_j^*$  are given by

$$\hat{\theta}_j^* = \int_0^1 Y_t \xi_j^*(dt), \quad j \in \mathbb{N}, \quad (3.18)$$

and the signed measure  $\xi_j^*(dt)$  is defined in Theorem 1. Moreover, the corresponding mean integrated squared error is given by

$$\text{MISE}(\hat{f}^*) = \frac{1}{1+c} \sum_{j=1}^{\infty} \theta_j^2 = \frac{1}{1+c} \int_0^1 f^2(t) dt,$$

where  $c$  is defined in (3.13).

Note that Theorem 1 is a theoretical result as it requires knowledge of the unknown regression function  $f$ . Nevertheless, we use it extensively in the following section to construct good estimators and corresponding optimal designs for series estimation in model (2.1).

**Remark 2.**

- (1) In model (2.1), with covariance kernel (3.3) and  $u(0) = 0$ , the observation  $Y_0$  at  $t = 0$  does not contain any errors. Therefore, the value of  $f(0)$  is known, and we can check whether case (B) or (C) of Theorem 1 holds.
- (2) The estimator given in Corollary 1 depends on the orthonormal system of the series expansion via the parameter  $\theta_j$ .
- (3) Using integration by parts, the resulting estimator  $\hat{\theta}_j^*$  in Corollary 1 can be represented as a stochastic integral. For example, in case (A) (where  $u(0) \neq 0$ ), the estimator can be represented as

$$(A) \quad \hat{\theta}_j^* = \frac{\theta_j}{1+c} \left\{ \int_0^1 \frac{d}{dt} \left[ \frac{f(t)}{v(t)} \right] \left( \frac{d}{dt} q(t) \right)^{-1} d \left( \frac{Y_t}{v(t)} \right) + \frac{f(0)}{u(0)} \frac{Y_0}{v(0)} \right\}, \quad (3.19)$$

where the constant  $c$  is defined in (3.13). Similarly, in case (B) (where  $u(0) = 0$  and  $f(0) = 0$ ), the estimator can be represented by

$$(B) \quad \hat{\theta}_j^* = \frac{\theta_j}{1+c} \left\{ \int_0^1 \frac{d}{dt} \left[ \frac{f(t)}{v(t)} \right] \left( \frac{d}{dt} q(t) \right)^{-1} d \left( \frac{Y_t}{v(t)} \right) \right\}. \tag{3.20}$$

Finally, in case (C) (where  $u(0) = 0$  and  $f(0) \neq 0$ ), the estimator directly reduces to

$$(C) \quad \hat{\theta}_j^* = \theta_j. \tag{3.21}$$

In the latter case, the estimator in (3.21) is not random, but fixed to the true, but unknown parameter  $\theta_j$ .

**Example 1.** A very popular orthonormal basis of  $L^2([0, 1])$  is given by the trigonometric functions  $\varphi_1(t) = 1$ ,

$$\varphi_j(t) = \begin{cases} \sqrt{2} \cos(2\pi kt), & j = 2k \\ \sqrt{2} \sin(2\pi kt), & j = 2k + 1 \end{cases}, \quad k = 1, 2, 3, \dots \tag{3.22}$$

Under the assumptions of Theorem 1, we assume that  $f$  and its derivative  $\dot{f}$  can be represented as a trigonometric series; that is,

$$f(t) = \theta_1 + \sum_{k=1}^{\infty} \sqrt{2} \cos(2\pi kt) \theta_{2k} + \sum_{k=1}^{\infty} \sqrt{2} \sin(2\pi kt) \theta_{2k+1}, \tag{3.23}$$

$$\dot{f}(t) = \bar{\theta}_1 + \sum_{k=1}^{\infty} \sqrt{2} \cos(2\pi kt) \bar{\theta}_{2k} + \sum_{k=1}^{\infty} \sqrt{2} \sin(2\pi kt) \bar{\theta}_{2k+1}. \tag{3.24}$$

Note that (under suitable assumptions) the Fourier coefficients in (3.23) and (3.24) are related by the equations

$$\bar{\theta}_1 = 0, \quad \bar{\theta}_{2k} = (2\pi k) \theta_{2k+1}, \quad \bar{\theta}_{2k+1} = -(2\pi k) \theta_{2k}. \tag{3.25}$$

If the error process  $\{\varepsilon_t : t \in [0, 1]\}$  in model (2.1) is given by a Brownian motion, we have  $u(t) = t$ ,  $v(t) = 1$  in the definition of the triangular kernel (3.3), and thus  $q(t) = t$ . A straightforward application of Corollary 1 (case (B)) yields the following for the optimal oracle estimator of the function  $f$ :

$$f^*(t) = \hat{\theta}_1^* + \sum_{k=1}^{\infty} \sqrt{2} \cos(2\pi kt) \hat{\theta}_{2k}^* + \sum_{k=1}^{\infty} \sqrt{2} \sin(2\pi kt) \hat{\theta}_{2k+1}^*, \tag{3.26}$$

where the estimated Fourier coefficients are given by

$$\hat{\theta}_j^* = \frac{\theta_j}{1+c} \int_0^1 \dot{f}(t) dY_t, \quad j \in \mathbb{N}, \quad (3.27)$$

(note that  $f(0) = f(1) = 0$ ). We thus also obtain an estimator of the function  $\dot{f}$  in model (3.2) by taking the derivative of  $f^*$  given in (3.26); that is,

$$\dot{f}^*(t) = - \sum_{k=1}^{\infty} (2\pi k) \sqrt{2} \sin(2\pi kt) \hat{\theta}_{2k}^* + \sum_{k=1}^{\infty} (2\pi k) \sqrt{2} \cos(2\pi kt) \hat{\theta}_{2k+1}^*. \quad (3.28)$$

Using the relation (3.25), the estimator in (3.27) can be rewritten as

$$\hat{\theta}_j^* = \begin{cases} -\frac{\bar{\theta}_{2k+1}}{2\pi k} \frac{1}{1+c} \int_0^1 \dot{f}(t) dY_t, & j = 2k, \\ \frac{\bar{\theta}_{2k}}{2\pi k} \frac{1}{1+c} \int_0^1 \dot{f}(t) dY_t, & j = 2k+1, \end{cases}$$

and the mean integrated squared error of the estimator  $\dot{f}^*$  in (3.28) is given by

$$\begin{aligned} \mathbb{E} \left[ \int_0^1 (\dot{f}^*(t) - \dot{f}(t))^2 dt \right] &= \sum_{j=2}^{\infty} \frac{\bar{\theta}_j^2}{(1+c)^2} \mathbb{E} \left[ \left( 1+c - \int_0^1 \dot{f}(t) dY_t \right)^2 \right] \\ &= \sum_{j=2}^{\infty} \frac{\bar{\theta}_j^2}{1+c} = \frac{\sum_{j=2}^{\infty} \bar{\theta}_j^2}{1 + \sum_{j=2}^{\infty} \bar{\theta}_j^2}, \end{aligned} \quad (3.29)$$

where we have used the representation  $c = \int_0^1 (\dot{f}(t))^2 dt = \sum_{j=1}^{\infty} \bar{\theta}_j^2 = \sum_{j=2}^{\infty} \bar{\theta}_j^2$  in the last equality. It might be of interest to compare this estimator with the linear oracle estimator

$$\dot{\check{f}}(t) = \sum_{j \in \mathbb{N}} \tilde{\theta}_j \varphi_j(t), \quad (3.30)$$

proposed in Tsybakov (2009, p.67), where  $\tilde{\theta}_j = \bar{\theta}_j^2 / (1 + \bar{\theta}_j^2) \int_0^1 \varphi_j(t) dY_t$  is used as the estimator of the Fourier coefficient  $\bar{\theta}_j$  ( $j = 1, 2, \dots$ ). This estimator is a shrinkage version of the unbiased estimator in (3.10), and the mean integrated squared error of  $\dot{\check{f}}$  is given by

$$\mathbb{E} \left[ \int_0^1 (\dot{\check{f}}(t) - \dot{f}(t))^2 dt \right] = \sum_{j=1}^{\infty} \frac{\bar{\theta}_j^2}{1 + \bar{\theta}_j^2}. \quad (3.31)$$

Comparing (3.29) and (3.31), we observe that the oracle estimator  $\dot{f}^*$ , constructed by applying Corollary 1, has a smaller mean integrated squared error than that

of the estimator  $\hat{f}$  defined in (3.30).

#### 4. Efficient Series Estimation from Correlated Data

In this section, we apply the results from the continuous-time model to construct optimal designs for a series estimation of the function  $f$  in model (2.1). In this transition from the continuous to the discrete model, we are faced with several challenges. First, the signed measure defining the optimal oracle estimator  $\hat{\theta}_j^*$  depends on the unknown function  $f$  through its Fourier coefficients and the constant  $c$ . Furthermore, the function  $f$  appears in the stochastic integrals in (3.19) and (3.20). Second, we need to address the problem that, even with preliminary knowledge of the function  $f$ , the stochastic integrals cannot be computed because the continuous-time process  $\{Y_t : t \in [0, 1]\}$  is not observable. In order to overcome these difficulties and construct an implementable estimator, that does not require preliminary knowledge of  $f$ , we proceed as follows. Roughly speaking, these steps consist of a two-stage estimation procedure, a truncation, and an appropriate approximation of the stochastic integrals by sums, which can be calculated from the available data. In the latter step of this procedure we also determine the optimal design points.

Throughout this section, we restrict ourselves to cases (A) and (B) of Theorem 1. For case (C), we simply propose to replace the parameter value (3.21) by the best linear unbiased estimator derived in Dette, Konstantinou and Zhigljavsky (2017).

##### 4.1. Truncation in the continuous-time model

In model (2.1), with  $n$  observations, only a finite number, say  $J$ , of Fourier coefficients in the series expansion (2.2) can be estimated. For this reason, we consider, for fixed  $J \in \mathbb{N}$ , the best  $L^2$ -approximation

$$f^{(J)}(t) = \sum_{j=1}^J \theta_j \varphi_j(t) = \Phi^{(J),T}(t) \theta^{(J)} \quad (4.1)$$

of the function  $f$  by functions from the span $\{\varphi_1, \dots, \varphi_J\}$  space, where the vectors  $\theta^{(J)}$  and  $\Phi^{(J)}$  are defined by  $\theta^{(J)} = (\theta_1, \dots, \theta_J)^T$  and  $\Phi^{(J)}(t) = (\varphi_1(t), \dots, \varphi_J(t))^T$ , respectively. We now replace the function  $f$  with the function  $f^{(J)}$  in the estimators  $\hat{\theta}_1^*, \dots, \hat{\theta}_J^*$ , defined in (3.19) and (3.20) for cases (A) and (B), respectively. In case (A), this gives the vector

$$\hat{\theta}^{(J),*} = \frac{\theta^{(J)}(\theta^{(J)})^T}{1 + c^{(J)}} \left\{ \int_0^1 \frac{d}{dt} \left[ \frac{\Phi^{(J)}(t)}{v(t)} \right] \left( \frac{d}{dt} q(t) \right)^{-1} d \left( \frac{Y_t}{v(t)} \right) + \frac{\Phi^{(J)}(0)}{u(0)} \frac{Y_0}{v(0)} \right\}, \quad (4.2)$$

where

$$c^{(J)} = (\theta^{(J)})^T C^{(J)} \theta^{(J)}, \quad (4.3)$$

and the  $J \times J$  matrix  $C^{(J)}$  is defined by

$$C^{(J)} = \int_0^1 \frac{d}{dt} \left[ \frac{\Phi^{(J)}(t)}{v(t)} \right] \left( \frac{d}{dt} \left[ \frac{\Phi^{(J)}(t)}{v(t)} \right] \right)^T \left( \frac{d}{dt} q(t) \right)^{-1} dt + \frac{\Phi^{(J)}(0)(\Phi^{(J)}(0))^T}{u(0)v(0)}. \quad (4.4)$$

Similarly, in case (B), we obtain

$$\hat{\theta}^{(J),*} = \frac{\theta^{(J)}(\theta^{(J)})^T}{1 + m^{(J)}} \left\{ \int_0^1 \frac{d}{dt} \left[ \frac{\Phi^{(J)}(t)}{v(t)} \right] \left( \frac{d}{dt} q(t) \right)^{-1} d \left( \frac{Y_t}{v(t)} \right) \right\}, \quad (4.5)$$

where

$$m^{(J)} = (\theta^{(J)})^T M^{(J)} \theta^{(J)}, \quad (4.6)$$

and the  $J \times J$  matrix  $M^{(J)}$  is given by

$$M^{(J)} = \int_0^1 \frac{d}{dt} \left[ \frac{\Phi^{(J)}(t)}{v(t)} \right] \left( \frac{d}{dt} \frac{\Phi^{(J)}(t)}{v(t)} \right)^T \left( \frac{d}{dt} q(t) \right)^{-1} dt. \quad (4.7)$$

The resulting estimators (4.2) and (4.5) still depend on the first  $J$  unknown Fourier coefficients  $\theta_1, \dots, \theta_J$ , and also depend on the full trajectory of the process  $\{Y_t : t \in [0, 1]\}$ . This dependence is removed in the following sections.

#### 4.2. Discrete approximation of stochastic integrals

In concrete applications, the integrals in (4.2) and (4.5) cannot be evaluated and have to be approximated from the given data. For this purpose, we assume that  $n$  observations  $Y_{t_1}, \dots, Y_{t_n}$  from model (2.1) at  $n$  distinct time points  $0 = t_1 < t_2 < \dots < t_{n-1} < t_n = 1$  are available. We consider the estimators

$$\hat{\theta}^{(J),n} = \frac{\theta^{(J)}(\theta^{(J)})^T}{1 + c^{(J)}} \left\{ \sum_{i=2}^n \mu_i \left( \frac{Y_{t_i}}{v(t_i)} - \frac{Y_{t_{i-1}}}{v(t_{i-1})} \right) + \frac{\Phi^{(J)}(0)}{u(0)} \frac{Y_0}{v(0)} \right\} \quad (4.8)$$

$$\hat{\theta}^{(J),n} = \frac{1}{1 + m^{(J)}} \theta^{(J)}(\theta^{(J)})^T \left\{ \sum_{i=2}^n \mu_i \left( \frac{Y_{t_i}}{v(t_i)} - \frac{Y_{t_{i-1}}}{v(t_{i-1})} \right) \right\}, \quad (4.9)$$

as approximations of the quantities in (4.2) and (4.5), respectively. Note that  $\hat{\theta}^{(J),*}$  depends on the full trajectory  $\{Y_t : t \in [0, 1]\}$ , while  $\hat{\theta}^{(J),n}$  is an approxi-

mation based on the sample  $\{Y_{t_i} : i = 1, \dots, n\}$ . In (4.8) and (4.9),  $\mu_2, \dots, \mu_n$  denote  $J$ -dimensional weights that depend on the time points  $0 = t_1 < t_2 < \dots < t_{n-1} < t_n = 1$ , and are chosen in an optimal way. In particular, we propose determining the weights  $\mu_2, \dots, \mu_n$ , such that the expected  $L^2$ -distance

$$\mathbb{E}[\|\hat{\theta}^{(J),*} - \hat{\theta}^{(J),n}\|^2] \tag{4.10}$$

between  $\hat{\theta}^{(J),*}$  and its discrete analogue  $\hat{\theta}^{(J),n}$  is minimized, where  $\|\cdot\|$  denotes the Euclidean norm in  $\mathbb{R}^J$ .

The following result provides an alternative expression of the expectation of this distance. The proof can be found in the Supplementary Material.

**Proposition 1.** *Assume that the conditions of Theorem 1 are satisfied. The Euclidean distance between the estimators  $\hat{\theta}^{(J),*}$  and  $\hat{\theta}^{(J),n}$  can be represented as*

$$\mathbb{E}[\|\hat{\theta}^{(J),*} - \hat{\theta}^{(J),n}\|^2] = k^{(J)} \{V(\mu_2, \dots, \mu_n) + B(\mu_2, \dots, \mu_n)\}, \tag{4.11}$$

where the quantities  $V$  and  $B$  are defined by

$$\begin{aligned} V(\mu_2, \dots, \mu_n) &= \text{tr} \left\{ \sum_{i=2}^n \int_{t_{i-1}}^{t_i} \left( \frac{d}{dt} \left[ \frac{\Phi^{(J)}(t)}{v(t)} \right] \right. \right. \\ &\quad \times \left. \left( \frac{d}{dt} q(t) \right)^{-1} - \mu_i \right) \left( \frac{d}{dt} \left[ \frac{\Phi^{(J)}(t)}{v(t)} \right] \left( \frac{d}{dt} q(t) \right)^{-1} - \mu_i \right)^T \left. \left( \frac{d}{dt} q(t) \right) dt \right\}, \\ B(\mu_2, \dots, \mu_n) &= \text{tr} \left\{ \sum_{i=2}^n \int_{t_{i-1}}^{t_i} \left( \frac{d}{dt} \left[ \frac{\Phi^{(J)}(t)}{v(t)} \right] \left( \frac{d}{dt} q(t) \right)^{-1} - \mu_i \right) \frac{d}{dt} \left[ \frac{f(t)}{v(t)} \right] dt \right. \\ &\quad \times \left. \left( \sum_{i=2}^n \int_{t_{i-1}}^{t_i} \left( \frac{d}{dt} \left[ \frac{\Phi^{(J)}(t)}{v(t)} \right] \left( \frac{d}{dt} q(t) \right)^{-1} - \mu_i \right) \left( \frac{d}{dt} \left[ \frac{f(t)}{v(t)} \right] \right) dt \right)^T \right\}, \end{aligned} \tag{4.12}$$

and  $k^{(J)}$  is given by  $\|\theta^{(J)}\|^4 / (1 + c^{(J)})^2$  in case (A) and  $\|\theta^{(J)}\|^4 / (1 + m^{(J)})^2$  in case (B).

Note that the expected  $L^2$ -distance in (4.11) only differs in the multiplicative factor  $k^{(J)}$  for cases (A) and (B), and that this factor does not depend on the vector-weights  $\mu_2, \dots, \mu_n$ . Therefore, optimal weights minimizing the expected  $L^2$ -distance can be determined without distinguishing between the two cases (A) and (B).

The function  $B$  in the criterion (4.11) still depends on the unknown regression function  $f$ , which we replace again by its truncation  $f^{(J)}$  defined in (4.1). The resulting criterion is given by

$$\Phi(\mu_2, \dots, \mu_n) = V(\mu_2, \dots, \mu_n) + B^{(J)}(\mu_2, \dots, \mu_n), \quad (4.13)$$

where

$$\begin{aligned} B^{(J)}(\mu_2, \dots, \mu_n) = & \operatorname{tr} \left\{ \sum_{i=2}^n \int_{t_{i-1}}^{t_i} \left( \frac{d}{dt} \frac{\Phi^{(J)}(t)}{v(t)} \left( \frac{d}{dt} q(t) \right)^{-1} - \mu_i \right) \frac{d}{dt} \frac{f^{(J)}(t)}{v(t)} dt \right. \\ & \left. \times \left( \sum_{i=2}^n \int_{t_{i-1}}^{t_i} \left( \frac{d}{dt} \left[ \frac{\Phi^{(J)}(t)}{v(t)} \right] \left( \frac{d}{dt} q(t) \right)^{-1} - \mu_i \right) \frac{d}{dt} \frac{f^{(J)}(t)}{v(t)} dt \right)^T \right\}. \end{aligned} \quad (4.14)$$

If we minimize the criterion  $\Phi$  in (4.13) without any further constraints, the resulting optimal weights depend on the unknown parameters  $\theta^{(J)}$ . We therefore impose a ‘‘controllable bias’’ condition, which ensures that the expected value of the approximation of the stochastic integral is equal (approximately equal) to the expected value of the stochastic integral.

We thus determine the optimal weights, such that the term  $B^{(J)}(\mu_2, \dots, \mu_n)$  in (4.13) vanishes for all potential Fourier coefficients  $\theta_1, \dots, \theta_J$  in the function  $f^{(J)}$ . Therefore, the optimal weights are obtained by minimizing  $\Phi$  in (4.13) under the constraint

$$\int_0^1 \frac{d}{dt} \left[ \frac{\Phi^{(J)}(t)}{v(t)} \right] \left( \frac{d}{dt} q(t) \right)^{-1} \frac{d}{dt} \left[ \frac{\Phi^{(J)}(t)}{v(t)} \right]^T dt = \sum_{i=2}^n \mu_i \int_{t_{i-1}}^{t_i} \frac{d}{dt} \left[ \frac{\Phi^{(J)}(t)}{v(t)} \right] dt. \quad (4.15)$$

In this situation, the criterion (4.13) reduces to the minimization of

$$\sum_{i=2}^n \int_{t_{i-1}}^{t_i} \operatorname{tr} \left\{ \left( \frac{d}{dt} \frac{\Phi^{(J)}(t)}{v(t)} \left( \frac{d}{dt} q(t) \right)^{-1} - \mu_i \right) \left( \frac{d}{dt} \frac{\Phi^{(J)}(t)}{v(t)} \left( \frac{d}{dt} q(t) \right)^{-1} - \mu_i \right)^T \frac{d}{dt} q(t) dt \right\}, \quad (4.16)$$

with respect to the weights  $\mu_2, \dots, \mu_n$  (depending on the time points  $0 = t_1 < t_2, \dots, t_{n-1} < t_n = 1$ ). In order to simplify this optimization, we introduce the following notation:

$$\beta_i = \frac{\Phi^{(J)}(t_i)/v(t_i) - \Phi^{(J)}(t_{i-1})/v(t_{i-1})}{\sqrt{q(t_i) - q(t_{i-1})}}, \quad \gamma_i = \mu_i \sqrt{q(t_i) - q(t_{i-1})}. \quad (4.17)$$

However, this does not reflect the dependence on the time points. Using the notation in (4.17), the approximation of the expected  $L^2$ -distance in (4.16) can be rewritten in terms of the quantities  $\gamma_2, \dots, \gamma_n$  as

$$\Psi(\gamma_2, \dots, \gamma_n) = -\operatorname{tr}(M^{(J)}) + \sum_{i=2}^n \gamma_i^T \gamma_i, \quad (4.18)$$

and the constraint (4.15) is given by

$$M^{(J)} = \sum_{i=2}^n \gamma_i \beta_i^T, \quad (4.19)$$

where  $M^{(J)}$  is the matrix defined in (4.7) (for both cases (A) and (B)). Note that the function  $\Psi$  and the constraint in (4.19) do not involve the function  $f$ , and only include assumptions related to the first  $J$  basis functions  $\varphi_1, \dots, \varphi_J$  used in the approximation  $f^{(J)}$ .

The resulting optimization problem (4.18) with constraint (4.19) has the same structure as an optimization problem considered in Dette, Konstantinou and Zhigljavsky (2017). From the results in the latter work, we obtain the solution

$$\gamma_i^* = M^{(J)} B^{-1} \beta_i, \quad i = 2, \dots, n, \quad (4.20)$$

where the matrix  $B$  is given by

$$B = \sum_{i=2}^n \beta_i \beta_i^T, \quad (4.21)$$

and  $M^{(J)}$  and  $\beta_i$  are defined in (4.7) and in (4.17), respectively. If the matrix  $B$  is singular, we replace the inverse  $B^{-1}$  in (4.20) with a generalized inverse  $B^-$ . Using the relation between  $\gamma_i$  and  $\mu_i$  in (4.17), we obtain the optimal weights  $\mu_i^* = (1/\sqrt{q(t_i) - q(t_{i-1})}) M^{(J)} B^{-1} \beta_i$  for  $i = 2, \dots, n$ . Note that these weights still depend on the design points  $t_2, \dots, t_{n-1}$ , which are determined next.

### 4.3. Optimal designs for series estimation

Using the optimal  $\gamma_2^*, \dots, \gamma_n^*$  given in (4.20) in the expression for the function  $\Psi$  defined in (4.18), we obtain an appropriate optimal design criterion for the choice of the time points  $0 = t_1 < t_2 < \dots < t_{n-1} < t_n = 1$ . More precisely, for the optimal weights, the function  $\Psi$  depends only on the design points and can be represented as the function

$$\tilde{\Psi}(t_2, \dots, t_{n-1}) = \text{tr}\{M^{(J)} B^{-1} M^{(J)}\}, \quad (4.22)$$

where the matrices  $B$  and  $M^{(J)}$  are defined in (4.21) and (4.7), respectively, and depend on  $0 = t_1 < t_2, \dots, t_{n-1} < t_n = 1$ . The optimal design is now determined by minimizing the function  $\tilde{\Psi}$ . This differs from the criterion considered in Dette, Konstantinou and Zhigljavsky (2017) for an unbiased linear estimation in the

linear regression model

$$Y_{t_i} = (\Phi^{(J)}(t_i))^T \theta + \varepsilon_{t_i}, \quad i = 1, \dots, n. \quad (4.23)$$

The optimal time points depend only on the first  $J$  basis functions, which are used for the estimator of the regression function  $f$ , and have to be determined numerically in all cases of practical interest. We present several examples in Section 5.

#### 4.4. The final estimate

With the optimal weights  $\mu_2^*, \dots, \mu_n^*$  determined in Section 4.2 and the optimal time points  $t_2^*, \dots, t_{n-1}^*$  determined in Section 4.3, the estimators in (4.8) and (4.9) corresponding to the cases (A) and (B) are given by

$$\hat{\theta}^{(J),n} = \frac{\theta^{(J)}(\theta^{(J)})^T}{1 + c^{(J)}} \left\{ M^{(J)} B^{-1} \sum_{i=2}^n \beta_i^* \left( \frac{Y_{t_i^*}}{v(t_i^*)} - \frac{Y_{t_{i-1}^*}}{v(t_{i-1}^*)} \right) + \frac{\Phi^{(J)}(0)}{u(0)} \frac{Y_0}{v(0)} \right\},$$

and

$$\hat{\theta}^{(J),n} = \frac{1}{1 + m^{(J)}} \theta^{(J)} (\theta^{(J)})^T M^{(J)} B^{-1} \sum_{i=2}^n \beta_i^* \left( \frac{Y_{t_i^*}}{v(t_i^*)} - \frac{Y_{t_{i-1}^*}}{v(t_{i-1}^*)} \right), \quad (4.24)$$

respectively, where  $\beta_i^* = (\Phi^{(J)}(t_i^*)/v(t_i^*) - \Phi^{(J)}(t_{i-1}^*)/v(t_{i-1}^*)) / \sqrt{q(t_i^*) - q(t_{i-1}^*)}$  ( $i = 2, \dots, n$ ).

For their application, we still require knowledge of the vector of Fourier coefficients  $\theta^{(J)}$  and the constants  $c^{(J)}$  and  $m^{(J)}$  defined in (4.3) and (4.6), respectively (note that these quantities also depend on  $\theta^{(J)}$ ). For this purpose, we propose using the linear unbiased estimate derived by Dette, Konstantinou and Zhigljavsky (2017) for the linear model (4.23). This estimate is defined as

$$\check{\theta}^{(J),n} = (C^{(J)})^{-1} \left\{ M^{(J)} B^{-1} \sum_{i=2}^n \beta_i^* \left( \frac{Y_{t_i^*}}{v(t_i^*)} - \frac{Y_{t_{i-1}^*}}{v(t_{i-1}^*)} \right) + \frac{\Phi^{(J)}(0)}{u(0)} \frac{Y_0}{v(0)} \right\}, \quad (4.25)$$

and the quantity  $c^{(J)}$  in (4.3) is estimated using  $\check{c}^{(J),n} = (\check{\theta}^{(J),n})^T C^{(J)} \check{\theta}^{(J),n}$ . A straightforward calculation shows that the resulting estimator for case (A) is given by

$$\hat{\theta}^{(J),n} = \frac{\check{\theta}^{(J),n} (\check{\theta}^{(J),n})^T}{1 + \check{c}^{(J),n}} \left\{ M^{(J)} B^{-1} \sum_{i=2}^n \beta_i^* \left( \frac{Y_{t_i^*}}{v(t_i^*)} - \frac{Y_{t_{i-1}^*}}{v(t_{i-1}^*)} \right) + \frac{\Phi^{(J)}(0)}{u(0)} \frac{Y_0}{v(0)} \right\}$$

$$= \frac{\check{\theta}^{(J),n}(\check{\theta}^{(J),n})^T}{1 + \check{c}^{(J),n}} C^{(J)} \check{\theta}^{(J),n} = \frac{\check{c}^{(J),n}}{1 + \check{c}^{(J),n}} \check{\theta}^{(J),n}, \tag{4.26}$$

which is a shrinkage version of the estimator  $\check{\theta}^{(J),n}$  in (4.25).

For case (B), similar arguments show that the estimator in (4.24) can also be rewritten in terms of the linear unbiased estimate  $\check{\theta}^{(J),n}$ ; that is,

$$\hat{\theta}^{(J),n} = \frac{\check{\theta}^{(J)}(\check{\theta}^{(J)})^T}{1 + \check{m}^{(J)}} M^{(J)} B^{-1} \sum_{i=2}^n \beta_i^* \left( \frac{Y_{t_i^*}}{v(t_i^*)} - \frac{Y_{t_{i-1}^*}}{v(t_{i-1}^*)} \right) = \frac{\check{m}^{(J),n}}{1 + \check{m}^{(J),n}} \check{\theta}^{(J),n},$$

where  $\check{m}^{(J),n} = (\check{\theta}^{(J),n})^T M^{(J)} \check{\theta}^{(J),n}$ . Here, the structure of the estimator  $\check{\theta}^{(J),n}$  depends on the structure of the basis functions contained in the vector  $\Phi^{(J)}$  (see Section 5 in Dette, Konstantinou and Zhigljavsky (2017) for more details).

**5. Numerical Results**

In this section, we illustrate the properties of the estimator and the corresponding optimal design derived in Section 4 by means of a small simulation study. We consider a Gaussian process, assuming both an exponential kernel and a Brownian motion as the error process in model (2.1). In both cases, we present the numerically calculated optimal time points with respect to the criterion defined in (4.22) and the corresponding simulated mean integrated squared errors for the proposed estimator

$$\hat{f}^{(J),n}(t) = \sum_{j=1}^J \hat{\theta}^{(J),n} \varphi_j(t) \tag{5.1}$$

and the estimator

$$\check{f}^{(J),n}(t) = \sum_{j=1}^J \check{\theta}^{(J),n} \varphi_j(t), \tag{5.2}$$

which is based on the best linear unbiased estimates in the truncated Fourier expansion.

Throughout this section, we use the trigonometric series defined in (3.22) as an orthonormal basis of  $L^2([0, 1])$ . We further assume that the unknown function  $f$  is symmetric on the interval  $[0, 1]$ , such that it is sufficient to use only the cosine functions in the series expansions of  $f$ . Note that this assumption is made for the sake of simplicity; similar results can be obtained for nonsymmetric functions using the full trigonometric basis from Example 1.

Consequently, the orthonormal system is given by  $\varphi_1(t) = 1$ ,  $\varphi_j(t) = \sqrt{2} \cos$

Table 1. Optimal time points for the series estimation minimizing the criterion (4.22). The covariance kernel is given by  $\exp(-L|s - t|)$ .

$L$	$n = 4$	$n = 7$
1	0.00, 0.25, 0.52, 1.00	0.00, 0.12, 0.27, 0.45, 0.57, 0.77, 1.00
5	0.00, 0.25, 0.51, 1.00	0.00, 0.12, 0.27, 0.45, 0.57, 0.76, 1.00

$2\pi(j - 1)t$  ( $j = 2, 3, \dots$ ). In Section 5.1 we consider the exponential kernel, whereas in Section 5.2, we concentrate on the Brownian motion.

### 5.1. The exponential kernel

We assume that the error process  $\{\varepsilon_t : t \in [0, 1]\}$  is a centered Gaussian process with an exponential kernel of the form  $K(s, t) = \exp(-L|s - t|)$ , where  $L \in \mathbb{R}^+$  is a given constant. This can be represented in the triangular form given in (3.3), with  $u(t) = \exp(Lt)$  and  $v(t) = \exp(-Lt)$ , and the function  $q$  is obtained as  $q(t) = u(t)/v(t) = \exp(2Lt)$ . Therefore, we have  $u(0) \neq 0$  (which corresponds to case (A)). The proposed estimator  $\hat{\theta}^{(J),n}$  is given by (4.26), and the corresponding estimators of the function  $f$  are defined in (5.1) and (5.2).

We first consider the exponential covariance kernel with  $L = 1$ . We assume that three basis functions, namely  $\varphi_1(t) = 1$ ,  $\varphi_2(t) = \cos(2\pi t)$ , and  $\varphi_3(t) = \cos(4\pi t)$ , are used in the series estimator, where  $n = 4$  and  $n = 7$  observations at different time points  $0 = t_1 < t_2 < \dots < t_{n-1} < t_n = 1$  can be taken. Note that one needs at least  $n = 4$  observations at different time points to guarantee that the matrix  $B$  in the preliminary estimator  $\check{\theta}^{(J),n}$  is nonsingular. The optimal points are determined by minimizing the criterion (4.22) using particle swarm optimization (see Clerc (2006) for details); the results are presented in the first row of Table 1.

We now evaluate the performance of the different estimators and the optimal time points by means of a simulation study. For the sake of comparison, we also consider non-optimized time points for the simulation, given by

$$0.00, 0.45, 0.90, 1.00 \quad (5.3)$$

$$0.00, 0.18, 0.36, 0.54, 0.72, 0.90, 1.00 \quad (5.4)$$

for the case  $n = 4$  and  $n = 7$ , respectively.

In the simulation study, we generate data according to model (2.1) using two regression functions

$$f(t) = 4t(t - 1), \quad (5.5)$$

$$f(t) = \sqrt{t(t - 1)} \quad (5.6)$$

Table 2. Simulated mean integrated squared errors of the estimators (5.1) and (5.2). The covariance kernel is given by  $\exp(-|s - t|)$ .

$f$	estimator	design ( $n = 4$ )		design ( $n = 7$ )	
		optimal	(5.3)	optimal	(5.4)
(5.5)	$\hat{f}^{(J),n}$	1.72	2.06	1.58	1.59
	$\check{f}^{(J),n}$	1.89	2.22	1.76	1.77
(5.6)	$\hat{f}^{(J),n}$	1.67	2.04	1.54	1.56
	$\check{f}^{(J),n}$	1.89	2.21	1.76	1.79

(note that both proposed functions are symmetric, with  $f(0) = f(1) = 0$ ). For each model, we determine the mean integrated squared error of the estimators  $\hat{f}^{(J),n}$  and  $\check{f}^{(J),n}$ , defined in (5.1) and (5.2), respectively. More precisely, if  $S$  denotes the number of simulation runs and  $\bar{f}_\ell$  is the estimator based on the  $\ell$ th run (either  $\hat{f}^{(J),n}$  and  $\check{f}^{(J),n}$ ), the simulated mean integrated squared error  $\text{MISE}_n = (1/S) \sum_{\ell=1}^S \int_0^1 (\bar{f}_\ell(t) - f(t))^2 dt$ , where  $f$ , the “true” regression function under consideration, is given by (5.5) or (5.6). All results are based on  $S = 1,000$  simulation runs.

For the case of the sample size  $n = 4$ , the resulting mean integrated squared error of the different estimators (and corresponding optimal time points) is shown in the left part of Table 2. For instance, the mean integrated squared error of the estimator  $\hat{f}^{(J),n}$  (based on the optimal design) is 1.72 if the true function is given by (5.5), but is 2.06 if the observations are taken according to the non-optimized design (5.3). Thus, the optimal design yields a 17% reduction in the mean integrated squared error. The optimal design also yields a reduction of 15% of the mean squared error of the preliminary estimator  $\check{f}^{(J),n}$  (although it is not constructed for this purpose). The results show that the new estimator  $\hat{f}^{(J),n}$  clearly outperforms the estimator  $\check{f}^{(J),n}$  in all cases under consideration (reduction of the mean squared error between 9% and 12%).

For the case of the sample size  $n = 7$ , the corresponding results are presented in the right part of Table 2, and we observe similar behavior. The new estimator  $\hat{f}^{(J),n}$  clearly outperforms  $\check{f}^{(J),n}$ , regardless of the design and model under consideration. On the other hand, the improvement from the choice of the design is less visible than when the sample size is  $n = 4$ . This means that the influence of the design on the performance of the estimators decreases with an increasing sample size. The reason for this observation is that in the models under consideration, the discrete model (2.1) already provides a good approximation of the continuous model (3.1) for the sample size  $n = 7$ . Because the full trajectory is available in this model, the design is negligible for sample sizes larger than 10. As a result,

Table 3. Simulated mean integrated squared errors of the estimators (5.1) and (5.2). The covariance kernel is given by  $\exp(-5|s - t|)$ .

$f$	estimator	design ( $n = 4$ )		design ( $n = 7$ )	
		optimal	(5.3)	optimal	(5.4)
(5.5)	$\hat{f}^{(J),n}$	0.65	2.13	0.47	0.51
	$\check{f}^{(J),n}$	0.77	2.30	0.58	0.62
(5.6)	$\hat{f}^{(J),n}$	0.64	2.09	0.43	0.43
	$\check{f}^{(J),n}$	0.81	2.30	0.59	0.59

a larger sample size does not decrease the mean integrated squared error substantially. A similar effect is observed by Dette, Konstantinou and Zhigljavsky (2017) for a linear regression model with correlated observations.

Next, we consider a situation in which the correlations between the observations are smaller, using the parameter  $L = 5$  for the exponential kernel. The time points minimizing the criterion  $\tilde{\Psi}$  in (4.22) are depicted in the second row of Table 1 for  $n = 4$  and  $n = 7$ . We observe that the optimal time points are similar to those for the constant  $L = 1$ . Further numerical results, not displayed for the sake of brevity, support these findings indicating robustness of the optimal design with respect to the choice of the parameter  $L$ . The simulated mean integrated squared error of the estimators  $\hat{f}^{(J),n}$  and  $\check{f}^{(J),n}$ , defined in (5.1) and (5.2), are displayed in Table 3 for sample sizes  $n = 4$  and  $n = 7$ . When the sample size is  $n = 4$ , the optimal design yields a substantial reduction in the mean squared errors of both estimators (between 65% and 70%). Compared with the case  $L = 1$  (see Table 2), the reduction is larger. When the sample size is  $n = 7$ , the mean integrated squared errors of the estimators based on the optimal time points are slightly smaller than those of the non-optimized time points. We observe again that the influence of the position of the time points, and thus of the design, decreases as the sample size  $n$  increases (see Table 3). A comparison of the estimators (5.1) and (5.2) shows again that the new estimator  $\hat{f}^{(J),n}$  outperforms  $\check{f}^{(J),n}$  in all cases under consideration (reduction of the mean squared error between 16% and 27%).

## 5.2. Brownian motion

We now consider the case in which the error process in (2.1) is given by a Brownian motion; that is  $K(s, t) = s \wedge t$ , which can be represented by  $K(s, t) = s$ ,  $s \leq t$ . Therefore, the functions  $u$  and  $v$  in (3.3) are given by  $u(t) = t$  and  $v(t) = 1$ , respectively, and the function  $q$  is obtained as  $q(t) = u(t)/v(t) = t$ . This situation corresponds to case (B), where  $u(0) = 0$  and  $f(0) = 0$ . The estimator

$\hat{\theta}^{(J),n}$  is given by

$$\hat{\theta}^{(J),n} = \frac{1}{1 + \check{m}^{(J)}} \check{\theta}^{(J)} (\check{\theta}^{(J)})^T M^{(J)} B^{-1} \sum_{i=2}^n \frac{(\Phi^{(J)}(t_i) - \Phi^{(J)}(t_{i-1}))^T}{\sqrt{t_i - t_{i-1}}} (Y_{t_i} - Y_{t_{i-1}}), \tag{5.7}$$

where the matrices  $M^{(J)}$  and  $B$  and the constant  $m^{(J)}$  are of the form

$$\begin{aligned} M^{(J)} &= \int_0^1 \dot{\Phi}^{(J)}(t) (\dot{\Phi}^{(J)}(t))^T dt, \\ B &= \sum_{i=2}^n \frac{(\Phi^{(J)}(t_i) - \Phi^{(J)}(t_{i-1})) (\Phi^{(J)}(t_i) - \Phi^{(J)}(t_{i-1}))^T}{t_i - t_{i-1}}, \\ \check{m}^{(J)} &= (\check{\theta}^{(J)})^T M^{(J)} \check{\theta}^{(J)}. \end{aligned}$$

Note that both the first row and the first column of the matrices  $M^{(J)}$  and  $B$  are zero (because  $\varphi_1(t) = 1$ ), such that both matrices are singular. Consequently, as proposed in Section 4, we use the generalized inverse

$$B^- = \begin{pmatrix} 0 & 0 \\ 0 & \tilde{B}^{-1} \end{pmatrix}$$

of  $B$ , where the matrix  $\tilde{B}$  is given by

$$\tilde{B} = \begin{pmatrix} \mathbf{0}_{(J-1)} & \mathbf{I}_{(J-1) \times (J-1)} \end{pmatrix} B \begin{pmatrix} \mathbf{0}_{(J-1)}^T \\ \mathbf{I}_{(J-1) \times (J-1)} \end{pmatrix}.$$

Here, the vector  $\mathbf{0}_{(J-1)}$  is of dimension  $(J - 1)$  with zero entries, where the matrix  $\mathbf{I}_{(J-1) \times (J-1)}$  is the  $(J - 1)$ , dimensional identity matrix. The estimator  $\check{\theta}^{(J),n}$  is obtained from Section 5.2 in Dette, Konstantinou and Zhigljavsky (2017),

$$\check{\theta}^{(J),n} = \underline{C}^{(J)} \sum_{i=2}^n \frac{(\Phi^{(J)}(t_i) - \Phi^{(J)}(t_{i-1}))^T}{\sqrt{t_i - t_{i-1}}} (Y_{t_i} - Y_{t_{i-1}}),$$

where the matrix  $\underline{C}^{(J)}$  is of the form

$$\underline{C}^{(J)} = \begin{pmatrix} 0 & -(\Phi^{(J)}(0))^T \begin{pmatrix} 0 \\ \tilde{B}^{-1} \end{pmatrix} \\ \mathbf{0}_{(J-1)} & \tilde{B}^{-1} \end{pmatrix}.$$

We now analyze the behavior of the resulting estimators of the function  $f$  if the first three basis functions are used for the series estimator and  $n = 4$  or

Table 4. Simulated mean integrated squared errors of the estimators (5.1) and (5.2). The error process is a Brownian motion.

$f$	estimator	design ( $n = 4$ )		design ( $n = 7$ )	
		optimal	(5.3)	optimal	(5.4)
(5.5)	$\hat{f}^{(J),n}$	0.16	0.41	0.13	0.14
	$\check{f}^{(J),n}$	0.15	0.43	0.12	0.12
(5.6)	$\hat{f}^{(J),n}$	0.13	0.45	0.11	0.11
	$\check{f}^{(J),n}$	0.15	0.48	0.12	0.13

$n = 7$  observations at different time points  $0 = t_1 < t_2 < \dots < t_{n-1} < t_n = 1$  are available. The optimal time points minimizing the criterion (4.22) derived in Section 4 are given by

$$0.00, 0.25, 0.47, 1.00 \quad (5.8)$$

$$0.00, 0.22, 0.28, 0.50, 0.72, 0.78, 1.00, \quad (5.9)$$

for sample sizes  $n = 4$  and  $n = 7$ , respectively. Note that the optimal time points (5.8) and (5.9) differ from those of the exponential kernel displayed in Table 1. This indicates that the position of the optimal time points depends on the structure of the covariance kernel.

The resulting mean integrated squared errors of the estimators  $\hat{f}^{(J),n}$  and  $\check{f}^{(J),n}$  are displayed in Table 4, where we again consider the comparative set of time points depicted in (5.3) and (5.4). We obtain similar results to those in Section 5.1. Specifically, for  $n = 4$ , we observe that the optimal design yields a substantial reduction in the mean squared errors of both estimators (see the left part of Table 4). When the sample size is  $n = 7$ , the difference between the optimal time points and the design (5.4) is less visible.

A comparison of the two estimators shows different behavior to that in Section 5.1; that is, unlike the case of an exponential kernel, when the error process is a Brownian motion, both estimators perform well and have similar (small) mean integrated squared errors (see Table 4).

Further simulation results are presented in the online Supplementary Material, where we consider the effect of the truncation parameter on the optimal design and on the two estimators. The results indicate some sensitivity of the optimal design with respect to the number of basis functions in the series estimator. Thus, an interesting problem for future research would be to construct optimal designs that address the uncertainty in the truncation parameter.

## Supplementary Material

The Supplementary Material contains further numerical results for different choices of the truncation parameter and proofs of Theorem 1 and Proposition 1.

## Acknowledgments

This work was supported in part by the Collaborative Research Center “Statistical modeling of nonlinear dynamic processes” (SFB 823, Teilprojekt C2) of the German Research Foundation (DFG). The authors would like to thank the two referees and the associate editor for their constructive comments on an earlier version of this paper.

## References

- Amo-Salas, M., López-Fidalgo, J. and López-Rios, V. (2012). Optimal designs for two nested pharmacokinetic models with correlated observations. *Communications in Statistics – Simulation and Computation* **41**, 944–963.
- Brown, L. D. and Low, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24**, 2384–2398.
- Clerc, M. (2006). *Particle Swarm Optimization*. Iste Publishing Company, London.
- Dette, H., Konstantinou, M. and Zhigljavsky, A. (2017). A new approach to optimal designs for correlated observations. *Ann. Statist.* **45**, 1579–1608.
- Dette, H., Pepelyshev, A. and Zhigljavsky, A. (2013). Optimal design for linear models with correlated observations. *Ann. Statist.* **41**, 143–176.
- Dette, H., Pepelyshev, A. and Zhigljavsky, A. (2016). Optimal designs in regression with correlated errors. *Ann. Statist.* **44**, 113–152.
- Dette, H. and Wiens, D. P. (2008). Robust designs for 3D shape analysis with spherical harmonic descriptors. *Statistica Sinica* **19**, 83–102.
- Doob, J. L. (1949). Heuristic Approach to the Kolmogorov-Smirnov Theorems. *Ann. Math. Statist.* **20**, 393–403.
- Efromovich, S. (1999). *Nonparametric Curve Estimation: Methods, Theory, and Applications*. Springer-Verlag, New York.
- Efromovich, S. (2008). Optimal Sequential Design in a Controlled Non-parametric Regression. *Scandinavian Journal of Statistics* **35**, 266–285.
- Grenander, U. (1950). Stochastic processes and statistical inference. *Ark. Mat.* **1**, 195–277.
- Harman, R. and Stulajter, F. (2010). Optimal prediction designs in finite discrete spectrum linear regression models. *Metrika* **72**, 281–294.
- Ibragimov, I. A. and Hasminskiĭ, R. Z. (1981). *Statistical Estimation*. Springer-Verlag, New York-Berlin.
- Müller, H.-G. (1984). Optimal designs for nonparametric kernel regression. *Statistics & Probability Letters* **2**, 285–290.
- Müller, W. G. and Stehlik, M. (2004). An example of  $D$ -optimal designs in the case of correlated errors. In *COMPSTAT 2004—Proceedings in Computational Statistics* 1543–1550. Physica, Heidelberg.

- Näther, W. and Simák, J. (2003). Effective observation of random processes. *Metrika* **58**, 71–84.
- Pázman, A. and Müller, W. G. (2001). Optimal design of experiments subject to correlated errors. *Statist. Probab. Lett.* **52**, 29–34.
- Pukelsheim, F. (2006). *Optimal Design of Experiments*. SIAM, Philadelphia.
- Rodríguez-Díaz, J. M. (2017). Computation of  $c$ -optimal designs for models with correlated observations. *Comp. Statist. and Data Analysis* **113**, 287–296.
- Sacks, J. and Ylvisaker, N. D. (1966). Designs for regression problems with correlated errors. *Ann. Math. Statist.* **37**, 66–89.
- Sacks, J. and Ylvisaker, N. D. (1968). Designs for regression problems with correlated errors; many parameters. *Ann. Statist.* **39**, 49–69.
- Stehlik, M., López-Fidalgo, J., Casero-Alonso, V. and Bukina, E. (2015). Robust integral compounding criteria for trend and correlation structures. *Stochastic Environmental Research and Risk Assessment* **29**, 379–395.
- Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer series in statistics. Springer, New York & London.
- Zhao, Z. and Yao, W. (2012). Sequential design for nonparametric inference. *Canadian Journal of Statistics* **40**, 362–377.

Kirsten Schorning

Technische Universität Dortmund, Fakultät für Statistik, 44221 Dortmund, Germany.

E-mail: schorning@statistik.tu-dortmund.de

Maria Konstantinou

Tickmill Europe Ltd, Kedron 9, Mesa Geitonia, P.C 4004 Limassol, Cyprus.

E-mail: m.konstantinou@tickmill.eu

Holger Dette

Fakultät für Mathematik, Ruhr-Universität Bochum, 44801 Bochum, Germany.

E-mail: holger.dette@rub.de

(Received December 2018; accepted November 2019)