Submitted to the Annals of Statistics

A NEW APPROACH TO OPTIMAL DESIGNS FOR CORRELATED OBSERVATIONS[*]

By Holger Dette, Maria Konstantinou and Anatoly Zhigljavsky

Ruhr-Universität Bochum and Cardiff University

This paper presents a new and efficient method for the construction of optimal designs for regression models with dependent error processes. In contrast to most of the work in this field, which starts with a model for a finite number of observations and considers the asymptotic properties of estimators and designs as the sample size converges to infinity, our approach is based on a continuous time model. We use results from stochastic analysis to identify the best linear unbiased estimator (BLUE) in this model. Based on the BLUE, we construct an efficient linear estimator and corresponding optimal designs in the model for finite sample size by minimizing the mean squared error between the optimal solution in the continuous time model and its discrete approximation with respect to the weights (of the linear estimator) and the optimal design points, in particular in the multi-parameter case.

In contrast to previous work on the subject, the resulting estimators and corresponding optimal designs are very efficient and easy to implement. This means that they are practically not distinguishable from the weighted least squares estimator and the corresponding optimal designs, which have to be found numerically by non-convex discrete optimization. The advantages of the new approach are illustrated in several numerical examples.

1. Introduction. The construction of optimal designs for dependent observations is a very challenging problem in statistics, because - in contrast to the independent case - the dependency yields non-convex optimization problems. As a consequence, classical tools of convex optimization theory as described, for example, in Pukelsheim (2006) are not applicable. Most of the discussion is restricted to very simple models and we refer to Dette, Kunert

[*]Supported by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Teilprojekt C2) of the German Research Foundation (DFG). The research of H. Dette reported in this publication was also partially supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R01GM107639.

MSC 2010 subject classifications: Primary 62K05; secondary 62M05

Keywords and phrases: linear regression, correlated observations, optimal design, Gaussian white mouse model, Doob representation, quadrature formulas

and Pepelyshev (2008); Kiseľák and Stehlík (2008); Harman and Štulajter (2010) for some exact optimal designs for linear regression models. Several authors have proposed to determine optimal designs using asymptotic arguments [see, for example, Sacks and Ylvisaker (1966, 1968), Bickel and Herzberg (1979), Näther (1985a), Zhigljavsky, Dette and Pepelyshev (2010)], but the resulting approximate optimal design problems are still non-convex and extremely difficult to solve. As a consequence, approximate optimal designs have mainly been determined analytically for the location model (in this case the corresponding optimization problems are in fact convex) and for a few one-parameter linear models [see Boltze and Näther (1982), Näther (1985a), Ch. 4, Näther (1985b), Pázman and Müller (2001) and Müller and Pázman (2003) among others].

Recently, substantial progress has been made in the construction of optimal designs for regression models with a dependent error process. Dette, Pepelyshev and Zhigljavsky (2013) determined (asymptotic) optimal designs for least squares estimation, under the additional assumption that the regression functions are eigenfunctions of an integral operator associated with the covariance kernel of the error process. Although this approach is able to deal with the multi-parameter case, the class of models for which approximate optimal designs can be determined explicitly is still rather small, because it refers to specific kernels with corresponding eigenfunctions. For this reason Dette, Pepelyshev and Zhigljavsky (2016) proposed a different strategy to obtain optimal designs and efficient estimators. Instead of constructing an optimal design for a particular estimator (such as least squares or weighted least squares), these authors proposed to consider the problem of optimizing the estimator and the design of experiment simultaneously. They constructed a class of estimators and corresponding optimal designs with a variance converging (as the sample size increases) to the optimal variance in the continuous time model. In other words, asymptotically these estimators achieve the same precision as the best linear unbiased estimator computed from the whole trajectory of the process. While this approach yields a satisfactory solution for one-dimensional parametric models using signed least squares estimators, it is not transparent and in many cases not efficient in the multi-parameter model. In particular, it is based on matrix-weighted linear estimators and corresponding designs which are difficult to implement in practice and do not yield the same high efficiencies as in the one-parameter case.

In this paper we present an alternative approach for the construction of estimators and corresponding optimal designs for regression models with dependent error processes, which has important advantages compared to the currently used methodology. First - in contrast to all other methods - the estimators with corresponding optimal designs proposed here are very easy to implement. Secondly, it is demonstrated that the new estimator and design yield a method which is practically not distinguishable from the best linear estimator (BLUE) with corresponding optimal design. Third, in many cases the new estimator and a uniform design are already very efficient.

Compared to most of the work in this field, which begins with a model for a finite number of observations and considers the asymptotic properties of estimators as the sample size converges to infinity, an essential difference of our approach is that it is directly based on the continuous time model. In Section 2 we derive the best linear unbiased estimate in this model using results about the absolute continuity of measures on the space $C([a,b])$. This yields a representation of the best linear estimator as a stochastic integral and provides an efficient tool for constructing estimators with corresponding optimal designs for finite samples which are practically not distinguishable from the optimal (weighted least squares) estimator and corresponding optimal design. We emphasize again that the latter design has to be determined by discrete non-convex optimization. To be more precise, in Section 3 we propose a weighted mean, say $\sum_{i=1}^{n} \mu_i Y_{t_i}$ (here Y_{t_i} denotes the response at the point t_i and n is the sample size), where the weights $\mu_1, \ldots, \mu_n$ (which are vectors in case of models with more than one parameter) and design points $t_1, \ldots, t_n$ are determined by minimizing the mean squared error between the optimal solution in the continuous time model (represented by a stochastic integral with respect to the underlying process) and its discrete approximation with respect to the weights (of the linear estimator) and the optimal design points. In Section 4 we discuss several examples and demonstrate the superiority of the new approach to the method which was recently proposed in Dette, Pepelyshev and Zhigljavsky (2016), in particular for multi-parameter models. Some more details on best linear unbiased estimation in the continuous time model are given in Section 5, where we discuss degenerate cases, which appear, for example, due to the presence of a constant term in the regression function. For a more transparent presentation of the ideas, some technical details are additionally deferred to the Appendix.

We finally note that this paper is a first approach which uses results from stochastic analysis in the context of optimal design theory. The combination of these two fields yields a practically implementable and satisfactory solution of optimal design problems for a broad class of regression models with dependent observations.

2. Optimal estimation in continuous time models. Consider a linear regression model of the form

$$Y_{t_i} = Y(t_i) = \theta^T f(t_i) + \varepsilon_{t_i}, \qquad i = 1, \dots, n\,, \tag{2.1}$$

where $\{\varepsilon_t \mid t \in [a,b]\}$ is a Gaussian process with $\mathbb{E}[\varepsilon_{t_i}] = 0$ and $K(t_i, t_j) = \mathbb{E}[\varepsilon_{t_i}\varepsilon_{t_j}]$ denoting the covariance between observations at the points t_i and t_j $(i, j = 1, \dots, n)$. Furthermore, $\theta = (\theta_1, \dots, \theta_m)^T$ is a vector of unknown parameters, $f(t) = (f_1(t), \dots, f_m(t))^T$ is a vector of continuously differentiable linearly independent functions, and the explanatory variables $t_1, \dots, t_n$ vary in a compact interval, say $[a,b]$. If $\mathbf{Y} = (Y_{t_1}, \dots, Y_{t_n})^T$ denotes the vector of observations, the weighted least squares estimator of θ is defined by

$$\widehat{\theta}_{WLSE} = (\mathbf{X}^T \mathbf{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{\Sigma}^{-1} \mathbf{Y},$$

where $\mathbf{X} = (f_p(t_j))_{j=1,\dots,n}^{p=1,\dots,m}$ is the $n \times m$ design matrix and the $n \times n$ matrix $\mathbf{\Sigma} = (K(t_i, t_j))_{i,j=1,\dots,n}$ is the matrix of variances/covariances. It is well known that $\hat{\theta}_{WLSE}$ is the BLUE in model (2.1). The corresponding minimal variance is given by

$$\mathrm{Var}(\widehat{\theta}_{WLSE}) = (\mathbf{X}^T \mathbf{\Sigma}^{-1}\, \mathbf{X})^{-1}, \tag{2.2}$$

and an optimal design for the estimation of the parameter vector θ in model (2.1) minimizes an appropriate real-valued functional of this matrix. As pointed out before, the direct minimization of this type of criterion is an extremely challenging non-convex discrete optimization problem and explicit solutions are not available in nearly all cases of practical interest. For this reason many authors propose to consider asymptotic optimal designs as the sample size n converges to infinity [see Sacks and Ylvisaker (1966, 1968), Bickel and Herzberg (1979), Näther (1985a), Zhigljavsky, Dette and Pepelyshev (2010)].

In the following discussion we consider - parallel to model (2.1) - its *continuous time* version, that is

$$Y_t = \theta^T f(t) + \varepsilon_t\ , \quad t \in [a,b], \tag{2.3}$$

where the full trajectory of the process $\{Y_t \mid t \in [a,b]\}$ can be observed and $\{\varepsilon_t \mid t \in [a,b]\}$ is a centered Gaussian process with continuous covariance kernel K, that is, $K(t,t') = \mathbb{E}[\varepsilon_t \varepsilon_{t'}]$. We will focus on *triangular kernels*, which are of the form

$$K(t,t') = u(t)v(t') \quad \text{for} \quad t \le t', \tag{2.4}$$

($K(t,t') = K(t',t)$ for $t > t'$), where $u(\cdot)$ and $v(\cdot)$ are some functions defined on the interval $[a,b]$. An alternative representation of K is given by

$$K(t,t') = v(t)v(t')\min\{q(t), q(t')\}; \quad (t,t' \in [a,b]),$$

where $q(t) = u(t)/v(t)$. We assume that the process $\{\varepsilon_t |\ t \in [a,b]\}$ is non-degenerate on the open interval (a,b), which implies that the function q is positive on the interval (a,b) and strictly increasing and continuous on $[a,b]$ [see Mehr and McFadden (1965) for more details]. Consequently, the functions u and v must have the same sign and can be assumed to be positive on the interval (a,b) without loss of generality. Note that the majority of covariance kernels considered in the literature belong to this class, see, for example, Näther (1985a); Zhigljavsky, Dette and Pepelyshev (2010) or Harman and Štulajter (2011). The simple triangular kernel

$$K(t,t') = t \wedge t',$$

is obtained for the choice $u(t) = t$ and $v(t) = 1$ and corresponds to the Brownian motion. As pointed out in Dette, Pepelyshev and Zhigljavsky (2016), the solutions of the optimal design problems with respect to different triangular kernels are closely related. In particular, if a best linear unbiased estimator (BLUE) for a particular triangular kernel has to be found for the continuous time model, it can be obtained by simple nonlinear transformation from the BLUE in a different continuous time model (on a possibly different interval) with a Brownian motion as error process (see Remark 2.1(b) below for more details). For this reason we will concentrate on the covariance kernel of the Brownian motion throughout this section. Our first result, presented in Theorem 2.1 provides the optimal estimator in the continuous time model (2.3), where the error process is given by a Brownian motion on the interval $[a,b]$, where $a > 0$ (the case $a = 0$ will be discussed in Section 5). We begin with a lemma which is crucial for the definition of the estimator. The proof can be found in the Appendix.

LEMMA 2.1. *Consider the continuous time linear regression model* (2.3) *on the interval* $[a,b]$, $a > 0$, *with a continuously differentiable vector of regression functions* f *and a Brownian motion as error process. Then the* $m \times m$ *matrix*

$$C = \int_a^b \dot{f}(t)\dot{f}^T(t)\,dt + \frac{f(a)f^T(a)}{a}, \tag{2.5}$$

is non-singular.

THEOREM 2.1. *Consider the continuous time linear regression model* (2.3) *on the interval* $[a,b]$, $a>0$, *with a continuously differentiable vector of regression functions* f *and a Brownian motion as error process. The best linear unbiased estimate is given by*

$$\hat{\theta}_{\mathrm{BLUE}} = C^{-1}\Big(\int_a^b \dot{f}(t)\,dY_t + \frac{f(a)}{a}Y_a\Big). \tag{2.6}$$

Moreover, the minimum variance is given by

$$C^{-1} = \Big(\int_a^b \dot{f}(t)\dot{f}^T(t)\,dt + \frac{f(a)f^T(a)}{a}\Big)^{-1}. \tag{2.7}$$

PROOF. Note that the continuous time model (2.3) can be written as a Gaussian white noise model

$$Y_t = \int_0^t s_1(u)\,du + \int_0^t d\varepsilon_u, \quad t\in[0,b],$$

where the function s_1 is defined as

$$s_1(u) = I_{[a,b]}(u)\theta^T\dot{f}(u) + I_{[0,a]}(u)\frac{\theta^T f(a)}{a}.$$

Let $\mathbb{P}_\theta$ and $\mathbb{P}_0$ denote the measure on $C([0,b])$ associated with the process $Y=\{Y_t|\ t\in[0,b]\}$ and $\{\varepsilon_t|\ t\in[0,b]\}$, respectively. From Theorem 1 in Appendix II of Ibragimov and Hasminskiĭ (1981), it follows that $\mathbb{P}_\theta$ is absolute continuous with respect to $\mathbb{P}_0$ with Radon-Nikodym derivative given by

$$\begin{aligned}\frac{d\mathbb{P}_\theta}{d\mathbb{P}_0}(Y) &= \exp\left\{\int_0^b s_1(t)\,dY_t - \frac{1}{2}\int_0^b s_1^2(t)\,dt\right\}\\ &= \exp\left\{\left(\int_a^b \theta^T\dot{f}(t)\,dY_t + \frac{\theta^T f(a)}{a}Y_a\right) - \frac{1}{2}\left(\int_a^b (\theta^T\dot{f}(t))^2\,dt + \frac{(\theta^T f(a))^2}{a}\right)\right\}.\end{aligned}$$

The maximum likelihood estimator can be determined by solving the equation

$$\frac{\partial}{\partial\theta}\log\frac{d\mathbb{P}_\theta}{d\mathbb{P}_0}(Y) = \int_a^b \dot{f}(t)\,dY_t + \frac{f(a)}{a}Y_a - \Big(\int_a^b \dot{f}(t)\dot{f}^T(t)\,dt + \frac{f(a)f^T(a)}{a}\Big)\theta = 0.$$

The solution coincides with the linear estimate (2.6), and a straightforward calculation, using Ito's formula and the fact that the random variables $\int_a^b \dot{f}(t)d\varepsilon_t$ and ε_a are independent, gives

$$\begin{aligned}\mathrm{Var}_\theta(\hat{\theta}_{\mathrm{BLUE}}) &= C^{-1}\mathbb{E}_\theta\Big[\Big(\int_a^b \dot{f}(t)d\varepsilon_t + \frac{f(a)}{a}\varepsilon_a\Big)\Big(\int_a^b \dot{f}(t)d\varepsilon_t + \frac{f(a)}{a}\varepsilon_a\Big)^T\Big]C^{-1}\\ &= C^{-1}\Big(\int_a^b \dot{f}(t)\dot{f}^T(t)dt + \frac{f(a)f^T(a)}{a}\Big)C^{-1} = C^{-1},\end{aligned}$$

where the matrix C is defined in (2.5). It has been shown in Dette, Pepelyshev and Zhigljavsky (2016) that C^{-1} is the variance/covariance matrix of the BLUE in the continuous time model, which proves Theorem 2.1. $\square$

REMARK 2.1.

(a) For a twice continuously differentiable vector of regression functions and Brownian motion as error process, Dette, Pepelyshev and Zhigljavsky (2016) determined the best linear estimator for the continuous time linear regression model (2.3) as

$$C^{-1}\Big\{\dot{f}(b)Y_b + \Big(\frac{f(a)}{a} - \dot{f}(a)\Big)Y_a - \int_a^b \ddot{f}(t)Y_t dt\Big\}. \tag{2.8}$$

Using integration by parts gives

$$\int_a^b \dot{f}(t)\,dY_t = \dot{f}(b)Y_b - \dot{f}(a)Y_a - \int_a^b \ddot{f}(t)Y_t\,dt,$$

and it is easily seen that the expression (2.8) coincides with (2.6). This means that a BLUE in the continuous time model (2.3) is even available under the weaker assumption of a once continuously differentiable function f.

(b) The best linear estimator in the continuous time model (2.3) with a general triangular kernel of the form (2.4) can easily be obtained from Appendix B in Dette, Pepelyshev and Zhigljavsky (2016). To be precise, consider a triangular kernel of the form (2.4), define

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

and consider the stochastic process

$$\varepsilon_t = \alpha(t)\tilde{\varepsilon}_{q(t)},$$

where $\{\tilde{\varepsilon}_{\tilde{t}} |\ \tilde{t} \in [\tilde{a}, \tilde{b}]\}$ is a Brownian motion on the interval $[\tilde{a}, \tilde{b}]$ and $\tilde{a} = q(a)$, $\tilde{b} = q(b)$. It follows from Doob (1949) that $\{\varepsilon_t |\ t \in [a, b]\}$ is a centered Gaussian process on the interval $[a, b]$ with covariance kernel (2.4). Moreover, if we consider the continuous time model

$$\tilde{Y}_{\tilde{t}} = \theta^T \tilde{f}(\tilde{t}) + \tilde{\varepsilon}_{\tilde{t}}, \quad \tilde{t} \in [\tilde{a}, \tilde{b}], \tag{2.9}$$

and use the transformations

$$\tilde{f}(\tilde{t}) = \frac{f(q^{-1}(\tilde{t}))}{v(q^{-1}(\tilde{t}))}\ ,\ \tilde{\varepsilon}_{\tilde{t}} = \frac{\varepsilon_{q^{-1}(\tilde{t})}}{v(q^{-1}(\tilde{t}))}\ ,\ \tilde{Y}_{\tilde{t}} = \frac{Y_t}{v(t)}, \tag{2.10}$$

then it follows from Dette, Pepelyshev and Zhigljavsky (2016) that the BLUE for the continuous time model (2.3) (with a general triangular covariance kernel) can be obtained from the BLUE in model (2.9) by the transformation $\tilde{t} = q(t)$. Therefore, an application of Theorem 2.1 gives for the best linear estimator in the continuous time model (2.3) with triangular covariance kernel of the form (2.4) the representation

$$\hat{\theta}_{\mathrm{BLUE}} = C^{-1}\Big[\int_a^b \frac{\dot{f}(t)v(t) - \dot{v}(t)f(t)}{\dot{u}(t)v(t) - u(t)\dot{v}(t)}\, d\left(\frac{Y_t}{v(t)}\right) + \frac{f(a)}{u(a)v(a)}Y_a\Big]\ ,$$

where the matrix C is given by

$$C = \int_a^b \frac{[\dot{f}(t)v(t) - \dot{v}(t)\dot{f}(t)][\dot{f}(t)v(t) - \dot{v}(t)\dot{f}(t)]^T}{v^2(t)[\dot{u}(t)v(t) - u(t)\dot{v}(t)]}\, dt + \frac{f(a)f^T(a)}{u(a)v(a)}.$$

(c) Using integration by parts it follows (provided that the functions f, u, and v are twice continuously differentiable) that the BLUE in the continuous time model (2.3) can be represented as

$$\hat{\theta}_{\mathrm{BLUE}} = \int_a^b Y_t\, \mu^*(dt),$$

where μ^* is a vector of signed measures defined by $\mu^*(dt) = P_a\delta_a + p(t)dt + P_b\delta_b$, δ_t denotes the Dirac measure at the point $t \in [0,1]$ and the "masses" P_a, P_b and the density p are given by

$$P_a = C^{-1}\frac{1}{u(a)}\frac{f(a)\dot{u}(a) - \dot{f}(a)u(a)}{\dot{u}(a)v(a) - u(a)\dot{v}(a)}\ ,\ P_b = C^{-1}\frac{1}{v(b)}\frac{\dot{f}(b)v(b) - \dot{v}(b)f(b)}{\dot{u}(b)v(b) - u(b)\dot{v}(b)},$$
$$p(t) = -C^{-1}\frac{d}{dt}\Big(\frac{1}{v(t)}\frac{\dot{f}(t)v(t) - \dot{v}(t)f(t)}{\dot{u}(t)v(t) - u(t)\dot{v}(t)}\Big)\frac{1}{v(t)},$$

respectively. Now, if $\hat{\theta}_n = \sum_{i=1}^n \omega_i Y_{t_i}$ denotes an unbiased linear estimate in model (2.1) with vectors $\omega_i \in \mathbb{R}^m$, we can represent this estimator as

$$\hat{\theta}_n = \int_a^b Y_t\, \hat{\mu}_n(dt),$$

in the continuous time model (2.3), where $\hat{\mu}_n$ is a discrete signed vector valued measure with "masses" ω_i at the points t_i. Consequently, we obtain from Theorem 2.1 that

$$C^{-1} = \mathrm{Var}(\hat{\theta}_{\mathrm{BLUE}}) \leq \mathrm{Var}(\hat{\theta}_n),$$

(in the Loewner ordering). In other words, C^{-1} is a lower bound for any linear estimator in the linear regression model (2.1).

3. Optimal estimators and designs for finite sample size. We have determined the BLUE and corresponding minimal variance/covariance matrix in the continuous time model (2.3). In the present section we now explain how the particular representation of the BLUE as a stochastic integral can be used to derive efficient estimators and corresponding optimal designs in the original model (2.1), which are practically not distinguishable from the BLUE in model (2.1) based on an optimal design. Our approach is based on a comparison of the mean squared error of the difference between the best linear unbiased estimator derived in Theorem 2.1 and a discrete approximation of the stochastic integral in (2.6). For the sake of a clear representation, we discuss the one-dimensional case first.

3.1. *One-parameter models.* Consider the estimator $\hat{\theta}_{\mathrm{BLUE}}$ defined by (2.6) for the continuous time model (2.3) with $m = 1$ and define an estimator $\hat{\theta}_n$ in the original regression model (2.1) by an approximation of the stochastic integral, that is

$$\hat{\theta}_n = C^{-1}\Big\{ \sum_{i=2}^{n} \omega_i \dot{f}(t_{i-1})(Y_{t_i} - Y_{t_{i-1}}) + \frac{f(a)}{a} Y_a \Big\}. \tag{3.1}$$

Here $a = t_1 < t_2 < \ldots < t_{n-1} < t_n = b$ are n design points in the interval $[a, b]$ and $\omega_2, \ldots, \omega_n$ are corresponding (not necessarily positive) weights. Obviously, the estimator depends on the weights ω_i only through the quantities $\mu_i = \omega_i \dot{f}(t_{i-1})$ and therefore we use the notation

$$\hat{\theta}_n = C^{-1}\Big\{ \sum_{i=2}^{n} \mu_i (Y_{t_i} - Y_{t_{i-1}}) + \frac{f(a)}{a} Y_a \Big\}, \tag{3.2}$$

in the following discussion. We will determine optimal weights $\mu_2^*, \ldots, \mu_n^*$ and optimal design points $t_2^*, \ldots, t_{n-1}^*$ minimizing the mean squared error $\mathbb{E}[(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)^2]$ between the estimators $\hat{\theta}_{\mathrm{BLUE}}$ and $\hat{\theta}_n$. Our first result provides an explicit expression for this quantity. The proof is omitted because we prove a more general result in the multi-parameter case (see Section A.3).

LEMMA 3.1. *Consider the continuous time model (2.3) in the one-dimensional case. If the assumptions of Theorem 2.1 are satisfied, then*

$$\begin{aligned} \mathbb{E}_\theta[(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)^2] = & C^{-1}\Big\{ \sum_{i=2}^{n} \int_{t_{i-1}}^{t_i} \big[\dot{f}(s) - \mu_i\big]^2 \, ds \\ & + \theta^2 \Big(\sum_{i=2}^{n} \int_{t_{i-1}}^{t_i} \big[\dot{f}(s) - \mu_i\big] \dot{f}(s)\, ds \Big)^2 \Big\} C^{-1}. \end{aligned} \tag{3.3}$$

In order to find "good"weights for the linear estimator $\hat{\theta}_n$ in (3.1) we propose to consider only estimators with weights $\mu_2, \dots, \mu_n$ such that the second term in (3.3) vanishes, that is

$$\sum_{i=2}^{n} \int_{t_{i-1}}^{t_i} \big[\dot{f}(s) - \mu_i\big] \dot{f}(s)\, ds = 0. \tag{3.4}$$

Theorem 3.1 shows that this condition is equivalent to the property that the estimator $\hat{\theta}_n$ in (3.1) is unbiased. Furthermore, under this constraint the minimization of $\mathbb{E}_\theta[(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)^2]$ is proven to be equivalent to the minimization of the variance of the proposed estimator, that is, $\mathbb{E}_\theta[(\hat{\theta}_n - \theta)^2]$. The general result for the multi-parameter case is proven in the Appendix (see section A.4) and therefore, the proof of Theorem 3.1 is omitted.

THEOREM 3.1. *The estimator $\hat{\theta}_n$ defined in (3.1) is unbiased if and only if the identity*

$$\int_a^b [\dot{f}(s)]^2\, ds = \sum_{i=2}^{n} \mu_i \int_{t_{i-1}}^{t_i} \dot{f}(s)\, ds = \sum_{i=2}^{n} \mu_i (f(t_i) - f(t_{i-1})), \tag{3.5}$$

is satisfied. Moreover, for any linear unbiased estimator of the form $\tilde{\theta}_n = \int_a^b g(s) dY_s$ we have

$$\mathbb{E}_\theta[(\tilde{\theta}_n - \theta)^2] = \mathbb{E}_\theta[(\tilde{\theta}_n - \hat{\theta}_{\mathrm{BLUE}})^2] + C^{-1}.$$

The following result describes the weights minimizing $\mathbb{E}[(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)^2]$ under the constraint (3.4).

LEMMA 3.2. *Consider the continuous time model (2.3) in the one-dimensional case. If the assumptions of Theorem 2.1 are satisfied, then the optimal weights minimizing $\mathbb{E}[(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)^2]$ in the class of all unbiased linear estimators of the form (3.1) are given by*

$$\mu_i^* = \kappa(t_1, \dots, t_n) \frac{f(t_i) - f(t_{i-1})}{t_i - t_{i-1}}\,, \tag{3.6}$$

where

$$\kappa(t_1, \dots, t_n) = \frac{\int_a^b [\dot{f}(s)]^2\, ds}{\sum_{j=2}^{n} [f(t_j) - f(t_{j-1})]^2 / (t_j - t_{j-1})}\,.$$

PROOF. Under the condition (3.4) the mean squared error simplifies to

$$\begin{aligned}\mathbb{E}_\theta[(\hat\theta_{\mathrm{BLUE}}-\hat\theta_n)^2] &= C^{-1}\Big\{\sum_{i=2}^n\int_{t_{i-1}}^{t_i}\big[\dot f(s)-\mu_i\big]^2\,ds\Big\}C^{-1}\\ &= C^{-1}\Big\{-\int_a^b[\dot f(s)]^2\,ds+\sum_{i=2}^n\mu_i^2(t_i-t_{i-1})\Big\}C^{-1}.\end{aligned}$$

Using Lagrangian multiplies to minimize this expression subject to the constraint (3.5) yields

$$\mu_i=\frac{\lambda[f(t_i)-f(t_{i-1})]}{2(t_i-t_{i-1})}\,,\qquad i=2,\dots,n,$$

where λ denotes the Lagrangian multiplier. Substituting this into (3.4) gives

$$\lambda/2=\frac{\int_a^b[\dot f(s)]^2\,ds}{\sum_{i=2}^n[f(t_i)-f(t_{i-1})]^2/(t_i-t_{i-1})}=\kappa(t_1,\dots,t_n).$$

Therefore, the optimal weights are given by (3.6). □

Inserting these optimal weights in the mean squared error gives the function

$$\begin{aligned}\mathbb{E}_\theta[(\hat\theta_{\mathrm{BLUE}}-\hat\theta_n)^2] =&C^{-1}\Big\{\Big(\int_a^b[\dot f(s)]^2ds\Big)^2\Big\{\sum_{i=2}^n\frac{(f(t_i)-f(t_{i-1}))^2}{t_i-t_{i-1}}\Big\}^{-1}\\ &-\int_a^b[\dot f(s)]^2ds\Big\}C^{-1},\end{aligned}$$

which finally has to be minimized by the choice of the design points $t_2,\dots,t_{n-1}$. Because we discuss the one-parameter case in this section and the matrix C does not depend on $t_2,\dots,t_n$, this optimization corresponds to the minimization of

$$\Phi(t_1,\dots,t_n)=\Big(\int_a^b[\dot f(s)]^2ds\Big)\Big\{\sum_{i=2}^n\frac{(f(t_i)-f(t_{i-1}))^2}{t_i-t_{i-1}}\Big\}^{-1}-1. \tag{3.7}$$

REMARK 3.1. Let

$$\begin{aligned}\mathrm{eff}(t_2,\dots,t_{n-1}) &= \frac{\mathrm{Var}_\theta(\hat\theta_{\mathrm{BLUE}})}{\mathrm{Var}_\theta(\hat\theta_n)]}=\frac{C^{-1}}{C^{-1}\int_a^b[\dot f(s)]^2ds\,\Phi(t_1,\dots,t_n)C^{-1}+C^{-1}}\\ &= \left(1+\frac{\Phi(t_1,\dots,t_n)}{1+\frac{f^2(a)}{a}/\int_a^b[\dot f(s)]^2ds}\right)^{-1},\end{aligned}$$

denote the efficiency of an estimator $\hat{\theta}_n$ defined by (3.1) with optimal weights. Note that from the proof of Lemma 3.2 it follows that the function Φ is non-negative for all $t_1, \ldots, t_n$. Consequently, minimizing Φ with respect to the design points means that $t_1 = a < t_2 < \ldots < t_{n-1} < t_n = b$ have to be determined such that

$$\sum_{i=2}^{n} \frac{(f(t_i) - f(t_{i-1}))^2}{t_i - t_{i-1}},$$

approximates the integral $\int_a^b [\dot{f}(s)]^2 ds$ most precisely (this produces an efficiency close to 1). Now, if f is sufficiently smooth, we have for any $\xi_i \in [t_{i-1}, t_i]$

$$\Big| \frac{(f(t_i) - f(t_{i-1}))^2}{t_i - t_{i-1}} - [\dot{f}(\xi_i)]^2 (t_i - t_{i-1}) \Big| \ \leq \ G,$$

for all $i = 2, \ldots, n$, where

$$G := 2 \max_{\xi \in [a,b]} |f'(\xi)| \max_{\xi \in [a,b]} |f''(\xi)| \cdot \max_{i=2,\ldots,n} |t_i - t_{i-1}|^2.$$

This gives

$$0 \leq A(t_1, \ldots, t_n) := \int_a^b \dot{f}^2(t) dt - \sum_{i=2}^{n} \frac{(f(t_i) - f(t_{i-1}))^2}{t_i - t_{i-1}} \leq (n-1)G.$$

As the function Φ has the representation

$$\Phi(t_1, \ldots, t_n) = \frac{A(t_1, \ldots, t_n)}{\int_a^b \dot{f}^2(s) ds - A(t_1, \ldots, t_n)},$$

it follows that (note that the expression on the right-hand side is increasing with $A(t_1, \ldots, t_n)$)

$$\Phi(t_1, \ldots, t_n) \leq \frac{(n-1) \cdot \max_{i=2,\ldots,n} |t_i - t_{i-1}|^2}{H(f) + (n-1) \cdot \max\limits_{i=2,\ldots,n} |t_i - t_{i-1}|^2}, \tag{3.8}$$

where

$$H(f) = \frac{\int_a^b \dot{f}^2(s) ds}{2 \max\limits_{\xi \in [a,b]} |\dot{f}(\xi)| \max\limits_{\xi \in [a,b]} |\ddot{f}(s)|}.$$

This shows that for most models a substantial improvement of the approximation by the choice of $t_2, \ldots, t_n$ can only be achieved if the sample size is small. For moderate or large sample sizes one could use the points $u_i = a + \frac{i-1}{n-1}(b-a)$, which gives already the estimate

$$\Phi(u_1, \dots, u_n) \le \frac{1}{1+(n-1)H(f)} = O\big(\frac{1}{n}\big).$$

Note that we consider worst case scenarios to obtain these estimates. Consequently, in many cases the design points can be chosen in an equidistant way provided that the weights of the estimator $\hat{\theta}_n$ are already chosen in an optimal way.

EXAMPLE 3.1. Consider the quadratic regression model $Y_t = \theta t^2 + \varepsilon_t$, where $t \in [a,b]$. Then $f(t) = t^2$, $\dot{f}(t) = 2t$, and the function Φ in (3.7) reduces to

$$\Phi(t_1, \dots, t_n) = \frac{4(b^3-a^3)}{3}\Big\{\sum_{i=2}^{n}(t_i+t_{i-1})^2(t_i-t_{i-1})\Big\}^{-1} - 1.$$

It follows by a straightforward computation that the optimal points are given by

$$t_i^* = a + \frac{i-1}{n-1}(b-a)\,; \qquad i = 1, \dots, n, \tag{3.9}$$

while the corresponding minimal value is

$$\Phi(t_1^*, \dots, t_n^*) = \frac{(a-b)^3}{4(n-1)^2(a^3-b^3)-(a-b)^3} \qquad (n \ge 2).$$

Note that this term is of order $O(\frac{1}{n^2})$. Remark 3.1 gives the bound

$$\Phi(t_1^*, \dots, t_n^*) \le \frac{1}{1+\frac{b^3-a^3}{2b}(n-1)} = O\big(\frac{1}{n}\big)\ ,$$

which shows that (3.8) is not necessarily sharp. For the efficiency we obtain

$$\mathrm{eff}(t_1^*, \dots, t_n^*) = 1 - \frac{4(a-b)^3(a^3-b^3)}{3a^3(a-b)^3+4(n-1)^2(a^3-b^3)(a-b)^3},$$

which is of order $1 - O(\frac{1}{n^2})$. On the other hand, if $f(t) = t^3$ the function Φ is given by

$$\begin{aligned}\Phi(t_1, \dots, t_n) &= \frac{9}{5}(b^5-a^5)\Big\{\sum_{i=2}^{n}(t_i-t_{i-1})(t_i^2+t_it_{i-1}+t_{i-1}^2)^2\Big\}^{-1} - 1 \\ &= \frac{(a-b)^2[5(n-1)^2(a^3-b^3)-(a-b)^3]}{9(n-1)^4(a^5-b^5)-(a-b)^2[5(n-1)^2(a^3-b^3)-(a-b)^3]},\end{aligned}$$

and optimal points have to be found numerically. However, we can evaluate the efficiency of the uniform design in (3.9), which is given by

$$\mathrm{eff}(t_1^*,\dots,t_n^*) = 1 - \frac{9(b^5-a^5)(a-b)^2[5(n-1)^2(a^3-b^3)-(a-b)^3]}{9(9b^5-4a^5)(a^5-b^5)(n-1)^4+5a^5(a-b)^2[5(n-1)^2(a^3-b^3)-(a-b)^3]},$$

$(n \geq 2)$ and also it is of order $1-O(\frac{1}{n^2})$. Thus, although the uniform design is not optimal, its efficiency (with respect to the continuous case) is extremely high.

3.2. *Multi-parameter models.* In this section we derive corresponding results for the multi-parameter case. If $m \geq 1$ we propose a linear estimator with matrix weights as an analogue of (3.1), that is

$$\begin{aligned}\hat{\theta}_n &= C^{-1}\Big\{\sum_{i=2}^n \Omega_i \dot{f}(t_{i-1})(Y_{t_i}-Y_{t_{i-1}}) + \frac{f(a)}{a}Y_a\Big\} \\ &= C^{-1}\Big\{\sum_{i=2}^n \mu_i(Y_{t_i}-Y_{t_{i-1}}) + \frac{f(a)}{a}Y_a\Big\},\end{aligned} \tag{3.10}$$

where C^{-1} is given in (2.7), $\Omega_2,\dots,\Omega_n$ are $m\times m$ matrices and $\mu_2 = \Omega_2\dot{f}(t_i),\dots,\mu_n = \Omega_n\dot{f}(t_{n-1})$ are m-dimensional vectors, which have to be chosen in an optimal way. For this purpose we derive a representation of the mean squared error between the best linear estimate in the continuous time model and its discrete approximation in the multi-parameter case first. The proof can be found in Appendix A.3.

LEMMA 3.3. *Consider the continuous time model* (2.3). *If the assumptions of Theorem 2.1 are satisfied, then*

$$\begin{aligned}\mathbb{E}_\theta[(\hat{\theta}_{\mathrm{BLUE}}-\hat{\theta}_n)(\hat{\theta}_{\mathrm{BLUE}}-\hat{\theta}_n)^T] &= C^{-1}\Big\{\sum_{i=2}^n\int_{t_{i-1}}^{t_i}\big[\dot{f}(s)-\mu_i\big]\big[\dot{f}(s)-\mu_i\big]^T ds \\ &+ \sum_{i=2}^n\int_{t_{i-1}}^{t_i}\big[\dot{f}(s)-\mu_i\big]\dot{f}^T(s)\,ds\,\theta\,\theta^T\sum_{j=2}^n\int_{t_{j-1}}^{t_j}\dot{f}(s)\big[\dot{f}(s)-\mu_j\big]^T ds\Big\}C^{-1}.\end{aligned} \tag{3.11}$$

In the following we choose optimal vectors (or equivalently matrices Ω_i) $\mu_i = \Omega_i\dot{f}(t_{i-1})$ and design points t_i, such that the linear estimate (3.10) is unbiased and the mean squared error matrix in (3.11) "becomes small". An alternative criterion is to replace $\mathbb{E}_\theta[(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)^T]$ by the mean squared error

$$\mathbb{E}_\theta[(\hat{\theta}_n - \theta)(\hat{\theta}_n - \theta)^T]$$

between the estimate $\hat{\theta}_n$ defined in (3.10) and the "true" vector of parameters. The following result shows that both optimization problems will yield the same solution in the class of all unbiased estimators. The proof can be found in Appendix A.4.

THEOREM 3.2. *The estimator $\hat{\theta}_n$ defined in (3.10) is unbiased if and only if the identity*

$$\int_a^b \dot{f}(s)\dot{f}^T(s)\,ds = \sum_{i=2}^n \mu_i \int_{t_{i-1}}^{t_i} \dot{f}^T(s)\,ds = \sum_{i=2}^n \mu_i (f(t_i) - f(t_{i-1}))^T, \tag{3.12}$$

is satisfied. Moreover, for any linear unbiased estimator of the form $\tilde{\theta}_n = \int_a^b g(s) dY_s$ we have

$$\mathbb{E}_\theta[(\tilde{\theta}_n - \theta)(\tilde{\theta}_n - \theta)^T] = \mathbb{E}_\theta[(\tilde{\theta}_n - \hat{\theta}_{\mathrm{BLUE}})(\tilde{\theta}_n - \hat{\theta}_{\mathrm{BLUE}})^T] + C^{-1}.$$

In order to describe a solution in terms of optimal "weights" μ_i^* and design points t_i^* we recall that the condition of unbiasedness of the estimate $\hat{\theta}_n$ in (3.10) is given by (3.12) and introduce the notation

$$\begin{aligned} \beta^{(i)} &= [f(t_i) - f(t_{i-1})]/\sqrt{t_i - t_{i-1}}, \\ \gamma^{(i)} &= \mu_i \sqrt{t_i - t_{i-1}}. \end{aligned} \tag{3.13}$$

It follows from Lemma 3.3 that for an unbiased estimate $\hat{\theta}_n$ the mean squared error has the representation

$$\mathbb{E}_\theta\big[(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)^T(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)\big] = -C^{-1}MC^{-1} + \sum_{i=2}^n C^{-1}\gamma^{(i)}\gamma^{(i)^T}C^{-1}, \tag{3.14}$$

which has to be "minimized" subject to the constraint

$$M = (m_{\ell,k})_{\ell,k}^m = \int_a^b \dot{f}(s)\dot{f}^T(s)ds = \sum_{i=2}^n \gamma^{(i)}\beta^{(i)^T}. \tag{3.15}$$

The following result shows that a minimization with respect to the weights μ_i (or equivalently γ_i) can actually be carried out with respect to the Loewner ordering.

THEOREM 3.3. *Assume that the assumptions of Theorem 2.1 are satisfied and that the matrix*

$$B = \sum_{i=2}^{n} \frac{[f(t_i) - f(t_{i-1})][f(t_i) - f(t_{i-1})]^T}{t_i - t_{i-1}},$$

is non-singular. Let $\mu_2^*, \dots, \mu_n^*$ denote $m \times 1$ vectors satisfying the equations

$$\mu_i^* = MB^{-1} \frac{f(t_i) - f(t_{i-1})}{t_i - t_{i-1}} \qquad i = 2, \dots, n, \tag{3.16}$$

then $\mu_2^*, \dots, \mu_n^*$ are optimal (vector) weights minimizing the mean squared error $\mathbb{E}_\theta[(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)^T]$ with respect to the Loewner ordering among all unbiased estimators of the form (3.10).

PROOF. Let A denote a positive definite $m \times m$ matrix and consider the problem of minimizing the linear criterion

$$\mathrm{tr}\left\{A \ \mathbb{E}_\theta[(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)^T]\right\},$$

subject to the constraint (3.15). Observing (3.14) this yields the Lagrange function

$$-\mathrm{tr}\{AC^{-1}MC^{-1}\} + \sum_{i=2}^{n} (C^{-1}\gamma^{(i)})^T A (C^{-1}\gamma^{(i)}) - \sum_{k,\ell=1}^{m} \lambda_{k,\ell} \Big(m_{k,\ell} - \sum_{i=2}^{n} \gamma_k^{(i)} \beta_\ell^{(i)} \Big),$$

where $C = (c_{k,\ell})_{k,\ell=1}^m$, $\gamma^{(i)} = (\gamma_1^{(i)}, \dots, \gamma_m^{(i)})^T$, $\beta^{(i)} = (\beta_1^{(i)}, \dots, \beta_m^{(i)})^T$ and $\Lambda = (\lambda_{k,\ell})_{k,\ell=1}^m$ is a matrix of Lagrange multipliers. This function is obviously convex with respect to $\gamma^{(2)}, \dots, \gamma^{(n)}$. Therefore, taking derivatives with respect to $\gamma_j^{(i)}$ $(j = 1, \dots, k)$ yields as necessary and sufficient for the extremum

$$\sum_{p=1}^{m} c^{p,j} \sum_{\ell=1}^{m} a_{p,\ell} \sum_{k=1}^{m} c^{\ell,k} \gamma_k^{(i)} + \sum_{p=1}^{m} \sum_{k=1}^{m} c^{p,k} \gamma_k^{(i)} \sum_{\ell=1}^{m} a_{p,\ell} c^{\ell,j} + \sum_{\ell=1}^{m} \lambda_{j,\ell} \beta_\ell^{(i)} = 0,$$

where $A = (a_{\ell,k})_{\ell,k=1}^m$ and $C^{-1} = (c^{\ell,k})_{\ell,k=1}^m$ is the inverse of the matrix C defined in (2.6). Rewriting this system of linear equations in matrix form gives

$$C^{-1}AC^{-1}\gamma^{(i)} + C^{-1}A^TC^{-1}\gamma^{(i)} + \Lambda\beta^{(i)} = 0 \qquad i = 2, \dots, n,$$

or equivalently

$$C^{-1}(A + A^T)C^{-1}\gamma^{(i)} = -\Lambda\beta^{(i)} \qquad i = 2, \dots, n.$$

Substituting this last expression in (3.15) and using the non-singularity of the matrices C and B yields for the matrix of Lagrangian multipliers

$$\Lambda = -C^{-1}(A + A^T)C^{-1}MB^{-1},$$

which finally gives

$$\gamma^{(i)} = MB^{-1}\beta^{(i)} \qquad i = 2, \dots, n.$$

Observing the notations in (3.13) shows that the optimal vector weights are given by (3.16). Thus the optimal weights in (3.16) do not depend on the matrix A and provide the solution for all linear optimality criteria. Consequently, using the matrices $A = vv^T + \varepsilon I_m$ with $v \in \mathbb{R}^m$, and considering the limit as $\varepsilon \to 0$, shows that the weights defined in (3.16) minimize $\mathbb{E}_\theta[(\hat{\theta}_{\text{BLUE}} - \hat{\theta}_n)(\hat{\theta}_{\text{BLUE}} - \hat{\theta}_n)^T]$ with respect to the Loewner ordering. $\square$

REMARK 3.2. If the matrix B in Theorem 3.3 is singular, the optimal vectors are not uniquely determined and we propose to replace the inverse B by its Moore-Penrose inverse.

Note that for fixed design points $t_1, \dots, t_n$ Theorem 3.3 yields universally optimal weights $\mu_2^*, \dots, \mu_n^*$ (with respect to the Loewner ordering) for estimators of the form (3.10) satisfying (3.12). On the other hand, a further optimization with respect to the Loewner ordering with respect to the choice of the points $t_1, \dots, t_n$ is not possible, and we have to apply a real valued optimality criterion for this purpose. More precisely, let $\hat{\theta}_n^*$ denote the estimator of the form (3.10) with optimal weights $\gamma^{*(i)} = \mu_i^* \sqrt{t_i - t_{i-1}}$ given by (3.16), then we choose $t_1, \dots, t_n$, such that

$$\operatorname{tr}\big(\mathbb{E}_\theta\big[(\hat{\theta}_{\text{BLUE}} - \hat{\theta}_n^*)^T(\hat{\theta}_{\text{BLUE}} - \hat{\theta}_n^*)\big]\big) = \operatorname{tr}\Big\{ -C^{-1}MC^{-1} + \sum_{i=2}^n C^{-1}\gamma^{*(i)}\gamma^{*(i)^T}C^{-1}\Big\}$$

$$= \operatorname{tr}\Big\{-C^{-1}MC^{-1} + C^{-1}M\Big(\sum_{i=2}^n \frac{(f(t_i) - f(t_{i-1})(f(t_i) - f(t_{i-1}))^T}{t_i - t_{i-1}}\Big)^{-1}MC^{-1}\Big\}$$

is minimal. The performance of this method will be illustrated in the following section.

4. Some numerical examples. In this section we illustrate our new methodology using several model and covariance kernel examples. Note that (under smoothness assumptions) our approach allows us to calculate a lower bound for the trace (or any other monotone functional) of the variance of any (unbiased) linear estimator for the parameter vector θ in model (2.1) [see Remark 2.1(c)]. Therefore we evaluate the quality of an estimator (with corresponding design), say $\hat{\theta}$, by the efficiency

$$\mathrm{eff}(\hat{\theta}) = \frac{\mathrm{tr}\{\mathrm{Var}_\theta(\hat{\theta}_{BLUE})\}}{\mathrm{tr}\{\mathrm{Var}_\theta(\hat{\theta})\}} = \frac{\mathrm{tr}(C^{-1})}{\mathrm{tr}\{\mathrm{Var}_\theta(\hat{\theta})\}}.$$

Throughout this section the estimator defined by (3.2) and Lemma 3.2 in the case of $m = 1$ and by (3.10) and Theorem 3.3 for $m > 1$, will be denoted by $\hat{\theta}_n^*$. As before the univariate and multivariate cases are studied separately.

4.1. *One-parameter models.* Consider model (2.1) with $m = 1$ and $n = 5$ observations in the interval $[a, b] = [1, 2]$, where the regression function is given by $f(t) = t^2$, $t^2 - 0.5$ and t^4 with kernel $k(s, t) = s \wedge t$. The discussion in Example 3.1 indicates that equally spaced design points provide already an efficient allocation for the new estimator $\hat{\theta}_n^*$ provided that the weights are chosen in an optimal way. Consequently, we compare the estimator $\hat{\theta}_{\mathrm{DPZ},n}$ (with a corresponding optimal design) proposed in Section 2.5 of Dette, Pepelyshev and Zhigljavsky (2016) with the BLUE for model (2.1), that is, the weighted least squares estimator, and also with the estimator defined by (3.2) and Lemma 3.2 both based on a uniform design. The latter two estimators are denoted by $\hat{\theta}_{\mathrm{WLSE}}^{\mathrm{uni}}$ and $\hat{\theta}_n^{*\mathrm{uni}}$, respectively, and we consider a uniform design with $n = 5$ equally spaced points. The corresponding efficiencies are displayed in Table 1.

TABLE 1
Efficiencies (in percent) of various estimators in the univariate linear regression model for $n = 5$ observations on the interval $[1, 2]$. $\hat{\theta}_{\mathrm{WLSE}}^{\mathrm{uni}}$ is the BLUE for model (2.1) based on a uniform design, $\hat{\theta}_n^{*\mathrm{uni}}$ is the estimator defined by (3.2) and Lemma 3.2 based on a uniform design and $\hat{\theta}_{\mathrm{DPZ},n}$ (with a corresponding design) is the estimator proposed in Dette, Pepelyshev and Zhigljavsky (2016).

$f(t)$	t^2	$t^2 - 0.5$	t^4
$\hat{\theta}_{\mathrm{BLUE},n}^{\mathrm{uni}}$	99.798	99.783	98.416
$\hat{\theta}_n^{*\mathrm{uni}}$	99.798	99.783	98.416
$\hat{\theta}_{\mathrm{DPZ},n}$	99.582	99.346	92.662

We observe that both $\hat{\theta}_{\mathrm{WLSE}}^{\mathrm{uni}}$ and $\hat{\theta}_n^{*\mathrm{uni}}$ have very good efficiencies and therefore we did not determine the optimal allocations for the two estimators. A comparison between both estimators shows that $\hat{\theta}_{\mathrm{WLSE}}^{\mathrm{uni}}$ and $\hat{\theta}_n^{*\mathrm{uni}}$ are practically not distinguishable. In all the cases considered, the efficiencies do not differ in the first 5 decimals. For example, for the function $f(t) = t^2 - 0.5$ we have

$$\mathrm{eff}(\hat{\theta}^{\mathrm{uni}}_{\mathrm{WLSE}}) = 0.99782609\ ,\ \mathrm{eff}(\hat{\theta}^{*\mathrm{uni}}_{n}) = 0.99782596\ .$$

The investigation of other one-dimensional examples showed a similar picture and details are omitted for the sake of brevity. Therefore, the new estimator $\hat{\theta}^*_n$ with a uniform design is not only highly efficient (even for small values of n), but most importantly, it is very close to the best achievable, that is, the weighted least squares estimator with a uniform design. The comparison with the estimator $\hat{\theta}_{\mathrm{DPZ},n}$ proposed in Dette, Pepelyshev and Zhigljavsky (2016) shows that the new approach still provides an improvement of an estimator which has efficiencies already above 90%, with the difference of efficiencies being small for $f(t) = t^2, t^2 - 0.5$ and large for $f(t) = t^4$.

4.2. *Models with $m > 1$ parameters.* We now compare the various estimators in the multi-parameter case. In particular, we consider two regression models given by

$$Y_t = (t, t^2, t^3)^T \theta + \varepsilon_t, \quad t \in [a, b], \tag{4.1}$$

$$Y_t = (\sin t, \cos t, \sin 2t, \cos 2t)^T\, \theta + \varepsilon_t, \quad t \in [a, b]. \tag{4.2}$$

For each one of these models we study two cases of the covariance kernel of the error process in model (2.1), namely $K(t, t') = \min\{t, t'\}$ and $K(t, t') = \exp\{-\lambda|t - t'|\}$. The sample size is again $n = 5$ and the design space is the interval $[1, 2]$.

It turns out that for these models and the particularly small sample size the uniform design does not yield similar high efficiencies as in the case $m = 1$ discussed in the previous section. For this reason we also calculate the corresponding optimal designs for $\hat{\theta}_{\mathrm{WLSE}}$ and the estimator $\hat{\theta}^*_n$ proposed in this paper [see (3.10) and Theorem 3.3] using the Particle swarm optimization (PSO) algorithm [see for example Clerc (2006) or Wong et al. (2015) among others].

If the error process is a Brownian motion, the optimal design of $\hat{\theta}^*_n$ is obtained by applying the PSO algorithm on the trace of the mean squared error $\mathbb{E}_\theta[(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)^T]$ given in (3.14) (or equivalently on the trace of $\mathbb{E}_\theta[(\hat{\theta}_n - \theta)(\hat{\theta}_n - \theta)^T]$), using the optimal weights μ^*_i, $i = 2, \ldots, n$, given in Theorem 3.3. In the case of the exponential kernel $K(t, t') = \exp\{-\lambda|t - t'|\}$ we follow the same procedure as before but for the transformed continuous time model given in (2.9). The optimal design for the initial model with the exponential covariance kernel can then be obtained by the transformation $\tilde{t} = q(t)$ applied on each one of the optimal design points the algorithm will yield (see Remark 2.1(b)). Minimizing (using the PSO method) the trace of $\mathrm{Var}(\hat{\theta}_{\mathrm{WLSE}})$ given in (2.2) for the corresponding variance/covariance matrix $\Sigma = (K(t_i, t_j))_{i,j=1,\dots,n}$ of the error process gives the optimal design for $\hat{\theta}_{\mathrm{WLSE}}$.

For the model and covariance kernel examples under consideration, the optimal designs for the estimators $\hat{\theta}_{WLSE}$ and $\hat{\theta}_n^*$ are presented in Table 2. The corresponding designs for the estimator $\hat{\theta}_{\mathrm{DPZ}}$ are chosen as described in Dette, Pepelyshev and Zhigljavsky (2016).

TABLE 2
Optimal five-point designs in the interval $[1, 2]$ for the estimators $\hat{\theta}_{\mathrm{WLSE}}$ and $\hat{\theta}_n^*$ for models (4.1) and (4.2) with two different covariance kernels.

		Optimal designs	
Model	**Kernel**	$\hat{\boldsymbol{\theta}}_{\mathbf{WLSE},n}$	$\hat{\boldsymbol{\theta}}_n^*$
(4.1)	$t \wedge t'$	[1, 1.466, 1.680, 1.852, 2]	[1, 1.444, 1.668, 1.846, 2]
	$\exp\{-\lvert t-t'\rvert\}$	[1, 1.474, 1.683, 1.852, 2]	[1, 1.459, 1.674, 1.847, 2]
(4.2)	$t \wedge t'$	[1, 1.111, 1.243, 1.800, 2]	[1, 1.120, 1.264, 1.802,2]
	$\exp\{-\lvert t-t'\rvert\}$	[1, 1.113, 1.245, 1.800, 2]	[1, 1.120, 1.263, 1.801, 2]

We observe that regardless of the model and the covariance kernel, the optimal designs for the estimators $\hat{\theta}_{\mathrm{WLSE}}$ and $\hat{\theta}_n^*$ are very similar. Furthermore, for the specific examples, the choice of covariance kernel does not affect the optimal design since for a given estimator, the two kernels yield the same design (up to 2 d.p.) for both models. In particular, the optimal designs are always supported at both end-points of the design space. For model (4.1), although the uniform design is not optimal, the middle points of the optimal design are somewhat spread in the interval $(1, 2)$, whereas in the case of model (4.2), more points are allocated closer to the lower bound $t = 1$ of the design space.

Table 3 gives the efficiencies of the three estimators $\hat{\theta}_{\mathrm{WLSE}}$, $\hat{\theta}_n^*$ and $\hat{\theta}_{\mathrm{DPZ},n}$ for the optimal design of each estimator (upper part) and the uniform design (lower part) with $n = 5$ equally-spaced observations. For model (4.1) and any of the two covariance kernels, if the uniform design is used both $\hat{\theta}_{\mathrm{WLSE}}$ and $\hat{\theta}_n^*$ estimators are very efficient. The efficiencies of course increase when observations are taken according to the optimal instead of the uniform design but remain below 90% when the four-dimensional model (4.2) is considered.

We also observe that the estimator $\hat{\theta}_n^*$ proposed in this paper has sub-

TABLE 3
Efficiencies (in percent) of the estimators $\hat{\theta}_{\mathrm{WLSE}}$, $\hat{\theta}_n^*$ and $\hat{\theta}_{\mathrm{DPZ},n}$ for models (4.1) and (4.2) and for two different covariance kernels of the error process. The design is the uniform or the optimal design with five observations

			Efficiencies		
	Model	**Kernel**	$\hat{\boldsymbol{\theta}}_{\mathbf{WLSE}}$	$\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}^*$	$\hat{\boldsymbol{\theta}}_{\mathbf{DPZ},\boldsymbol{n}}$
optimal design	(4.1)	$t \wedge t'$	96.77	96.71	82.14
		$\exp\{-\lvert t-t'\rvert\}$	96.72	96.65	79.60
	(4.2)	$t \wedge t'$	83.98	83.40	70.91
		$\exp\{-\lvert t-t'\rvert\}$	83.47	82.95	71.57
uniform design	(4.1)	$t \wedge t'$	94.35	93.82	76.38
		$\exp\{-\lvert t-t'\rvert\}$	94.07	93.46	75.10
	(4.2)	$t \wedge t'$	73.13	73.12	70.91
		$\exp\{-\lvert t-t'\rvert\}$	72.56	72.46	71.57

stantially larger efficiencies than $\hat{\theta}_{\mathrm{DPZ},n}$ (always well below 90%) and thus the new approach provides a substantial improvement and is additionally much easier to implement for multi-parameter models than that introduced in Dette, Pepelyshev and Zhigljavsky (2016). Finally, the estimators $\hat{\theta}_{\mathrm{WLSE}}$ and $\hat{\theta}_n^*$ have similar efficiencies regardless of the underlying design. We therefore conclude that the alternative approach proposed in this paper provides estimators with corresponding optimal designs which are practically not distinguishable from the optimal estimator and corresponding design for finite sample.

TABLE 4
Efficiencies (in percent) of the estimators $\hat{\theta}_{\mathrm{WLSE}}$, $\hat{\theta}_n^*$ and $\hat{\theta}_{\mathrm{DPZ},n}$ for models (4.1) and (4.2) and for two different covariance kernels of the error process. The design is the uniform or the optimal design with ten observations

			Efficiencies		
	Model	**Kernel**	$\hat{\boldsymbol{\theta}}_{\mathbf{WLSE}}$	$\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}^*$	$\hat{\boldsymbol{\theta}}_{\mathbf{DPZ},\boldsymbol{n}}$
optimal design	(4.1)	$t \wedge t'$	99.39	99.32	95.38
		$\exp\{-\lvert t-t'\rvert\}$	99.36	99.35	94.28
	(4.2)	$t \wedge t'$	96.64	96.56	92.03
		$\exp\{-\lvert t-t'\rvert\}$	96.20	95.68	92.51
uniform design	(4.1)	$t \wedge t'$	98.89	98.87	94.42
		$\exp\{-\lvert t-t'\rvert\}$	98.83	98.81	93.83
	(4.2)	$t \wedge t'$	94.62	94.61	89.76
		$\exp\{-\lvert t-t'\rvert\}$	94.47	94.46	88.01

As pointed out by a referee, it is of interest to also consider larger values for the sample size. Note that for one-parameter models, all estimators have efficiencies well above 90% already for $n = 5$ and therefore, we only examine the multi-parameter case. Table 4 is the analogous of Table 3 for the case of sample size $n = 10$.

As expected, an increase in the sample size results in the efficiencies of all three estimators to be increased. Furthermore, as in the case of $n = 5$, the proposed estimator $\hat{\theta}_n^*$ has substantially larger efficiencies that $\hat{\theta}_{\mathrm{DPZ},n}$ and is practically not distinguishable from the BLUE for finite sample size, that is, $\hat{\theta}_{\mathrm{WLSE}}$. We also observe that even though the optimal design performs best, the uniform design produces efficiencies above 90% for all model and covariance kernel examples. However, the estimator $\hat{\theta}_n^*$ with equally spaced time-points is efficient only provided that the weights are chosen in an optimal way, that is, as in Theorem 3.3.

5. Degenerate models.

So far we have considered the continuous regression model (2.3) with a covariance kernel of the form (2.4) satisfying $u(a) \neq 0$. If $u(a) = 0$, then the variance of the observation at $t = a$ is 0 and all formulas of Section 2 and 3 degenerate in this case. The estimator $\hat{\theta}_{\mathrm{BLUE}}$ in the continuous time model and its discrete approximation (3.10) are not well defined and the results of previous sections cannot be applied. In this section, we indicate how the methodology can be extended to the case $u(a) = 0$. For the sake of brevity we only consider the continuous time model with a Brownian motion as error process, since the transformation (2.10) which reduces any model with the covariance kernel (2.4) to the case of Brownian motion can still be applied.

The main idea is to construct the BLUE $\hat{\theta}_{\mathrm{BLUE}}$ in the continuous time model (2.3) on the interval $[0, b]$ by a sequence of estimators $\hat{\theta}_{\mathrm{BLUE},a}$ for the same model on the interval $[a, b]$, where $a \to 0$. For this purpose we make the dependence of some quantities in the following discussion more explicit. For example, we write C_a for the matrix C defined in (2.5). The three different cases of degeneracy are discussed below.

5.1. *Models with no intercept, that is* $1 \notin \mathrm{span}\{f_1, \ldots, f_m\}$.

By Lemma A.1 in Section A.1, if $1 \notin \mathrm{span}\{f_1, \ldots, f_m\}$ then the matrix

$$M_a = \int_a^b \dot{f}(s) \dot{f}^T(s) ds, \tag{5.1}$$

is non-singular for all $a \in [0, b)$. In particular, M_0^{-1} exists. Additionally, in this case, for any $a > 0$ the inverse of the matrix

$$C_a = \int_a^b \dot{f}(t)\dot{f}^T(t)\,dt + \frac{f(a)f^T(a)}{a} = M_a + \frac{f(a)f^T(a)}{a},$$

can be expressed in the form

$$C_a^{-1} = M_a^{-1} - \frac{M_a^{-1}f(a)f^T(a)M_a^{-1}}{a + f^T(a)M_a^{-1}f(a)}\,. \tag{5.2}$$

We now discuss the cases $f(0) \neq 0$ and $f(0) = 0$ separately.

THEOREM 5.1. *Consider the continuous time linear regression model* (2.3) *on the interval* $[0, b]$ *with a continuously differentiable vector* f *of regression functions. If each component of* f *is of bounded variation,* $1 \notin \operatorname{span}\{f_1, \ldots, f_m\}$ *and* $f(0) \neq 0 \in \mathbb{R}^m$*, then the estimator*

$$\hat{\theta}_{\mathrm{BLUE}} = \underline{C} \int_0^b \dot{f}(t)\,dY_t + \frac{M_0^{-1}f(0)}{f^T(0)M_0^{-1}f(0)}Y_0\,, \tag{5.3}$$

is the best linear unbiased estimator, where

$$\underline{C} = \lim_{a\to 0} C_a^{-1} = M_0^{-1} - \frac{M_0^{-1}f(0)f^T(0)M_0^{-1}}{f^T(0)M_0^{-1}f(0)} = \mathrm{Var}(\hat{\theta}_{\mathrm{BLUE}})\,.$$

PROOF. For any $a > 0$ the BLUE $\hat{\theta}_{\mathrm{BLUE},a}$ in the continuous time model (2.3) on the interval $[a, b]$ is given by

$$\hat{\theta}_{\mathrm{BLUE},a} = C_a^{-1}\Big(\int_a^b \dot{f}(t)\,dY_t + \frac{f(a)}{a}Y_a\Big). \tag{5.4}$$

As $a \to 0$,

$$\lim_{a\to 0} C_a^{-1}\int_a^b \dot{f}(t)\,dY_t = \underline{C}\int_0^b \dot{f}(t)\,dY_t, \tag{5.5}$$

and

$$\begin{aligned}\lim_{a\to 0} C_a^{-1}\frac{f(a)}{a} &= \lim_{a\to 0}\Big(M_a^{-1}\frac{f(a)}{a} - \frac{M_a^{-1}f(a)f^T(a)M_a^{-1}f(a)}{a(a + f^T(a)M_a^{-1}f(a))}\Big)\\ &= \lim_{a\to 0}\frac{M_a^{-1}f(a)}{a + f^T(a)M_a^{-1}f(a))} = \frac{M_0^{-1}f(0)}{f^T(0)M_0^{-1}f(0)}.\end{aligned}$$

Hence the left-hand side of (5.3) is the limit of the estimators $\hat{\theta}_{\mathrm{BLUE},a}$ as $a \to 0$. The covariance matrix of this estimator is obtained by Ito's formula and the fact that $\varepsilon_0 = 0$, that is

$$\mathrm{Var}(\hat{\theta}_{\mathrm{BLUE}}) = \underline{C}\left[\int_0^b \dot{f}(t)\dot{f}^T(t)\,dt\right]\underline{C} = \underline{C}M_0\underline{C} = I - \frac{M_0^{-1}f(0)f^T(0)}{f^T(0)M_0^{-1}f(0)}\underline{C} = \underline{C}\,.$$

In order to prove that the derived estimator (5.3) is in fact BLUE we use Theorem 2.3 in Näther (1985a), which states that an unbiased estimator of the form $\hat{\theta} = \int_a^b Y_t dG(t)$ with covariance matrix $C = \mathrm{Var}(\hat{\theta})$ is BLUE in model (2.1) if the identity

$$\int_a^b K(s,t)dG(s) = Cf(t), \tag{5.6}$$

holds for all $t \in [a,b]$. Here G is a vector measure on the interval $[a,b]$. In the present case $a = 0$ and $K(s,t) = \min(s,t)$, and in order to prove that the estimator (5.3) is indeed BLUE we use the representation

$$\int_0^b \dot{f}(t)\,dY_t = \dot{f}(b)Y_b - \dot{f}(0)Y_0 - \int_0^b Y_t d\dot{f}(t),$$

for the stochastic integral $\int_0^b \dot{f}(t)\,dY_t$. This defines the vector measure dG in an obvious manner, that is, it has mass $\underline{C}\dot{f}(b)$ at the point b, the density $-\underline{C}\ddot{f}(t)$ for $t \in [0,b]$ and some mass at the point 0. The validity of (5.6) for $\hat{\theta}_{\mathrm{BLUE}}$ and $\underline{C}$ now follows from

$$\begin{aligned} &-\int_0^b \min(s,t)d\dot{f}(s) = -\int_0^t s d\dot{f}(s) - t\int_t^b d\dot{f}(s) \\ &= -[t\dot{f}(t) - f(t) + f(0)] - t[\dot{f}(b) - \dot{f}(t)] = -f(0) + f(t) - t\dot{f}(b), \end{aligned}$$

by noting that $\underline{C}f(0) = 0$ and that the weight at b cancels out. □

If $f(0) = 0 \in \mathbb{R}^m$, the observation at $t = 0$ necessarily gives $Y_0 = 0$ and provides no further information about the parameter vector θ. We obtain the following result.

THEOREM 5.2. *Consider the continuous time linear regression model (2.3) on the interval $[0,b]$ with a continuously differentiable vector f of regression functions. If each component of f is of bounded variation, $1 \notin \mathrm{span}\{f_1,\dots,f_m\}$ and $f(0) = 0 \in \mathbb{R}^m$, then*

$$\hat{\theta}_{\mathrm{BLUE}} = M_0^{-1}\int_0^b \dot{f}(t)\,dY_t\,, \tag{5.7}$$

and

$$\mathrm{Var}(\hat{\theta}_{\mathrm{BLUE}}) = M_0^{-1}.$$

PROOF. Since for any $p = 1, \dots, m$ the function $f_p(t)$ is continuously differentiable on $[0, b]$, the limit $\lim_{t\to 0} f_p(t)/t$ is necessarily finite, possibly 0. Using this and the fact that $f(0) = 0$, the representation (5.2) gives $\lim_{a\to 0} C_a^{-1} = M_0^{-1}$, and the limit of $\hat{\theta}_{\mathrm{BLUE},a}$ defined in (5.4) is obviously (5.7). The covariance matrix of this estimator is again obtained by an application of Ito's formula and its optimality follows by similar arguments as given in the proof of Theorem 5.1. □

5.2. *Models with an intercept, that is* $1 \in \mathrm{span}\{f_1, \dots, f_m\}$. Without loss of generality, we may assume $f_1(t) = 1$ for all $t \in [0, b]$ and rewrite the original regression model (2.3) as

$$Y_t = \theta_1 + \tilde{\theta}^T \tilde{f}(t) + \varepsilon_t, \qquad t \in [0, b],$$

where $\tilde{\theta} = (\theta_2, \dots, \theta_m)^T$ and $\tilde{f}(t) = (f_2(t), \dots, f_m(t))^T)$. Note that the observation at $t = 0$ is error-free and gives $Y_0 = \theta_1 + \tilde{\theta}^T \tilde{f}(0)$. By subtracting we obtain

$$Y_t - Y_0 = \tilde{\theta}^T(\tilde{f}(t) - \tilde{f}(0)) + \varepsilon_t. \tag{5.8}$$

Note that $1 \notin \mathrm{span}\{\tilde{f}_2(t) - \tilde{f}_2(0), \dots, \tilde{f}_m(t) - \tilde{f}_m(0)\}$ and $\tilde{f}(t) - \tilde{f}(0)$ is obviously 0 at $t = 0$. For computing the BLUE for $\tilde{\theta}$ and its covariance matrix in model (5.8) we can apply Theorem 5.2 and obtain

$$\tilde{\theta}_{\mathrm{BLUE}} = \tilde{M}_0^{-1} \int_0^b \dot{\tilde{f}}(t)\, d(Y_t)\,, \tag{5.9}$$

$$\mathrm{Var}(\tilde{\theta}_{\mathrm{BLUE}}) = \tilde{M}_0^{-1} = \left[\int_0^b \dot{\tilde{f}}(t)\dot{\tilde{f}}^T(t)dt\right]^{-1}. \tag{5.10}$$

Finally, the BLUE for θ_1 is given by $\hat{\theta}_1 = Y_0 - \tilde{\theta}^T_{\mathrm{BLUE}}\tilde{f}(0)$. Noting that Y_0 is a constant, we obtain $\mathrm{cov}(\hat{\theta}_1, \hat{\theta}_p) = -\tilde{f}^T(0)\tilde{M}_0^{-1}e_p \quad (p = 2, \dots, m)$, where e_p is the p-th coordinate vector. The variance of $\hat{\theta}_1$ is given by $\mathrm{Var}(\hat{\theta}_1) = \tilde{f}^T(0)\tilde{M}_0^{-1}\tilde{f}(0)$.

5.3. *Optimal designs.* In the previous sections we have derived the BLUE $\hat{\theta}_{\mathrm{BLUE}}$ in the continuous time model (2.3) for the three cases of degeneracy. The corresponding linear unbiased estimators with optimal designs in the original model (2.1) can be obtained using a discrete approximation of the stochastic integral in the representation of the corresponding BLUE and following similar arguments as those presented in Section 3.

For example, in the case of Theorem 5.2, that is, for models with no intercept and $f(0) = 0 \in \mathbb{R}^m$, we define the approximate linear estimator

$$(5.11) \qquad \hat{\theta}_n = M_0^{-1} \sum_{i=2}^{n} \Omega_i \dot{f}(t_{i-1})(Y_{t_i} - Y_{t_{i-1}}) = M_0^{-1} \sum_{i=2}^{n} \mu_i (Y_{t_i} - Y_{t_{i-1}}).$$

Then using the condition (equivalent to the property that $\hat{\theta}_n$ is unbiased)

$$\sum_{i=2}^{n} \mu_i (f(t_i) - f(t_{i-1}))^T = \sum_{i=2}^{n} \mu_i \int_{t_{i-1}}^{t_i} \dot{f}^T(s)\, ds = M_0 = \int_0^b \dot{f}(s) \dot{f}^T(s)\, ds,$$

and following along the same lines as in the proof of Theorem 3.3, we obtain the optimal (vector) weights $\mu_2^*, \dots, \mu_n^*$ satisfying the equations

$$\mu_i^* = M_0 B^{-1} \frac{f(t_i) - f(t_{i-1})}{t_i - t_{i-1}}, \quad i = 2, \dots, n,$$

which minimize the mean squared error $\mathbb{E}_\theta[(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)^T]$ with respect to the Loewner ordering among all unbiased estimators of the form (5.11).

The optimal weights for the two other cases of degeneracy can be derives in a similar manner. Finally, as before the optimal time-points $t_1^*, \dots, t_n^*$ are chosen numerically such that the $\mathrm{tr}\big(\mathbb{E}_\theta\big[(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n^*)^T (\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n^*)\big]\big)$ is minimized, where $\hat{\theta}_n^*$ denotes the linear unbiased estimator with optimal weights for the corresponding case of degenerate model.

APPENDIX A: APPENDIX: MORE TECHNICAL DETAILS

A.1. An auxiliary result.

LEMMA A.1. *Let $f(t) = (f_1(t), \dots, f_m(t))^T$ be a vector of continuously differentiable linearly independent functions on the interval $[a, b]$ with $0 \leq a < b$ and define $M = \int_a^b \dot{f}(s) \dot{f}^T(s) ds$.*

1. *The matrix M is non-singular if and only if $1 \notin \mathrm{span}\{f_1, \dots, f_m\}$.*
2. *If $1 \in \mathrm{span}\{f_1, \dots, f_m\}$ then $\mathrm{rank}(M) = m - 1$.*

PROOF.

1. Obviously the non-singularity of M implies that $1 \notin \text{span}\{f_1, \ldots, f_m\}$. To prove the converse we consider the equation

$$a_1 \dot{f}_1(t) + \ldots a_m \dot{f}_m(t) = 0, \quad \forall t \in [a, b], \tag{A.1}$$

for scalars $a_1, \ldots, a_m$. This equation is satisfied if and only if for some a_0 we have

$$a_0 + a_1 f_1(t) + \ldots a_m f_m(t) = 0, \quad \forall t \in [a, b]\,. \tag{A.2}$$

By the assumption, the functions $f_1, \ldots, f_m$ are linearly independent on the interval $[a, b]$ and $1 \notin \text{span}\{f_1, \ldots, f_m\}$, which implies that the $m+1$ functions $1, f_1, \ldots, f_m$ are also linearly independent on $[a, b]$. Consequently, the equation (A.2) has only the trivial solution $a_0 = a_1 = \ldots = a_m = 0$ which yields that the equation (A.1) has only trivial solution $a_1 = \ldots = a_m = 0$. Therefore, the functions $\dot{f}_1(t), \ldots, \dot{f}_m(t)$ are linearly independent on the interval $[a, b]$ and the non-singularity of M follows from basic results on Gramian matrices [see Akhiezer and Glazman (1993), p. 12].

2. To prove the second part assume now that $1 \in \text{span}\{f_1, \ldots, f_m\}$. Since $f_1, \ldots, f_m$ are linearly independent we may assume without loss of generality that $f_1(t)$ is constant for all $t \in [a, b]$. In this case, $\dot{f}_1 = 0$ and $1 \notin \text{span}\{f_2, \ldots, f_m\}$ and part (1) shows that the $(m-1) \times (m-1)$ submatrix of the matrix $(\int_a^b f_k(s) f_l(s) ds)_{k,l=2,\ldots,m}$ has full rank, which implies that $\text{rank}(M) = m - 1$.

□

A.2. Proof of Lemma 2.1. If $1 \notin \text{span}\{f_1, \ldots, f_m\}$ if follows from Lemma A.1 in Section A.1 that the matrix M is non-singular and hence positive definite, which implies $C > 0$. If $1 \in \text{span}\{f_1, \ldots, f_m\}$ we may assume without loss of generality that $f_1(t) \equiv 1$. As the functions $f_2, \ldots, f_m$ are linearly independent and $1 \notin \text{span}\{f_2, \ldots f_m\}$ it follows that

$$M = \int_a^b \dot{f}(t) \dot{f}^T(t) dt = \begin{pmatrix} 0 & 0 \\ 0 & \tilde{M} \end{pmatrix},$$

where (by Lemma A.1) the matrix $\tilde{M} = (\int_a^b \dot{f}_k(t) \dot{f}_l^T(t) dt)_{k,l=2}^m$ has rank $m-1$. Define $f(t) = (1, \tilde{f}(t)^T)$, where $\tilde{f}^T(t) = (f_2, \ldots, f_m)$ and assume that the matrix C is singular. Then there exists a vector $z = (z_1, \tilde{z}^T) \in \mathbb{R}^m \setminus \{0\}$ with $\tilde{z} \in \mathbb{R}^{m-1}$ such that

$$z^T C z = z^T M z + \frac{z^T f(a) f^T(a) z}{a} = \tilde{z}^T \tilde{M} \tilde{z} + (z^T f(a))^2 / a \;= 0.$$

As both terms in the sum are nonnegative we have $\tilde{z}^T\tilde{M}\tilde{z} = 0$ and $z^T f(a) = 0$. Since $\tilde{M}$ is a positive definite matrix we obtain $\tilde{z} = 0 \in \mathbb{R}^{m-1}$. The equation $z^T f(a) = 0$ then becomes $z_1 f_1(0) = 0$ implying $z_1 = 0$ and hence $z = 0 \in \mathbb{R}^m$. This yields a contradiction to the assumption that the matrix C is singular and proves Lemma 2.1.

A.3. Proof of Lemma 3.3. Define the random variables

$$X_i = \int_{t_{i-1}}^{t_i} [\dot{f}(s) - \mu_i]\, dY_s, \qquad i = 2, \dots, n.$$

From the definition of $\hat{\theta}_{\mathrm{BLUE}}$ and $\hat{\theta}_n$ in (2.6) and (3.10), respectively, we have

$$\mathbb{E}_\theta[(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)^T] = C^{-1}\mathbb{E}_\theta\Big[\sum_{i=2}^n X_i \sum_{j=2}^n X_j^T\Big] C^{-1}.$$

Observing the fact that the random variables $X_2, \dots, X_n$ are independent we obtain

$$\mathbb{E}_\theta\Big[\sum_{i=2}^n X_i \sum_{i=2}^n X_i^T\Big] = \sum_{i=2}^n \mathbb{E}_\theta\big[(X_i - \mathbb{E}_\theta[X_i])(X_i - \mathbb{E}_\theta[X_i])^T\big] + \sum_{i=2}^n \mathbb{E}_\theta[X_i] \sum_{j=2}^n \mathbb{E}_\theta[X_j^T].$$

Ito's isometry yields

$$\mathbb{E}_\theta[X_i] = \int_{t_{i-1}}^{t_i} [\dot{f}(s) - \mu_i]\dot{f}^T(s)\theta ds, \qquad i = 2, \dots, n,$$

and

$$\begin{aligned}
\mathbb{E}_\theta[(X_i - \mathbb{E}_\theta[X_i])(X_i - \mathbb{E}_\theta[X_i])^T] &= \mathbb{E}_\theta\Big[\int_{t_{i-1}}^{t_i} [\dot{f}(s) - \mu_i]\, d\varepsilon_s \int_{t_{i-1}}^{t_i} [\dot{f}(s) - \mu_i]^T\, d\varepsilon_s\Big] \\
&= \int_{t_{i-1}}^{t_i} [\dot{f}(s) - \mu_i][\dot{f}(s) - \mu_i]^T\, ds.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}_\theta\Big[\sum_{i=2}^n X_i \sum_{i=2}^n X_i^T\Big] &= \sum_{i=2}^n \int_{t_{i-1}}^{t_i} [\dot{f}(s) - \mu_i][\dot{f}(s) - \mu_i]^T\, ds \\
&\quad + \sum_{i=2}^n \int_{t_{i-1}}^{t_i} [\dot{f}(s) - \mu_i]\dot{f}^T(s)\theta\, ds \sum_{j=2}^n \int_{t_{j-1}}^{t_j} \theta^T \dot{f}(s)[\dot{f}(s) - \mu_j]^T\, ds,
\end{aligned}$$

which proves the assertion.

A.4. Proof of Theorem 3.2.. Standard calculations show that

$$\mathbb{E}_\theta[\hat{\theta}_n] = C^{-1}\Big[\sum_{i=2}^{n} \mu_i (f(t_i) - f(t_{i-1}))^T + \frac{f(a) f^T(a)}{a}\Big]\theta.$$

Observing the definition of the matrix C in (2.7) it follows that the estimator $\hat{\theta}_n$ defined in (3.1) is unbiased if and only if the identity (3.12) is satisfied. In order to prove the second part of Theorem 3.2 we use the decomposition

$$\mathbb{E}_\theta[(\tilde{\theta}_n - \theta)(\tilde{\theta}_n - \theta)^T] = E_1 + E_2 + E_2^T + E_3, \tag{A.3}$$

where the terms E_1, E_2 and E_3 are defined by

$$\begin{aligned}
E_1 &= \mathbb{E}_\theta[(\tilde{\theta}_n - \hat{\theta}_{\mathrm{BLUE}})(\tilde{\theta}_n - \hat{\theta}_{\mathrm{BLUE}})^T],\\
E_2 &= \mathbb{E}_\theta[(\tilde{\theta}_n - \hat{\theta}_{\mathrm{BLUE}})(\hat{\theta}_{\mathrm{BLUE}} - \theta)^T],\\
E_3 &= \mathbb{E}_\theta[(\hat{\theta}_{\mathrm{BLUE}} - \theta)(\hat{\theta}_{\mathrm{BLUE}} - \theta)^T].
\end{aligned}$$

By Theorem 2.1 we have

$$E_3 = C^{-1} = \left[\int_a^b \dot{f}(s)\dot{f}^T(s)\,ds + \frac{f(a)f^T(a)}{a}\right]^{-1}.$$

Using the definition of $\tilde{\theta}_n$ and $\hat{\theta}_{\mathrm{BLUE}}$ in (2.6), yields

$$\begin{aligned}
C(\tilde{\theta}_n - \hat{\theta}_{\mathrm{BLUE}}) &= C\int_a^b g(s)\,dY_s - \int_a^b \dot{f}(s)\,dY_s - \frac{f(a)}{a}Y_a\\
&= C\int_a^b g(s)\dot{f}^T(s)\theta\,ds + C\int_a^b g(s)\,d\varepsilon_s - \int_a^b \dot{f}(s)\dot{f}^T(s)\theta\,ds - \int_a^b \dot{f}(s)\,d\varepsilon_s\\
&\quad - \frac{f(a)f^T(a)}{a}\theta - \frac{f(a)}{a}\varepsilon_a\\
&= \int_a^b [Cg(s) - \dot{f}(s)]\,d\varepsilon_s - \frac{f(a)}{a}\varepsilon_a,
\end{aligned}$$

where the last identity follows from the fact that $\tilde{\theta}_n$ is unbiased, that is,

$$\int_a^b g(s)\dot{f}^T(s)ds = I. \tag{A.4}$$

On the other hand

$$\begin{aligned}
C(\hat{\theta}_{\mathrm{BLUE}} - \theta) &= \int_a^b \dot{f}(s)\,dY_s + \frac{f(a)}{a}Y_a - \int_a^b \dot{f}(s)\dot{f}^T(s)\,ds\theta - \frac{f(a)f^T(a)}{a}\theta\\
&= \int_a^b \dot{f}(s)\,d\varepsilon_s + \frac{f(a)}{a}\varepsilon_a.
\end{aligned}$$

Therefore we obtain for the term E_2 the representation

$$\begin{aligned}
E_2 &= C^{-1}\Big\{\mathbb{E}_\theta\Big[\Big(\int_a^b [Cg(s) - \dot f(s)]\,d\varepsilon_s - \frac{f(a)}{a}\varepsilon_a\Big)\Big(\int_a^b \dot f(s)\,d\varepsilon_s + \frac{f(a)}{a}\varepsilon_a\Big)^T\Big]\Big\}C^{-1} \\
&= C^{-1}\Big\{\mathbb{E}_\theta\Big[\int_a^b [Cg(s) - \dot f(s)]\,d\varepsilon_s \int_a^b \dot f^T(s)\,d\varepsilon_s\Big] - \mathbb{E}_\theta\Big[\frac{f(a)}{a}\varepsilon_a\varepsilon_a^T\frac{f^T(a)}{a}\Big]\Big\}C^{-1} \\
&= C^{-1}\Big[\int_a^b [Cg(s) - \dot f(s)]\dot f^T(s)\,ds - \frac{f(a)f^T(a)}{a}\Big]C^{-1} \\
&= C^{-1}\Big[C - \int_a^b \dot f(s)\dot f^T(s)\,ds - \frac{f(a)f^T(a)}{a}\Big]C^{-1} = 0,
\end{aligned}$$

where the last identity is again a consequence of (A.4). Hence it follows from (A.3)

$$\mathbb{E}_\theta[(\tilde\theta_n - \theta)(\tilde\theta_n - \theta)^T] = \mathbb{E}_\theta[(\tilde\theta_n - \hat\theta_{\mathrm{BLUE}})(\tilde\theta_n - \hat\theta_{\mathrm{BLUE}})^T] + C^{-1},$$

which proves the assertion of Theorem 3.2.

ACKNOWLEDGEMENTS

The authors would like to thank Kirsten Schorning for her constructive comments on an earlier version of this manuscript and Martina Stein who typed parts of this paper with considerable technical expertise. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Parts of this paper have been written while the authors were visiting the Isaac Newton Institute, Cambridge, and we would like to thank the institute for its hospitality. We would also like to thank two referees for their constructive comments on an earlier version of this paper.

REFERENCES

Akhiezer, N. I. and Glazman, I. M. (1993). *Theory of linear operators in Hilbert space.* Dover Publications, Inc., New York. MR1255973

Bickel, P. J. and Herzberg, A. M. (1979). Robustness of design against autocorrelation in time. I. Asymptotic theory, optimality for location and linear regression. *The Annals of Statistics* **7** 77–95. MR515685

Boltze, L. and Näther, W. (1982). On effective observation methods in regression models with correlated errors. *Math. Operationsforsch. Statist. Ser. Statist.* **13** 507–519. MR682028

Clerc, M. (2006). *Particle swarm optimization.* ISTE, London. MR2269598

Dette, H., Kunert, J. and Pepelyshev, A. (2008). Exact optimal designs for weighted least squares analysis with correlated errors. *Statistica Sinica* **18** 135–154. MR2384982

Dette, H., Pepelyshev, A. and Zhigljavsky, A. (2013). Optimal design for linear models with correlated observations. *The Annals of Statistics* **41** 143–176. MR3059413

DETTE, H., PEPELYSHEV, A. and ZHIGLJAVSKY, A. (2016). Optimal designs in regression with correlated errors. *The Annals of Statistics* **44** 113–152. MR3449764

DOOB, J. L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. *Annals of Mathematical Statistics* **20** 393–403. MR0030732

HARMAN, R. and ŠTULAJTER, F. (2010). Optimal prediction designs in finite discrete spectrum linear regression models. *Metrika* **72** 281–294. MR2725102

HARMAN, R. and ŠTULAJTER, F. (2011). Optimality of equidistant sampling designs for the Brownian motion with a quadratic drift. *Journal of Statistical Planning and Inference* **141** 2750–2758. MR2787742

IBRAGIMOV, I. A. and HASMINSKIĬ, R. Z. (1981). *Statistical estimation*. Springer-Verlag, New York-Berlin. MR620321

KISEĽÁK, J. and STEHLÍK, M. (2008). Equidistant and D-optimal designs for parameters of Ornstein-Uhlenbeck process. *Statistics & Probability Letters.* **78** 1388–1396. MR2453793

MEHR, C. B. and MCFADDEN, J. A. (1965). Certain properties of Gaussian processes and their first-passage times. *Journal of the Royal Statistical Society. Series B.* **27** 505–522. MR0199885

MÜLLER, W. G. and PÁZMAN, A. (2003). Measures for designs in experiments with correlated errors. *Biometrika* **90** 423–434. MR1986657

NÄTHER, W. (1985a). *Effective observation of random fields*. BSB B. G. Teubner Verlagsgesellschaft, Leipzig. MR863287

NÄTHER, W. (1985b). Exact design for regression models with correlated errors. *Statistics* **16** 479–484. MR803486

PÁZMAN, A. and MÜLLER, W. R. G. (2001). Optimal design of experiments subject to correlated errors. *Statistics & Probability Letters* **52** 29–34. MR1820047

PUKELSHEIM, F. (2006). *Optimal design of experiments*. SIAM, Philadelphia, PA. MR2224698

SACKS, J. and YLVISAKER, D. (1966). Designs for regression problems with correlated errors. *Annals of Mathematical Statistics* **37** 66–89. MR0192601

SACKS, J. and YLVISAKER, D. (1968). Designs for regression problems with correlated errors; many parameters. *Annals of Mathematical Statistics* **39** 49–69. MR0220424

WONG, W. K., CHEN, R. B., HUANG, C. C. and WANG, W. C. (2015). A Modified Particle Swarm Optimization Technique for Finding Optimal Designs for Mixture Models. *PLoS ONE* **10** e0124720.

ZHIGLJAVSKY, A., DETTE, H. and PEPELYSHEV, A. (2010). A new approach to optimal design for linear models with correlated observations. *Journal of the American Statistical Association.* **105** 1093–1103. MR2752605

FAKULTÄT FÜR MATHEMATIK
RUHR-UNIVERSITÄT BOCHUM
44780 BOCHUM
GERMANY
E-MAIL: holger.dette@ruhr-uni-bochum.de
maria.konstantinou@ruhr-uni-bochum.de

SCHOOL OF MATHEMATICS
CARDIFF UNIVERSITY
CARDIFF, CF24 4AG
UK
E-MAIL: zhigljavskyaa@cardiff.ac.uk