Submitted to the Annals of Statistics

THE BLUE IN CONTINUOUS-TIME REGRESSION MODELS WITH CORRELATED ERRORS

By Holger Dette*, Andrey Pepelyshev† and Anatoly Zhigljavsky†

Ruhr-Universität Bochum and Cardiff University*†

In this paper the problem of best linear unbiased estimation is investigated for continuous-time regression models. We prove several general statements concerning the explicit form of the best linear unbiased estimator (BLUE), in particular when the error process is a smooth process with one or several derivatives of the response process available for construction of the estimators. We derive the explicit form of the BLUE for many specific models including the cases of continuous autoregressive errors of order two and integrated error processes (such as integrated Brownian motion). The results are illustrated on many examples.

1. Introduction.

Consider a continuous-time linear regression model of the form

$$y(t) = \theta^T f(t) + \epsilon(t)\,, \quad t \in \mathcal{T} \subseteq [A, B], \tag{1.1}$$

where $\theta \in \mathbb{R}^m$ is a vector of unknown parameters, $f(t) = (f_1(t), \dots, f_m(t))^T$ is a vector of linearly independent functions on $\mathcal{T}$, and $\epsilon = \{\epsilon(t) | t \in [A, B]\}$ is a random error process with $\mathbb{E}[\epsilon(t)] = 0$ for all $t \in [A, B]$ and covariances $\mathbb{E}[\epsilon(t)\epsilon(s)] = K(t, s)$. We will assume that ϵ has continuous (in the mean-square sense) derivatives $\epsilon^{(i)}$ $(i = 0, 1, \dots, q)$ up to order q, where q is a non-negative integer. Finally, $\mathcal{T}$ is the set where the observations of $y(t)$ and perhaps derivatives of $y(t)$ are available. This often occurs in practice, in particular, in the geophysical determination of gravity anomalies and the satellite gradiometry (Freeden, 1999), computer experiments (Morris et al., 1993; Stein, 2012) and global optimization (Osborne et al., 2009). For a detailed discussion on different types of derivatives of random processes we refer to Yaglom (1987).

The main aim of this paper is studying the best linear unbiased estimator (BLUE) of the parameters θ in the general setting and in many specific instances. Understanding of the explicit form of the BLUE has profound

MSC 2010 subject classifications: Primary 62M10 Secondary 62M09

Keywords and phrases: linear regression, correlated observations, signed measures, optimal design, BLUE, AR processes, continuous autoregressive model

significance on general estimation theory and on asymptotically optimal design for (at least) three reasons. Firstly, the efficiency of the ordinary least squares estimator, the discrete BLUE and other unbiased estimators can be computed exactly. Secondly, as pointed out in a series of papers (Sacks and Ylvisaker, 1966, 1968, 1970), the explicit form of the BLUE is the key ingredient for constructing the (asymptotically) optimal exact designs in the regression model

$$y(t_i) = \theta^T f(t_i) + \epsilon(t_i)\,, \quad A \le t_1 < t_2 \ldots < t_{N-1} < t_N \le B\,, \tag{1.2}$$

with $\mathbb{E}[\epsilon(t_i)\epsilon(t_j)] = K(t_i, t_j)$. Thirdly, simple and very efficient estimators for the parameter θ in the regression model (1.2) can be derived from the continuous BLUE, like the extended signed least squares estimator investigated in Dette et al. (2013, 2016) and the estimators based on approximation of stochastic integrals proposed in Dette et al. (2017). In contrast to our previous work, which had its focus on the construction of optimal designs, this paper concentrates on the specific properties of BLUE in the continuous time model (1.1); more discussion can be found in Section 2.8.

There are many classical papers dealing with construction of the BLUE, mainly in the case of a non-differentiable error process; that is, in model (1.1) with $q = 0$. In this situation, it is well understood that solving specific instances of an equation of Wiener-Hopf type

$$\int_{\mathcal{T}} K(t,s)\zeta(dt) = f(s)\,, \;\; \forall s \in \mathcal{T}, \tag{1.3}$$

for an m-dimensional vector ζ of signed measures implies an explicit construction of the BLUE in the continuous-time model (1.1). This equation was first considered in a seminal paper of Grenander (1950) for the case of the location-scale model $y(t) = \theta + \epsilon(t)$, i.e. $m = 1$, $f_1(t) = 1$. For a general regression model with $m \ge 1$ regression functions (and $q = 0$), the BLUE was extensively discussed in Grenander (1954) and Rosenblatt (1956) who considered stationary processes in discrete time, where the spectral representation of the error process was heavily used for the construction of the estimators. In this and many other papers including Kholevo (1969) and Hannan (1975) the subject of the study was concentrated around the spectral representation of the estimators and hence the results in these references are only applicable to very specific models. A more direct investigation of the BLUE in the location scale model (with $q = 0$) can be found in Hájek (1956), where equation (1.3) for the BLUE was solved for a few simple kernels. The most influential paper on properties of continuous BLUE and its relation to the reproducing kernel Hilbert spaces (RKHS) is Parzen (1961). A relation between discrete and continuous BLUE has been further addressed in Anderson (1970). An excellent survey of classical results on the BLUE is given in the book of Näther (1985), Sect. 2.3 and Chapter 4 (for the location scale model). Formally, Theorem 2.3 of Näther (1985) includes the case when the derivatives of the process $y(t)$ are available ($q \geq 0$); this is made possible by the use of generalized functions which may contain derivatives of the Dirac delta-function. This theorem, however, provides only a sufficient condition for an estimator to be the BLUE. The examples, where the explicit form of the BLUE was known before the publication of the monograph by Näther (1985), are listed in Sect. 2.3 of his book. In most of these examples either a Markovian structure of the error process is assumed or the one-dimensional location scale model is studied. Section 2.6 of our paper updates this list and gives a short outline of previously known cases where the explicit form of the BLUE was known until now.

There was also an extensive study of the relation between solutions of the Wiener-Hopf equations and the BLUE through the RKHS theory, see Parzen (1961); Sacks and Ylvisaker (1966, 1968, 1970) for an early or Ritter (2000) for a more recent reference. If $q = 0$ then the main RKHS assumption is usually formulated as the existence of a solution, say ζ_0, of equation (1.3), where the measure ζ_0 is continuous and has no atoms, see Berlinet and Thomas-Agnan (2011) for the RKHS theory. As shown in the present paper, this almost never happens for the commonly used covariance kernels and regression functions (a single general exception from this observation is given in Proposition 2.3). The case when the covariance kernel K is imprecisely known is carefully considered in (Näther, 1985, Ch.10); see also Loh et al. (2000); Anderes (2010); Stein (2012) for some discussions concerning the problem of estimation of covariance kernels.

Note also that the numerical construction of the continuous BLUE is difficult even for $q = 0$ and $m = 1$, see e.g. Ramm and Charlot (1980) and a remark on p.80 in Sacks and Ylvisaker (1966). For $q > 0$, the problem of numerical construction of the BLUE is severely ill-posed and hence is extremely hard.

The main purpose of this paper is to provide further insights into the structure of the BLUE (and its covariance matrix) from the observations $\{Y(t)|t \in \mathcal{T}\}$ (and its q derivatives) in continuous-time regression models of the form (1.1), where the set $\mathcal{T} \subseteq [A, B]$ defines the region where the process is observed. By generalizing the celebrated Gauss-Markov theorem, we derive new characterizations for the BLUE. Our results require minimal assumptions regarding the regression function and the error process. Important examples, where the BLUE can be determined explicitly, include general integrated processes (in particular, integrated Brownian motion) and continuous autoregressive processes including the Matérn kernels with parameters 3/2 and 5/2.

The remaining part of this paper is organized as follows. In Section 2 we develop a consistent general theory of best linear unbiased estimation using signed matrix measures and derive several important characterizations and properties of the BLUE. In particular, in Theorem 2.1 we provide necessary and sufficient conditions for an estimator to be BLUE when $q \geq 0$; in Theorem 2.2 such conditions are derived for $q = 0$, $\mathcal{T} \subset \mathbb{R}^d$ with $d \geq 1$ and very general assumptions about the vector of regression functions $f(\cdot)$ and the covariance kernel $K(\cdot, \cdot)$.

Section 3 is devoted to models where the error process has one derivative. In particular, we derive an explicit form of the BLUE, see Theorems 3.1 and 3.2, and obtain the BLUE for specific types of smooth kernels. In Section 3.4 we consider regression models with a continuous-time autoregressive (AR) error process of order 2 (i.e. CAR(2)) in more detail. Moreover, in practice the corresponding discrete-time regression model (1.2) is used. Therefore, in an online supplement [see Dette et al. (2018)] we exemplarily demonstrate that the covariance matrix of the BLUE in this model can be obtained as a limit of the covariance matrices of the BLUE in the discrete regression model (1.2) with observations at equidistant points and a discrete AR(2) error process. In Section 4 we give some insight into the structure of the BLUE when the error process is more than once differentiable. Some numerical illustrations are given in Section 5, while technical proofs can be found in Section 6.

2. General linear estimators and the BLUE.

2.1. *Linear estimators and their properties.* Consider the regression model (1.1) with covariance kernel $K(t, s) = \mathbb{E}[\epsilon(t)\epsilon(s)]$. Suppose that we can observe the process $\{y(t) | t \in \mathcal{T}\}$ and, if $q > 0$, also its mean square derivatives $\{y^{(i)}(t) | t \in \mathcal{T}\}$ for $i = 1, \ldots, q$. The set $\mathcal{T}$ is a Borel subset of some interval $[A, B]$ with $-\infty \leq A < B \leq \infty$. This is possible when the kernel $K(t, s)$ is q times continuously differentiable on the square $[A, B] \times [A, B]$ and the vector-function $f(t) = (f_1(t), \ldots, f_m(t))^T$ is q times differentiable on the interval $[A, B]$ with derivatives $f^{(1)}, \ldots f^{(q)}$ ($f^{(0)} = f$). We will also assume throughout that the functions $f_1, \ldots, f_m$ are linearly independent on $\mathcal{T}$.

Let $Y(t) = \{(y^{(0)}(t), \ldots, y^{(q)}(t))^T\}$ be the observation vector containing the process $y(t) = y^{(0)}(t)$ and its q derivatives. Denote by $\mathbf{Y}_{\mathcal{T}} = \{Y(t) : \; t \in \mathcal{T}\}$ the set of all available observations. The general linear estimator of the parameter θ in the regression model (1.1) can be defined as

$$\hat{\theta}_G = \int_{\mathcal{T}} G(dt)Y(t) = \sum_{i=0}^{q} \int_{\mathcal{T}} y^{(i)}(t)G_i(dt), \tag{2.1}$$

where $G(dt) = (G_0(dt), \dots, G_q(dt))$ is a matrix of size $m \times (q+1)$. The columns of this matrix are signed vector-measures $G_0(dt), \dots, G_q(dt)$ defined on Borel subsets of $\mathcal{T}$ (all vector-measures in this paper are signed and have length m).

The following lemma shows a simple way of constructing unbiased estimators; this lemma will also be used for deriving the BLUE in many examples. The proof is given in Section 6.

LEMMA 2.1. *Let $\zeta_0, \dots, \zeta_q$ be some signed vector-measures defined on $\mathcal{T}$ such that the $m \times m$ matrix*

$$C = \sum_{i=0}^{q} \int_{\mathcal{T}} \zeta_i(dt) \left(f^{(i)}(t) \right)^T \tag{2.2}$$

is non-degenerate. Define $G = (G_0, \dots, G_q)$, where G_i are the signed vector-measures and $G_i(dt) = C^{-1}\zeta_i(dt)$ for $i = 0, \dots, q$. Then the estimator $\hat{\theta}_G$ is unbiased.

Note that the matrix C defined in (2.2) plays the role of an information matrix; this can also be seen from Corollary 2.1 below.

The covariance matrix of any unbiased estimator $\hat{\theta}_G$ of the form (2.1) is

$$\begin{aligned} \operatorname{Var}(\hat{\theta}_G) &= \int_{\mathcal{T}} \int_{\mathcal{T}} G(dt)\mathbf{K}(t,s)G^T(ds) \\ &= \sum_{i=0}^{q} \sum_{j=0}^{q} \int_{\mathcal{T}} \int_{\mathcal{T}} \frac{\partial^{i+j} K(t,s)}{\partial t^i \partial s^j} G_i(dt) G_j^T(ds) \,, \end{aligned} \tag{2.3}$$

where

$$\mathbf{K}(t,s) = \left(\frac{\partial^{i+j} K(t,s)}{\partial t^i \partial s^j} \right)_{i,j=0}^{q} = \left(\mathbb{E}[\epsilon^{(i)}(t)\epsilon^{(j)}(s)] \right)_{i,j=0}^{q}$$

is the matrix consisting of the derivatives of K.

2.2. *The BLUE.* The (continuous) BLUE is defined as follows. If there exists a set of signed vector-measures, say $G = (G_0, \dots, G_q)$, such that the estimator $\hat{\theta}_G = \int_{\mathcal{T}} G(dt)Y(t)$ is unbiased and $\mathrm{Var}(\hat{\theta}_H) \geq \mathrm{Var}(\hat{\theta}_G)$ in the sense of Loewner ordering, where $\hat{\theta}_H = \int_{\mathcal{T}} H(dt)Y(t)$ is any other linear unbiased estimator which uses the observations $\mathbf{Y}_{\mathcal{T}}$, then $\hat{\theta}_G$ is called the best linear unbiased estimator (BLUE) for the regression model (1.1) using the set of observations $\mathbf{Y}_{\mathcal{T}}$. The BLUE depends on the kernel K, the vector-function f, the set $\mathcal{T}$ and the number q of available derivatives of the process $\{y(t) | t \in \mathcal{T}\}$. The notation "continuous BLUE" highlights that estimation is performed for continuous observations.
The following theorem is a generalization of the celebrated Gauss-Markov theorem (which is usually formulated for the case when $q = 0$ and $\mathcal{T}$ is finite) and gives a necessary and sufficient condition for an estimator to be the BLUE. In this theorem and below we denote the partial derivatives of the kernel $K(t, s)$ with respect to the first component by

$$K^{(i)}(t,s) = \frac{\partial^i K(t,s)}{\partial t^i}\,.$$

The proof of the theorem can be found in Section 6.

THEOREM 2.1. *Consider the regression model (1.1), where the error process $\{\epsilon(t) | t \in [A, B]\}$ has a covariance kernel $K(\cdot, \cdot) \in C^q([A, B] \times [A, B])$ for some $q \geq 0$. Suppose that the process $\{y(t) | t \in [A, B]\}$ along with its q derivatives can be observed at all $t \in \mathcal{T} \subseteq [A, B]$. Assume also that all components of $f(\cdot)$ are q times differentiable.*
An unbiased estimator $\hat{\theta}_G = \int_{\mathcal{T}} G(dt)Y(t)$ is BLUE if and only if the equality

$$\sum_{i=0}^{q} \int_{\mathcal{T}} K^{(i)}(t,s) G_i(dt) = Df(s), \tag{2.4}$$

is fulfilled for all $s \in \mathcal{T}$, where D is some $m \times m$ matrix. In this case, $D = \mathrm{Var}(\hat{\theta}_G)$ with $\mathrm{Var}(\hat{\theta}_G)$ defined in (2.3).

Corollary 2.1 is weaker than Theorem 2.1, where the covariance matrix of the BLUE is not assumed to be non-degenerate, but will be very useful in further considerations.

COROLLARY 2.1. *Let the assumptions of Theorem 2.1 be satisfied and let $\zeta_0, \dots, \zeta_q$ be signed vector-measures defined on $\mathcal{T}$ such that the matrix C defined in (2.2) is non-degenerate. Define $G = (G_0, \dots, G_q)$, $G_i(dt) =$*

$C^{-1}\zeta_i(dt)$ for $i = 0, \ldots, q$. The estimator $\hat{\theta}_G = \int_{\mathcal{T}} G(dt)Y(t)$ is the BLUE if and only if

$$\sum_{i=0}^{q} \int_{\mathcal{T}} K^{(i)}(t,s)\zeta_i(dt) = f(s) \tag{2.5}$$

for all $s \in \mathcal{T}$. In this case, the covariance matrix of $\hat{\theta}_G$ is $\mathrm{Var}(\hat{\theta}_G) = C^{-1}$.

In the following sections we derive sufficient conditions for (2.4) and (2.5); see, for example, Sections 3.1, 3.3 and 4.2.

2.3. *Grenander's theorem and its generalizations.*
When $\mathcal{T} = [A, B]$, $q = 0$, $m = 1$ and the regression model (1.1) is the location-scale model $y(t) = \alpha + \varepsilon(t)$, Theorem 2.1 is known as Grenander's theorem [see Grenander (1950) and Section 4.3 in Näther (1985)]. In this special case Grenander's theorem has been generalised by Näther (1985) to the case when $\mathcal{T} \subset \mathbb{R}^d$ [see Theorem 4.3 in this reference]. For the case of one-dimensional processes, Theorem 2.1 generalizes Grenander's theorem to arbitrary m-parameter regression models of the form (1.1) and the case of arbitrary $q \geq 0$. Another generalization of Grenander's theorem is given below; it deals with a general m-parameter regression model (1.1) with a continuous error process (i.e. $q = 0$) and a d-dimensional set $\mathcal{T} \subset \mathbb{R}^d$; that is, the case where $y(t)$ is a random field.

THEOREM 2.2. *Consider the regression model $y(t) = \theta^T f(t) + \epsilon(t)$, where $t \in \mathcal{T} \subset \mathbb{R}^d$, the error process $\epsilon(t)$ has covariance kernel $K(\cdot,\cdot)$ and $f : \mathcal{T} \to \mathbb{R}^m$ is a vector of bounded integrable and linearly independent functions. Suppose that the process $y(t)$ can be observed at all $t \in \mathcal{T}$ and let G be a signed vector-measure on $\mathcal{T}$, such that the estimator $\hat{\theta}_G = \int_{\mathcal{T}} G(dt)Y(t)$ is unbiased. $\hat{\theta}_G$ is a BLUE if and only if the equality*

$$\int_{\mathcal{T}} K(t,s)G(dt) = Df(s)$$

holds for all $s \in \mathcal{T}$ for some $m \times m$ matrix D. In this case, $D = \mathrm{Var}(\hat{\theta}_G)$, where $\mathrm{Var}(\hat{\theta}_G)$ is the covariance matrix of the estimator $\hat{\theta}_G$ defined by (2.3).

The proof of this theorem is a simple extension of the proof of Theorem 2.1 with $q = 0$ to a general set $\mathcal{T} \subset \mathbb{R}^d$ and left to the reader.

2.4. *Properties of the BLUE.*

(P1) Let $\hat{\theta}_{G_1}$ and $\hat{\theta}_{G_2}$ be BLUEs for the same regression model (1.1) and the same q but for two different design sets $\mathcal{T}_1$ and $\mathcal{T}_2$ such that $\mathcal{T}_1 \subseteq \mathcal{T}_2$. Then $\mathrm{Var}(\hat{\theta}_{G_1}) \geq \mathrm{Var}(\hat{\theta}_{G_2})$.

(P2) Let $\hat{\theta}_{G_1}$ and $\hat{\theta}_{G_2}$ be BLUEs for the same regression model (1.1) and the same design set $\mathcal{T}$ but for two different values of q, say, q_1 and q_2, where $0 \leq q_1 \leq q_2$. Then $\mathrm{Var}(\hat{\theta}_{G_1}) \geq \mathrm{Var}(\hat{\theta}_{G_2})$.

(P3) Let $\hat{\theta}_G$ with $G = (G_0, \ldots, G_q)$ be a BLUE for the regression model (1.1), design space $\mathcal{T}$ and given $q \geq 0$. Define $g(t) = Lf(t)$, where L is a non-degenerate $m \times m$ matrix, and a signed vector-measure $H = (H_0, \ldots, H_q)$ with $H_i(dt) = L^{-1}G_i(dt)$ for $i = 0, \ldots, q$. Then $\hat{\theta}_H$ is a BLUE for the regression model $y(t) = \beta^T g(t) + \varepsilon(t)$ with the same $y(t)$, $\varepsilon(t)$, $\mathcal{T}$ and q. The covariance matrix of $\hat{\theta}_H$ is $L^{-1}\mathrm{Var}(\hat{\theta}_G)L^{-1^T}$.

(P4) If $\mathcal{T} = [A, B]$ and a BLUE $\hat{\theta}_G$ is defined by the matrix-measure G that has smooth enough continuous parts, then we can choose another representation $\hat{\theta}_H$ of the same BLUE, which is defined by the matrix-measure $H = (H_0, H_1, \ldots, H_q)$ with vector-measures $H_1, \ldots, H_q$ having no continuous parts.

(P5) Let $\zeta_0, \ldots, \zeta_q$ satisfy the equation (2.5) for all $s \in \mathcal{T}$, for some vector-function $f(\cdot)$, design set $\mathcal{T}$ and given $q \geq 0$. Define $C = C_f$ by (2.2). Let $g(\cdot)$ be some other q times differentiable vector-function on the interval $[A, B]$. Assume that the signed vector-measures $\eta_0, \ldots, \eta_q$ satisfy the equation

$$\sum_{i=0}^{q} \int_{\mathcal{T}} K^{(i)}(t,s)\eta_i(dt) = g(s)\,, \quad \forall s \in \mathcal{T}; \tag{2.6}$$

that is, the equation (2.5) for the vector-function $g(\cdot)$, the same design set $\mathcal{T}$ and the same q. Define $C_g = \sum_{i=0}^{q} \int_{\mathcal{T}} g^{(i)}(t)\eta_i^T(dt)$, which is the matrix (2.2) with η_i substituted for ζ_i and $g(\cdot)$ substituted for $f(\cdot)$. If the matrix $C = C_f + C_g$ is non-degenerate, then we define the set of signed vector-measures $G = (G_0, \ldots, G_q)$ by $G_i = C^{-1}(\zeta_i + \eta_i)$, $i = 0, \ldots, q$, yielding the estimator $\hat{\theta}_G$. This estimator is a BLUE for the regression model $y(t) = \theta^T[f(t) + g(t)] + \varepsilon(t)$, $t \in \mathcal{T}$.

Properties (P1)–(P3) are obvious. The property (P4) is a particular case of the discussion of Case (5) in Section 2.5. To prove (P5) we simply add the equations (2.5) and (2.6) and then use Corollary 2.1.

We believe that the properties (P4) and (P5) have never been noticed before and both these properties are very important for understanding best linear

unbiased estimators in the continuous-time regression model (1.1) and especially for constructing a BLUE for new models from the cases when a BLUE is known for simpler models.
As an example, set $g(t) = c$, where c is a constant vector of size m and let f be arbitrary for which we know the BLUE with $C = C_f$. Assume that we also know the BLUE for the location-scale model which gives us the associated matrix $C = C_g$ of rank 1 for the model $\theta^T g(t)$. Then, assuming that the matrix $C_f + C_g$ is non-degenerate we can use property (P5) to construct BLUE for $\theta^T(f(t) + c)$. In particular, if all functions in the vector f are not constant and C_f is non-degenerate then $C_f + C_g$ is non-degenerate. This observation constitutes an important part in the proof of Theorem 3.2, which allows obtaining the explicit form of the BLUE for integrated error processes from the explicit form of the BLUE for the corresponding non-integrated errors (which is an easier problem). In this particular application of property (P5), the vector-function g is used to correct the constant terms in functions $f_1, \dots, f_m$ as the latter ones are integrals $\int_a^t \psi_i(s)ds$ of some other functions ψ_i and hence contain undesirable constant terms.

2.5. *Existence and uniqueness of the BLUE.*
Let us classify different situations.
(1) If functions $f_1, \dots, f_m$ are linearly dependent on $\mathcal{T}$ then the BLUE does not exist as the unbiasedness condition cannot be satisfied. This is the reason why we assume that $f_1, \dots, f_m$ are linearly independent on $\mathcal{T}$.
(2) If $\mathcal{T}$ is a discrete set $\mathcal{T} = \{t_1, \dots, t_N\}$ and the kernel K is strictly positive definite then the BLUE exists for any $q \geq 0$. It is uniquely defined for $q = 0$ as the matrix $(K(t_i, t_j))_{i,j=1}^N$ is always non-degenerate. However, for $q > 0$ the BLUE may not be uniquely defined as the matrix of covariances and cross-covariances between observations and derivatives may not be non-degenerate; this is similar to the continuous case considered below in case (5).
(3) If all functions $f_1, \dots, f_m$ belong to the RKHS associated with K and additionally satisfy some extra smoothness conditions then the BLUE exits and can be found using results of Parzen (1961). A serious difficulty with this approach is the fact that for the majority of kernels there is no known expression for the scalar products in the RKHS. Also, the RKHS-based approach is not applicable when the covariance matrix of the BLUE is degenerate. There are many examples when the BLUE exists and can be found using Theorem 2.1 with functions $f_1, \dots, f_m$ which may not belong to the RKHS. As a simple example, set $\mathcal{T} = [0, 1]$, $q = m = 1$, $f(t) = t - t^2/2$, $K(t, s) = \min(t, s)^2(3 \max(t, s) - \min(t, s))/6$, see (3.5). Define $G_0(dt) = 0$ and $G_1(dt) = \delta_0(dt)$, where $\delta_a(dt)$ is the Dirac delta-measure concentrated

at the point a. Then equation (2.4) is satisfied with $D = 0$, and as the estimator $\hat{\theta}_G$ is obviously unbiased $G = (G_0, G_1)$ defines the BLUE .
(4) If $\mathcal{T} = [A, B]$ and the derivatives $y^{(j)}(t)$ are available for $j = 0, \ldots, p < q$ then the BLUE exists only for very specific functions f. Assume, for example, that $p = 0$ so that only values of $y(t)$ are available. Then the class of respective functions is $g(t) = \tilde{f}(t) + \sum_i c_i K(a_i, t)$, where $a_i \in [A, B]$, $c_i \in \mathbb{R}$ and $\tilde{f}$ is defined by (2.8). It seems to be in contradiction with the fact that discrete BLUE estimators always exist even when $p = 0$. This can be explained by the behaviour of these discrete BLUEs when the uniform N-point grid approximating a continuous $\mathcal{T} = [A, B]$ gets finer and finer: the discrete BLUE weights at the points close to A and B are trying to create approximations for all q derivatives of y at A and B and hence have the order of N^q (in absolute values). Therefore, (a) the sequence of discrete measures diverge, and (b) the covariance matrices of the discrete BLUEs converge very slowly; they do converge to the covariance matrix of the continuous BLUE which would use all q derivatives of $y(t)$. To increase efficiency of discrete BLUEs we would advice to always place $q - 1$ distinct design points very close to A and B, in addition to A and B themselves.
(5) Assume $\mathcal{T} = [A, B]$, $q > 0$ and the values of derivatives $y^{(j)}(t)$ for $j = 0, \ldots, q$ are available. In this case, if f is smooth enough then the BLUE is not uniquely defined. More generally, we will show that if $\mathcal{T} = [A, B]$ then, under additional smoothness conditions of f, for a given set of signed vector-measures $G = (G_0, G_1, \ldots, G_q)$ on $\mathcal{T}$ we can find another set of measures $H = (H_0, H_1, \ldots, H_q)$ such that the signed vector-measures $H_1, \ldots, H_q$ have no continuous parts but the expectations and covariance matrices of the estimators $\hat{\theta}_G$ and $\hat{\theta}_H$ coincide.
For this purpose, let $G_0, \ldots, G_q$ be some signed vector-measures and assume that for some $i \in \{1, \ldots, m\}$, the signed measure $G_i(dt)$ has the form $G_i(dt) = Q_i(dt) + \varphi_i(t)dt$, where $Q_i(dt)$ is a signed vector-measure and $\varphi_i \in C^i([A, B])$. Define the matrix $H = (H_0, \ldots, H_q)$, where the columns of H are the following signed vector-measures:

$$\begin{aligned}
H_0(dt) &= G_0(dt) + (-1)^i \left[{\varphi_i}^{(i)}(t)dt - \varphi_i^{(i-1)}(A)\delta_A(dt) + \varphi_i^{(i-1)}(B)\delta_B(dt) \right], \\
H_j(dt) &= G_j(dt) + (-1)^{i-j-1} \left[\varphi_i^{(i-j-1)}(A)\delta_A(dt) - \varphi_i^{(i-j-1)}(B)\delta_B(dt) \right]
\end{aligned}$$

for $j = 1, \ldots, i-1$; $H_i(dt) = Q_i(dt)$, $H_j(dt) = G_j(dt)$, for $j = i+1, \ldots, q$. The proof of the following result is given in Section 6.

LEMMA 2.2. *In the notation above, the expectations and covariance matrices of the estimators $\hat{\theta}_G = \int G(dt)Y(t)$ and $\hat{\theta}_H = \int H(dt)Y(t)$ coincide.*

Lemma 2.2 shows that the sets of measures $G = (G_0, \dots, G_q)$ and $H = (H_0, \dots, H_q)$ produce estimators $\hat{\theta}_G$ and $\hat{\theta}_H$ of the form (2.1) with the same covariance matrix. Therefore, we can restrict the search of linear unbiased estimators to estimators $\hat{\theta}_G$ such that the components $G_1, \dots, G_q$ of G have no continuous parts. To achieve this, by a repeated use of Lemma 2.2 we negate the absolutely continuous parts of measures G_i one-by-one, for $i = q, q-1, \dots, 1$. A family of different BLUE-measures is shown in Example 3.1.

2.6. *Examples of the BLUE for non-differentiable error processes.* For the sake of completeness we first consider the case when the errors in model (1.1) follow a Markov process, which is a very common class of correlation kernels and includes as a particular case the kernels of continuous autoregressive errors of order 1. In presenting these results we follow Näther (1985) and Dette et al. (2016).

PROPOSITION 2.1. *Consider the regression model (1.1) with f twice differentiable and covariance kernel $K(t,s) = u(t)v(s)$ for $t \le s$ and $K(t,s) = v(t)u(s)$ for $t > s$; here $u(\cdot)$ and $v(\cdot)$ are twice differentiable positive functions such that $q(t) = u(t)/v(t)$ is monotonically increasing. Define the signed vector-measure $\zeta(dt) = z_A\delta_A(dt) + z_B\delta_B(dt) + z(t)dt$ with*

$$\begin{aligned} z_A &= \frac{1}{v^2(A)q'(A)}\Big[\frac{f(A)u'(A)}{u(A)} - f'(A)\Big], \\ z(t) &= -\frac{1}{v(t)}\Big[\frac{h'(t)}{q'(t)}\Big]', \quad z_B = \frac{h'(B)}{v(B)q'(B)}, \end{aligned}$$

where ψ' denotes a derivative of a function ψ, the vector-function $h(\cdot)$ is defined by $h(t) = f(t)/v(t)$. Assume that the matrix $C = \int_{\mathcal{T}} f(t)\zeta^T(dt)$ is non-degenerate. Then the estimator $\hat{\theta}_G$ with $G(dt) = C^{-1}\zeta(dt)$ is a BLUE with covariance matrix C^{-1}.

In the following statement we provide an explicit expression for the BLUE for one special case of non-Markovian covariance kernel. The proof is given in Section 6.

PROPOSITION 2.2. *Consider the regression model (1.1) on the interval $\mathcal{T} = [A,B]$ with errors having the covariance function $K(t,s) = 1 + \lambda_1 t - \lambda_2 s$, where $t \le s$, $\lambda_1 \ge \lambda_2 > 0$ and $\lambda_2(B-A) \le 1$. Define the signed vector-*

measure $\zeta(dt) = z_A\delta_A(dt) + z_B\delta_B(dt) + z(t)dt$ *by*

$$z(t) = -\frac{f^{(2)}(t)}{\lambda_1+\lambda_2}, \quad z_A = \Big(-f^{(1)}(A) + \frac{\lambda_1^2 f(A) + \lambda_1\lambda_2 f(B)}{\lambda_1+\lambda_2+\lambda_1^2 A - \lambda_2^2 B)}\Big)/(\lambda_1+\lambda_2),$$
$$z_B = \Big(f^{(1)}(B) + \frac{\lambda_1\lambda_2 f(A) + \lambda_2^2 f(B)}{\lambda_1+\lambda_2+\lambda_1^2 A - \lambda_2^2 B)}\Big)/(\lambda_1+\lambda_2)$$

and suppose that the matrix $C = \int_{\mathcal{T}} f(t)\zeta^T(dt)$ *is non-degenerate. Then the estimator* $\hat{\theta}_G$ *with* $G(dt) = C^{-1}\zeta(dt)$ *is a BLUE with covariance matrix* C^{-1}*.*

If $\lambda_1 = \lambda_2$ and $[A,B] = [0,1]$ in Proposition 2.2 then we obtain the case

$$K(t,s) = \max(1-\lambda|t-s|, 0)\,. \tag{2.7}$$

Optimal designs for this covariance kernel (with $\lambda = 1$) have been considered in [Sect. 6.5 in Näther (1985)], Müller and Pázman (2003) and Fedorov and Müller (2007).

EXAMPLE 2.1. Consider the regression model (1.1) on the interval $\mathcal{T} = [0,1]$ with errors having the covariance kernel (2.7) with $\lambda \leq 1$. Define the signed vector-measure

$$\zeta(dt) = \big[-\tfrac{f^{(1)}(0)}{2\lambda} + f_\lambda\big]\delta_0(dt) + \big[\tfrac{f^{(1)}(1)}{2\lambda} + f_\lambda\big]\delta_1(dt) - \big[\tfrac{f^{(2)}(t)}{2\lambda}\big]dt\,,$$

where $f_\lambda = (f(0)+f(1))/(4-2\lambda)$. Assume that the matrix $C = \int_{\mathcal{T}} f(t)\zeta^T(dt)$ is non-degenerate. Then the estimator $\hat{\theta}_G$ with $G(dt) = C^{-1}\zeta(dt)$ is a BLUE; the covariance matrix of this estimator is given by C^{-1}.

Next we consider the case when the regression functions are linear combinations of eigenfunctions from Mercer's theorem. Note that a similar approach was used in Dette et al. (2013) for the construction of optimal designs for the signed least squares estimators. Let $\mathcal{T} = [A,B]$; consider the integral operator $T_K(h)(\cdot) = \int_A^B K(t,\cdot)h(t)dt$ on $L_2([A,B])$, which defines a symmetric, compact self-adjoint operator. In this case Mercer's Theorem [see e.g. Kanwal (1997)] shows that there exist a countable number of orthonormal eigenfunctions $\phi_1, \phi_2, \ldots$ with positive eigenvalues $\lambda_1, \lambda_2, \ldots$ of the integral operator T_K. The next statement follows directly from Corollary 2.1.

PROPOSITION 2.3. *Let* $\phi_1, \phi_2, \ldots$ *be the eigenfunctions of the integral operator* $T_K(\cdot)$ *and* $f(t) = \sum_{\ell=1}^\infty q_\ell\phi_\ell(t)$ *for some sequence* $\{q_\ell\}_{\ell\in\mathbb{N}}$ *in* $\mathbb{R}^m$*.*

Assume that the matrix $C = \sum_{\ell=1}^{\infty} \lambda_\ell^{-1} q_\ell q_\ell^T$ *is non-degenerate and the vector sum* $\sum_{\ell=1}^{\infty} \lambda_\ell^{-1} q_\ell \phi_\ell(t)$ *converges for all* t*. Then the estimator* $\hat{\theta}_G$ *with*

$$G(dt) = C^{-1} \sum_{\ell=1}^{\infty} \lambda_\ell^{-1} q_\ell \phi_\ell(t) dt$$

is a BLUE with covariance matrix C^{-1}*.*

Proposition 2.3 provides a way of constructing covariance kernels for which the measure defining the BLUE does not have any atoms. An example of such kernels is the following.

EXAMPLE 2.2. Consider the regression model (1.1) with $m = 1$, $f(t) \equiv 1$, $t \in \mathcal{T} = [-1, 1]$, and the covariance kernel $K(t, s) = 1 + \kappa p_{\alpha,\beta}(t) p_{\alpha,\beta}(s)$, where $\kappa > 0, \alpha, \beta > -1$ are some constants and $p_{\alpha,\beta}(t) = \frac{\alpha-\beta}{2} + (1 + \frac{\alpha+\beta}{2})t$ is the Jacobi polynomial of degree 1. Then the estimator $\hat{\theta}_G$ with $G(dt) = \text{const} \cdot (1-t)^\alpha (1+t)^\beta dt$ is a BLUE.

2.7. *BLUE for functions from the class* $\mathrm{SY}(K)$. Equation (1.3) is related to the work of Sacks and Ylvisaker (1966, 1968, 1970), who used a RKHS approach to construct asymptotically optimal designs for linear regression models with correlated observations. To be precise denote by $H(K)$ the RKHS of functions on $\mathcal{T}$ associated with the kernel K and by $\mathrm{SY}(K)$ the class of functions $h \in H(K)$ of the form $h(\cdot) = \int_{\mathcal{T}} K(s, \cdot)\phi(s) ds$ for some continuous function ϕ on $\mathcal{T}$. The functions from $\mathrm{SY}(K)$ are often referred to as the functions satisfying the Sacks-Ylvisaker conditions, see Ritter et al. (1995); Ritter (2000).

Assume that all components f_i of f belong to $\mathrm{SY}(K)$ so that $f_i(\cdot) = \int_{\mathcal{T}} K(s, \cdot)\phi_i(s) ds$ for some continuous functions $\phi_i(\cdot)$. Set $\phi = (\phi_1, \ldots, \phi_m)^T$. Corollary 2.1 then implies that if the matrix $C = \int_{\mathcal{T}} \phi(t) f^T(t) dt$ is non-degenerate then the estimator $\hat{\theta}_G = \int_{\mathcal{T}} G(dt) Y(t)$ is the BLUE; here $G = (G_0, \ldots, G_q)$ with $G_0(dt) = C^{-1}\phi(t)dt$ and $G_i = 0$, for $i = 1, \ldots, q$. This implies that if all components f_i of f belong to $\mathrm{SY}(K)$, then the BLUE measure for f can be chosen so that it has no atoms and no weights assigned to any derivatives of $y(t)$.

Assume now that $\mathcal{T} = [A, B]$, the vector-function f is smooth enough and all components f_i of f belong to $H(K)$ but not necessarily to $\mathrm{SY}(K)$. As shown in Section 2.5, we can choose vector-measures $\zeta_i(dt)$ $(i = 0, 1, \ldots, q)$ satisfying (2.5) so that there are no continuous parts in the measures $\zeta_i(dt)$, $i = 1, \ldots, q$. Formally, this can be expressed as $\zeta_0(dt) = z_A^{(0)} \delta_A(dt) +$

$z_B^{(0)}\delta_B(dt)+z(t)dt$ and $\zeta_i(dt) = z_A^{(i)}\delta_A(dt)+z_B^{(i)}\delta_B(dt)$ for $i = 1,\dots,q$, where $z(t)$ is some continuous function on $[A, B]$ and $z_A^{(i)}$, $z_B^{(i)}$ $(i = 0, 1,\dots,q)$ are some vectors. Define

$$\tilde{f}(t) = f(t) - \sum_{i=0}^{q} z_A^{(i)}\, K^{(i)}(A,t) - \sum_{i=0}^{q} z_B^{(i)}\, K^{(i)}(B,t)\,. \tag{2.8}$$

From (2.5), all components of $\tilde{f}$ belong to $SY(K)$ and $\tilde{f}(\cdot) = \int_A^B K(s,\cdot)z(s)ds$. Summarizing, for any sufficiently smooth $f \in H(K)$, the function $\tilde{f} \in SY(K)$ exists and is uniquely defined. The BLUE measures for f and $\tilde{f}$ can be chosen so that the measure for $\tilde{f}$ has no atoms and the continuous components of the BLUE measures for f and $\tilde{f}$ are proportional; we may call such $\tilde{f} \in SY(K)$ a representative of $f \in H(K)$ in $SY(K)$. Note also that the above discussion shows that the functions in $SY(K)$ have, as a rule, a very peculiar form.

2.8. *Signed least squares estimators and the BLUE.* The ordinary least square estimator (OLSE) of θ in the model (1.1) for the design measure ξ is given by $\hat{\theta}_{OLSE} = \int M^{-1}f(t)Y(t)\xi(dt)$ with $M = \int f(t)f^T(t)\xi(dt)$ and covariance matrix $D(\hat{\theta}_{OLSE}) = M^{-1}\left[\int\int K(t,s)f(t)f^T(s)\xi(dt)\xi(ds)\right]M^{-1}$. Assume that for some probability density $p(t)$ on $\mathcal{T}$ and some non-degenerate $m \times m$ matrix Λ we have $\int K(t,s)p(t)f(t)dt = \Lambda f(s)$ for all $s \in \mathcal{T}$. In this case, for the continuous design $\xi(dt) = p(t)dt$ we obtain

$$D(\hat{\theta}_{OLSE}) = M^{-1}\left[\int \Lambda f(s)f^T(s)p(s)ds\right]M^{-1} = M^{-1}\Lambda.$$

At the same time, the condition (2.5) with $q = 0$ is satisfied by the measure $\zeta_0(dt) = \Lambda^{-1}f(t)p(t)dt$ and hence from Corollary 2.1 we deduce that $\zeta_0(dt)$ gives the BLUE with covariance matrix

$$\left[\int \zeta_0(dt)f^T(t)\right]^{-1} = \left[\int \Lambda^{-1}f(t)f^T(t)p(t)dt\right]^{-1} = M^{-1}\Lambda.$$

This implies that in this case the OLSE with design $\xi(dt) = p(t)dt$ coincides with the continuous BLUE.

Matrix-weighted estimators (MWE) introduced in Dette et al. (2016) generalize the OLSE by giving specific $m \times m$ matrix weights to all points $t \in \mathcal{T}$. They showed that if $f \in H(K)$ and $q = 0$, then the optimal MWE is also the BLUE. If all matrix weights contain only $-1, 0$ or 1, then the MWE becomes the (generalized) signed least square estimator (SLSE). It was also shown that if the number of observations N tends to infinity then for a suitable sequence of designs the asymptotic covariance matrix of the discrete SLSE converges to the covariance matrix of the BLUE. However, unless $f \in SY(K)$, the rate convergence is extremely slow as many design points are required to emulate weights which the BLUE measure assigns to the end-points A and B. These results are generalizable to the case of differentiable kernels as considered in this paper.

2.9. *BLUE and energy minimization.* The problem of constructing the continuous BLUE generalizes the problem of the so-called energy minimization problem (see e.g. Sejdinovic et al. (2013) and Székely and Rizzo (2013) for details), which for a given (conditionally) positive definite kernel $K(s,t)$ is the minimization problem

$$\int_{\mathcal{T}} \int_{\mathcal{T}} K(s,t) G(ds) G(dt) \to \min_{G \in \mathcal{G}} \tag{2.9}$$

where $\mathcal{G}$ is the set of finite signed measures on $\mathcal{T}$ such that $\int_{\mathcal{T}} G(dt) = 1$ (signed measures of total mass 1). This is exactly the problem of construction of the continuous BLUE for the case $m = 1$, $q = 0$, general $\mathcal{T}$ and the location-scale regression model with $f(t) = 1$. For a general f with $m = 1$ and $q = 0$, the unbiasedness condition for a general linear estimator (2.1) is $\int_{\mathcal{T}} f(t) G(dt) = 1$ and it reduces to $\int_{\mathcal{T}} G(dt) = 1$ when $f(t) = 1$.
On the other hand, if $f(t) \neq 0$ for all $t \in \mathcal{T}$ then we can define $\tilde{G}(dt) = f(t) G(dt)$ and $\tilde{K}(s,t) = K(s,t)/(f(s)f(t))$. Then the problem of constructing the BLUE for the model (1.1) with $m = 1$, $q = 0$ and general $\mathcal{T}$ is exactly the energy minimization problem (2.9) for the kernel $\tilde{K}(s,t)$, assuming that it is also a positive definite kernel.

3. BLUE for processes with trajectories in $C^1[A, B]$.

In this section, we assume that the error process is exactly once continuously differentiable (in the mean-square sense).

3.1. *A general statement.* Consider the regression model (1.1) and a linear estimator in the form

$$\hat{\theta}_{G_0,G_1} = \int_{\mathcal{T}} y(t) G_0(dt) + \int_{\mathcal{T}} y^{(1)}(t) G_1(dt), \tag{3.1}$$

where $G_0(dt)$ and $G_1(dt)$ are signed vector-measures. The following corollary is a specialization of Corollary 2.1 when $q = 1$.

COROLLARY 3.1. *Consider the regression model (1.1) with the covariance kernel $K(t,s)$ and such that $y^{(1)}(t)$ exists in the mean-square sense for all $t \in [A,B]$. Suppose that $y(t)$ and $y^{(1)}(t)$ can be observed at all $t \in \mathcal{T}$. Assume that there exist vector-measures ζ_0 and ζ_1 such that the equality*

$$\int_{\mathcal{T}} K(t,s)\zeta_0(dt) + \int_{\mathcal{T}} K^{(1)}(t,s)\zeta_1(dt) = f(s),$$

is fulfilled for all $s \in \mathcal{T}$, and such that the matrix

$$C = \int_{\mathcal{T}} f(t)\zeta_0^T(dt) + \int_{\mathcal{T}} f^{(1)}(t)\zeta_1^T(dt)$$

is non-degenerate. Then the estimator $\hat{\theta}_{G_0,G_1}$ defined in (3.1) with $G_i = C^{-1}\zeta_i$ ($i=0,1$) is a BLUE with covariance matrix C^{-1}.

The next theorem provides sufficient conditions for vector-measures of some particular form to define a BLUE by (3.1) for the case $\mathcal{T} = [A,B]$. This theorem, which is proved in Section 6, will be useful for several choices of the covariance kernel below. Assume that $s_3 = K^{(3)}(s-,s) - K^{(3)}(s+,s)$ is a non-zero constant; here $K^{(j)}(s-,s)$ and $K^{(j)}(s+,s)$ are one-sided j-th derivatives of K at the diagonal. Define the vector-function

$$z(t) \;=\; (\tau_0 f(t) - \tau_2 f^{(2)}(t) + f^{(4)}(t))/s_3,$$

and vectors

$$\begin{aligned}
z_A &= \big(f^{(3)}(A) - \gamma_{1,A} f^{(1)}(A) + \gamma_{0,A} f(A)\big)/s_3,\\
z_B &= \big(-f^{(3)}(B) + \gamma_{1,B} f^{(1)}(B) + \gamma_{0,B} f(B)\big)/s_3,\\
z_{1,A} &= \big(-f^{(2)}(A) + \beta_{1,A} f^{(1)}(A) - \beta_{0,A} f(A)\big)/s_3,\\
z_{1,B} &= \big(f^{(2)}(B) + \beta_{1,B} f^{(1)}(B) + \beta_{0,B} f(B)\big)/s_3,
\end{aligned}$$

where $\tau_0, \tau_2, \gamma_{0,A}, \gamma_{1,A}, \beta_{0,A}, \beta_{1,A}, \gamma_{0,B}, \gamma_{1,B}, \beta_{0,B}, \beta_{1,B}$ are some constants. Define the functions

$$\begin{aligned}
J_1(s) &= -\gamma_{1,A}K(A,s) + \beta_{1,A}K^{(1)}(A,s) + \tau_2 K(A,s) - K^{(2)}(A,s),\\
J_2(s) &= \;\;\gamma_{0,A}K(A,s) - \beta_{0,A}K^{(1)}(A,s) - \tau_2 K^{(1)}(A,s) + K^{(3)}(A,s),\\
J_3(s) &= -\gamma_{1,B}K(B,s) + \beta_{1,B}K^{(1)}(B,s) - \tau_2 K(B,s) + K^{(2)}(B,s),\\
J_4(s) &= \;\;\gamma_{0,B}K(B,s) - \beta_{0,B}K^{(1)}(B,s) + \tau_2 K^{(1)}(B,s) - K^{(3)}(B,s).
\end{aligned} \tag{3.2}$$

THEOREM 3.1. *Consider the regression model* (1.1) *on the interval* $\mathcal{T} = [A, B]$ *with errors having the covariance kernel* $K(t, s)$. *Suppose that the vector of regression functions* f *is four times differentiable and the kernel* $K(t, s)$ *is once differentiable for all* $t, s \in [A, B]$ *and is four times differentiable for* $t \neq s$ *such that* $s_3 \neq 0$ *and* $K^{(i)}(s-, s) - K^{(i)}(s+, s) = 0$, $i = 0, 1, 2$. *Using the notation of the previous paragraph define the vector-measures* $\zeta_0(dt) = z_A\delta_A(dt) + z_B\delta_B(dt) + z(t)dt$ *and* $\zeta_1(dt) = z_{1,A}\delta_A(dt) + z_{1,B}\delta_B(dt)$. *Assume that there exist constants* $\tau_0, \tau_2, \gamma_{0,A},\ \gamma_{1,A}, \beta_{0,A},\ \beta_{1,A}, \gamma_{0,B}, \gamma_{1,B}, \beta_{0,B}, \beta_{1,B}$ *such that (i) the identity*

$$\tau_0 K(t, s) - \tau_2 K^{(2)}(t, s) + K^{(4)}(t, s) \equiv 0 \tag{3.3}$$

holds for all $t, s \in [A, B]$ *with* $t \neq s$, *(ii) the identity* $J_1(s) + J_2(s) + J_3(s) + J_4(s) \equiv 0$ *holds for all* $s \in [A, B]$, *and (iii) the matrix* $C = \int_{\mathcal{T}} f(t)\zeta_0^T(dt) + \int_{\mathcal{T}} f^{(1)}(t)\zeta_1^T(dt)$ *is non-degenerate. Then the estimator* $\hat{\theta}_{G_0,G_1}$ *defined in* (3.1) *with* $G_i(dt) = C^{-1}\zeta_i(dt)$ $(i = 0, 1)$ *is a BLUE with covariance matrix* C^{-1}.

3.2. *Two examples for integrated error processes.* In this section we illustrate the application of our results calculating the BLUE when errors follow an integrated Brownian motion and an integrated process with triangular-shape kernel. All results of this section can be verified by a direct application of Theorem 3.1. We first consider the case of Brownian motion, where the integrated covariance kernel is given by

$$\begin{aligned} K(t, s) &= \int_a^t \int_a^s \min(t', s')dt'ds' \\ &= \frac{\max(t, s)(\min(t, s)^2 - a^2)}{2} - \frac{a^2(\min(t, s) - a)}{2} - \frac{\min(t, s)^3 - a^3}{6} \qquad (3.4) \end{aligned}$$

and $0 \leq a \leq A$.

PROPOSITION 3.1. *Consider the regression model* (1.1) *with i covariance kernel* (3.4) *and suppose that* f *is four times differentiable on the interval* $[A, B]$. *Define the signed vector-measures* $\zeta_0(dt) = z_A\delta_A(dt) + z_B\delta_B(dt) + z(t)dt$ *and* $\zeta_1(dt) = z_{1,A}\delta_A(dt) + z_{1,B}\delta_B(dt)$, *where* $z(t) = f^{(4)}(t)$,

$$\begin{aligned} z_A &= f^{(3)}(A) - \frac{6(A + a)}{(A + 3a)(A - a)^2} f^{(1)}(A) + \frac{12A}{(A + 3a)(A - a)^3} f(A), \\ z_{1,A} &= -f^{(2)}(A) + \frac{4(A + 2a)}{(A + 3a)(A - a)} f^{(1)}(A) - \frac{6(A + a)}{(A + 3a)(A - a)^2} f(A), \\ z_B &= -f^{(3)}(B), \quad z_{1,B} = f^{(2)}(B). \end{aligned}$$

Assume that the matrix $C = \int_A^B f(t)\zeta_0^T(dt) + \int_{\mathcal{T}} f^{(1)}(t)\zeta_1^T(dt)$ *is non-degenerate. Then the estimator* $\hat{\theta}_{G_0,G_1}$ *defined in* (3.1) *with* $G_i(dt) = C^{-1}\zeta_i(dt)$ *is a BLUE with covariance matrix* C^{-1}.

The next example is a particular case of Proposition 3.1 when $a = 0$.

EXAMPLE 3.1. Consider the regression model (1.1) on $\mathcal{T} = [A, B]$ with the covariance kernel (3.4) with $a = 0$:

$$K(t,s) = \min(t,s)^2(3\max(t,s) - \min(t,s))/6\,. \tag{3.5}$$

Suppose that f is differentiable four times. Define the vector-measures $\zeta_0(dt) = z_A\delta_A(dt) + z_B\delta_B(dt) + z(t)dt$ and $\zeta_1(dt) = z_{1,A}\delta_A(dt) + z_{1,B}\delta_B(dt)$, where

$$\begin{aligned} z_A &= f^{(3)}(A) - \frac{6}{A^2}f^{(1)}(A) + \frac{12}{A^3}f(A), \\ z_{1,A} &= -f^{(2)}(A) + \frac{4}{A}f^{(1)}(A) - \frac{6}{A^2}f(A), \\ z_B &= -f^{(3)}(B), \quad z_{1,B} = f^{(2)}(B), \;\; z(t) = f^{(4)}(t). \end{aligned} \tag{3.6}$$

If $C = \int_A^B f(t)\zeta_0^T(dt) + \int_A^B f^{(1)}(t)\zeta_1^T(dt)$ is non-degenerate then the estimator $\hat{\theta}_{G_0,G_1}$ with $G_i(dt) = C^{-1}\zeta_i(dt)$ is a BLUE with covariance matrix C^{-1}.

As shown in Section 2.5, the expressions (3.6) are not the only expressions defining the BLUE; indeed, using Lemma 2.2, we can construct many other measures defining a BLUE. Specifically, let $\psi(t)$ be a vector of arbitrary differentiable functions on $\mathcal{T}$. Define the vector-measures $\zeta_{0,\psi}(dt) = \zeta_0(dt) - \psi(A)\delta_A(dt) + \psi(B)\delta_B(dt) + \psi^{(1)}(t)dt$ and $\zeta_{1,\psi}(dt) = \zeta_1(dt) + \psi(t)$. Then the matrix C does not depend on the choice of ψ and all estimators $\hat{\theta}_{G_{0,\psi},G_{1,\psi}}$ with $G_{i,\psi}(dt) = C^{-1}\zeta_{i,\psi}(dt)$ are BLUE. In particular, if $\psi(t) \equiv 0$, then we get the expression (3.6), where the derivative of $y(t)$ at the interior points of $[A, B]$ is not used. However, if we choose $\psi(t)$ such that $\psi^{(1)}(t) = -f^{(4)}(t)$ for all $t \in [A, B]$, then the estimator $\hat{\theta}_{G_{0,\psi},G_{1,\psi}}$ would not use observations of the process $y(t)$ but instead use the observations of the derivative $y^{(1)}(t)$ at the interior points of the interval $[A, B]$; the corresponding BLUE is defined by the vector-measures

$$\begin{aligned} \zeta_0(dt) &= \Big[-\frac{6}{A^2}f^{(1)}(A) + \frac{12}{A^3}f(A)\Big]\delta_A(dt), \\ \zeta_1(dt) &= -\Big[f^{(2)}(A) - \frac{4}{A}f^{(1)}(A) + \frac{6}{A^2}f(A)\Big]\delta_A(dt) \\ &\quad + f^{(2)}(B)\delta_B(dt) - f^{(3)}(t)dt. \end{aligned} \tag{3.7}$$

In particular, for the location-scale model with $f(t) \equiv 1$ and arbitrary ψ we obtain $\zeta_{0,\psi}(dt) = 12/A^3\delta_A(dt) - \psi(A)\delta_A(dt) + \psi(B)\delta_B(dt) + \psi^{(1)}(t)dt$ and $\zeta_{1,\psi}(dt) = -6/A^2\delta_A(dt) + \psi(t)$. This gives different BLUE-defining measures G but the value $C = \big[12/A^3 - \psi(A) + \psi(B) + \int_A^B \psi^{(1)}(t)dt\big] = 12/A^3$ (the inverse of the BLUE variance) does not depend on the choice of ψ.

Consider now the integrated triangular-shape kernel

$$
\begin{aligned}
K(t,s) &= \int_0^t \int_0^s \max\{0, 1-\lambda|t'-s'|\}dt'ds' \\
&= ts - \lambda\min(t,s)\Big(3\max(t,s)^2 - 3ts + 2\min(t,s)^2\Big)/6. \qquad (3.8)
\end{aligned}
$$

Proposition 3.2. *Consider the regression model (1.1) on $\mathcal{T} = [A,B]$ with integrated covariance kernel (3.8), where $\lambda(B-A) < 1$. Suppose that f is four times differentiable. Define the signed vector-measures $\zeta_0(dt) = z_A\delta_A(dt) + z_B\delta_B(dt) + z(t)dt$ and $\zeta_1(dt) = z_{1,A}\delta_A(dt) + z_{1,B}\delta_B(dt)$, where $z(t) = f^{(4)}(t)/(2\lambda)$ and*

$$
\begin{aligned}
z_A &= \Big[f^{(3)}(A) - \frac{6\kappa_2}{A^2\kappa_4}f^{(1)}(A) + \frac{6\lambda}{A\kappa_4}f^{(1)}(B) + \frac{12\kappa_1}{A^3\kappa_4}f(A)\Big]/(2\lambda),\\
z_{1,A} &= \Big[-f^{(2)}(A) + \frac{4\kappa_3}{A\kappa_4}f^{(1)}(A) - \frac{2\lambda}{\kappa_4}f^{(1)}(B) - \frac{6\kappa_2}{A^2\kappa_4}f(A)\Big]/(2\lambda),\\
z_{1,B} &= \Big[f^{(2)}(B) - \frac{2\lambda}{\kappa_4}f^{(1)}(A) + \frac{4\lambda}{\kappa_4}f^{(1)}(B) + \frac{6\lambda}{A\kappa_4}f(A)\Big]/(2\lambda),\\
z_B &= -f^{(3)}(B)/(2\lambda), \quad \kappa_j = A\lambda - jB\lambda + 2j.
\end{aligned}
$$

Assume that the matrix $C = \int_A^B f(t)\zeta_0^T(dt) + \int_A^B f^{(1)}(t)\zeta_1^T(dt)$ is non-degenerate. Then the estimator $\hat\theta_{G_0,G_1}$ defined in (3.1) with $G_i(dt) = C^{-1}\zeta_i(dt)$ is a BLUE with covariance matrix C^{-1}.

3.3. *Explicit form of the BLUE for the integrated processes.* We conclude this section establishing a direct link between the BLUE for models with non-differentiable error processes and the BLUE for regression models with an integrated kernel of the form (3.11) below. Note that this extends the class of kernels considered in Sacks and Ylvisaker (1970) in a nontrivial way. Consider the regression model (1.1) with a non-differentiable error process with covariance kernel $R(t,s)$ and the BLUE $\hat\theta_{G_0} = \int_{\mathcal{T}} y(t)G_0(dt)$. From Corollary 2.1 we have for the vector-measure $\zeta_0(dt)$ satisfying (2.5) and defining the BLUE

$$
\int_A^B R(t,s)\zeta_0(dt) = f(s) \qquad (3.9)
$$

and $\mathrm{Var}(\hat\theta_{G_0}) = C^{-1} = \big(\int_{\mathcal T} f(t)\zeta_0^T(dt)\big)^{-1}$. The unbiasedness condition for the measure $G_0(dt) = C^{-1}\zeta_0(dt)$ is

$$\int_{\mathcal T} f(t)G_0^T(dt) = I_m.$$

Define the integrated process as follows:

$$\widetilde{y}(t) = \int_a^t y(u)du, \quad \widetilde{f}(t) = \int_a^t f(u)du, \quad \widetilde{\varepsilon}(t) = \int_a^t \varepsilon(u)du$$

with some $a \le A$ (meaning that the regression vector-function and the error process are defined on $[a,B]$ but observed on $[A,B]$) so that

$$\widetilde{f}^{(1)}(t) = f(t), \quad \widetilde{y}^{(1)}(t) = y(t), \quad \widetilde{\varepsilon}^{(1)}(t) = \varepsilon(t)\ .$$

Consider the regression model

$$\tilde y(t) = \theta^T \tilde f(t) + \tilde\varepsilon(t), \tag{3.10}$$

which has the integrated covariance kernel

$$K(t,s) = \int_a^t \int_a^s R(u,v)dudv. \tag{3.11}$$

The proof of the following result is given in Section 6.

THEOREM 3.2. *Let the vector-measure ζ_0 satisfy the equality (3.9) and define the BLUE $\hat\theta_{G_0}$ with $G_0(dt) = C^{-1}\zeta_0(dt)$ in the regression model (1.1) with covariance kernel $R(\cdot,\cdot)$. Let the measures η_0, η_1 satisfy the equality*

$$\int_{\mathcal T} K(t,s)\eta_0(dt) + \int_{\mathcal T} K^{(1)}(t,s)\eta_1(dt) = 1 \tag{3.12}$$

for all $s \in \mathcal T$. Define the vector-measures $\tilde\zeta_0 = -c\eta_0$ and $\tilde\zeta_1 = -c\eta_1 + \zeta_0$, where the vector c is given by $c = \int_a^A [\int_A^B R(t,s)\zeta_0(dt) - f(s)]ds$. Then the estimator $\hat\theta_{\tilde G_0, \tilde G_1}$ defined in (3.1) with $\tilde G_i(dt) = \tilde C^{-1}\tilde\zeta_i(dt)$ $(i = 1, 2)$, where

$$\tilde C = \int \tilde f(t)\tilde\zeta_0^T(dt) + \int \tilde f^{(1)}(t)\tilde\zeta_1^T(dt),$$

is a BLUE in the regression model (3.10) with kernel (3.11).

Repeated application of Theorem 3.2 extends the results to the case of several times integrated processes.
If $a = A$ in (3.11) we have $c = 0$ in Theorem 3.2; in this case, the statement of Theorem 3.2 can be proved easily. Moreover, in this case the class of kernels defined by (3.11) is exactly the class of kernels considered in equation (1.5) and (1.6) of Sacks and Ylvisaker (1970) for once differentiable processes ($k = 1$ in their notation). We emphasize that the class of kernels considered here is much richer than the class considered in this reference.

3.4. *BLUE for AR(2) errors.* Consider the continuous-time regression model (1.1), which can be observed at all $t \in [A, B]$, where the error process is a continuous autoregressive (CAR) process of order 2. Formally, a CAR(2) process is defined as a solution of the linear stochastic differential equation of the form

$$d\varepsilon^{(1)}(t) = \tilde{a}_1 \varepsilon^{(1)}(t) + \tilde{a}_2 \varepsilon(t) + \sigma_0^2 dW(t), \tag{3.13}$$

where $\tilde{a}_1$ and $\tilde{a}_2$ are constants, $\mathrm{Var}(\varepsilon(t)) = \sigma^2$ and $W(t)$ is a standard Wiener process, [see Brockwell et al. (2007)]. Note that the process $\{\varepsilon(t) | t \in [A, B]\}$ defined by (3.13) has a continuous derivative and, consequently, the process $\{y(t) = \theta^T f(t) + \varepsilon(t) |\ t \in [A, B]\}$, is a continuously differentiable process with drift on the interval $[A, B]$. In this section we derive the explicit form for the continuous BLUE using Theorem 3.1. An alternative approach would be to use the coefficients of the equation (3.13) as indicated in Parzen (1961). There are in fact three different forms of the autocorrelation function $\rho(t) = K(0, t)$ of CAR(2) processes [see e.g. formulas (14)–(16) in He and Wang (1989)], which are given by

$$\rho_1(t) = \frac{\lambda_2}{\lambda_2 - \lambda_1} e^{-\lambda_1 |t|} - \frac{\lambda_1}{\lambda_2 - \lambda_1} e^{-\lambda_2 |t|}\,, \tag{3.14}$$

where $\lambda_1 \neq \lambda_2$, $\lambda_1 > 0$, $\lambda_2 > 0$, by

$$\rho_2(t) = e^{-\lambda |t|} \Big\{ \cos(\omega |t|) + \frac{\lambda}{\omega} \sin(\omega |t|) \Big\}\,, \tag{3.15}$$

where $\lambda > 0$, $\omega > 0$, and by

$$\rho_3(t) = e^{-\lambda |t|} (1 + \lambda |t|)\,, \tag{3.16}$$

where $\lambda > 0$. Note that the kernel (3.16) is widely known as Matérn kernel with parameter 3/2, which has numerous applications in spatial statistics [see Rasmussen and Williams (2006)] and computer experiments [see Pronzato and Müller (2012)]. In the following results, which are proved in Section 6.7, we specify the BLUE for the CAR(2) model.

PROPOSITION 3.3. *Consider the regression model* (1.1) *with CAR(2) errors, where the covariance kernel* $K(t,s) = \rho(t-s)$ *has the form* (3.14). *Suppose that* f *is a vector of linearly independent, four times differentiable functions on the interval* $[A,B]$. *Then the conditions of Theorem 3.1 are satisfied for* $s_3 = 2\lambda_1\lambda_2(\lambda_1+\lambda_2)$, $\tau_0 = \lambda_1^2\lambda_2^2$, $\tau_2 = \lambda_1^2+\lambda_2^2$, $\beta_{j,A} = \beta_{j,B} = \beta_j$ *and* $\gamma_{j,A} = \gamma_{j,B} = \gamma_j$ *for* $j = 0,1$, *where* $\beta_1 = \lambda_1+\lambda_2$, $\gamma_1 = \lambda_1^2+\lambda_1\lambda_2+\lambda_2^2$, $\beta_0 = \lambda_1\lambda_2$ *and* $\gamma_0 = \lambda_1\lambda_2(\lambda_1+\lambda_2)$.

PROPOSITION 3.4. *Consider the regression model* (1.1) *with CAR(2) errors, where the covariance kernel* $K(t,s) = \rho(t-s)$ *has the form* (3.15). *Suppose that* f *is a vector of linearly independent, four times differentiable functions. Then the conditions of Theorem 3.1 hold for* $s_3 = 4\lambda(\lambda^2+\omega^2)$, $\tau_0 = (\lambda^2+\omega^2)^2$, $\tau_2 = 2(\lambda^2-\omega^2)$, $\beta_{j,A} = \beta_{j,B} = \beta_j$ *and* $\gamma_{j,A} = \gamma_{j,B} = \gamma_j$ *for* $j = 0,1$, *where* $\beta_1 = 2\lambda$, $\gamma_1 = \gamma_1 = 3\lambda^2-\omega^2$, $\beta_0 = \lambda^2+\omega^2$ *and* $\gamma_0 = 2\lambda(\lambda^2+\omega^2)$.

The BLUE for the covariance kernel in the form (3.16) is obtained from either Proposition 3.3 with $\lambda_1 = \lambda_2 = \lambda$ or Proposition 3.4 with $\omega = 0$.

REMARK 3.1. In the online supplement Dette et al. (2018) we consider the regression model (1.2) with a discrete AR(2) error process. Although the discretised CAR(2) process follows an ARMA$(2,1)$ model rather than an AR(2) [see He and Wang (1989)] we will be able to establish the connection between the BLUE in the discrete and continuous-time models and hence derive the limiting form of the discrete BLUE and its covariance matrix.

4. Models with more than once differentiable error processes. If $\mathcal{T} = [A,B]$ and $q > 1$ then solving the Wiener-Hopf type equation (2.5) numerically is virtually impossible in view of the fact that the problem is severely ill-posed. Derivation of explicit forms of the BLUE for smooth kernels with $q > 1$ is hence extremely important. We did not find any general results on the form of the BLUE in such cases. In particular, the well-known paper Sacks and Ylvisaker (1970) dealing with these kernels does not contain any specific examples. In Theorem 3.2 we have already established a general result that can be used for deriving explicit forms for the BLUE for $q > 1$ times integrated kernels, which can be used repeatedly for this purpose. We can also formulate a result similar to Theorem 3.1. However, already for $q = 2$, even a formulation of such theorem would take a couple of pages and hence its usefulness would be very doubtful.
In this section, we indicate how the general methodologies developed in the previous sections can be extended to error processes with $q > 1$ by two

examples: twice integrated Brownian motion and CAR(p) error models with $p \geq 3$, but other cases can be treated very similarly.

4.1. *Twice integrated Brownian motion.*

PROPOSITION 4.1. *Consider the regression model* (1.1) *where the error process is the twice integrated Brownian motion with the covariance kernel*

$$K(t,s) = t^5/5! - st^4/4! + s^2t^3/12, \ t < s.$$

Suppose that f is 6 times differentiable and define the vector-measures $\zeta_0(dt) = z_A\delta_A(dt) + z_B\delta_B(dt) + z(t)dt, \ \zeta_1(dt) = z_{1,A}\delta_A(dt) + z_{1,B}\delta_B(dt), \ \zeta_2(dt) = z_{2,A}\delta_A(dt) + z_{2,B}\delta_B(dt)$, *where* $z(t) = -f^{(6)}(t)$,

$$\begin{aligned}
z_A &= (-A^5 f^{(5)}(A) + 60A^2 f^{(2)}(A) - 360Af^{(1)}(A) + 720f(A))/A^5, \\
z_{1,A} &= (A^4 f^{(4)}(A) - 36A^2 f^{(2)}(A) + 192Af^{(1)}(A) - 360f(A))/A^4, \\
z_{2,A} &= (-A^3 f^{(3)}(A) + 9A^2 f^{(2)}(A) - 36Af^{(1)}(A) + 60f(A))/A^3, \\
z_B &= f^{(5)}(B), \quad z_{1,B} = -f^{(4)}(B), \quad z_{2,B} = f^{(3)}(B).
\end{aligned}$$

Then the estimator $\hat{\theta}_{G_0,G_1,G_2}$ *defined by* (2.1) *(for* $q = 2$*) with* $G_i(dt) = C^{-1}\zeta_i(dt)$ $(i = 0,1,2)$,

$$C = \int_{\mathcal{T}} f(t)\zeta_0^T(dt) + \int_{\mathcal{T}} f^{(1)}(t)\zeta_1^T(dt) + \int_{\mathcal{T}} f^{(2)}(t)\zeta_2^T(dt),$$

is the BLUE with covariance matrix C^{-1}.

4.2. *CAR(p) models with p≥* 3. Consider the regression model (1.1), which can be observed at all $t \in [A,B]$ and the error process has the continuous autoregressive (CAR) structure of order p. Formally, a CAR(p) process is a solution of the linear stochastic differential equation of the form

$$d\varepsilon^{(p-1)}(t) = \tilde{a}_1\varepsilon^{(p-1)}(t) + \ldots + \tilde{a}_p\varepsilon(t) + \sigma_0^2 dW(t),$$

where $\mathrm{Var}(\varepsilon(t)) = \sigma^2$ and W is a standard Wiener process, [see Brockwell et al. (2007)]. Note that the process ε has continuous derivatives $\varepsilon^{(1)}(t), \ldots, \varepsilon^{(p-1)}(t)$ at the point t and, consequently, the process $\{y(t) = \theta^T f(t) + \varepsilon(t) | \ t \in [A,B]\}$ is continuously differentiable $p-1$ times on the interval $[A,B]$ with drift $\theta^T f(t)$. Define the vector-functions

$$z(t) = (\tau_0 f(t) + \tau_2 f^{(2)}(t) + \ldots + f^{(2p)}(t))/s_{2p-1},$$

and vectors

$$z_{j,A} = \sum_{l=0}^{2p-j-1} \gamma_{l,j,A} f^{(j)}(A)/s_{2p-1},$$
$$z_{j,B} = \sum_{l=0}^{2p-j-1} \gamma_{l,j,B} f^{(j)}(B)/s_{2p-1}$$

for $j = 0, 1, \ldots, p-1$, where $s_{2p-1} = K^{(2p-1)}(s-, s) - K^{(2p-1)}(s+, s)$.

PROPOSITION 4.2. *Consider the regression model (1.1) with CAR(p) errors. Define the vector-measures*

$$\zeta_0(dt) = z_{0,A}\delta_A(dt) + z_{0,B}\delta_B(dt) + z(t)dt,$$
$$\zeta_j(dt) = z_{j,A}\delta_A(dt) + z_{j,B}\delta_B(dt), \; j = 1, \ldots, p-1,$$

for $j = 1, \ldots, p-1$. Then there exist constants $\tau_0, \tau_2 \ldots, \tau_{2(p-1)}$ and $\gamma_{l,j,A}, \gamma_{l,j,B}$, such that the estimator $\hat{\theta}_{G_0,G_1,\ldots,G_{p-1}}$ defined by (2.1) (for $q = p-1$) with $G_j(dt) = C^{-1}\zeta_j(dt)$ $(i = 0, 1, \ldots, p-1)$,

$$C = \int_{\mathcal{T}} f(t)\zeta_0^T(dt) + \sum_{j=1}^{p-1} \int_{\mathcal{T}} f^{(j)}(t)\zeta_j^T(dt),$$

is a BLUE with covariance matrix C^{-1}.

Let us consider the construction of a BLUE for model (1.1) with a CAR(3) error process in more detail. One of several possible forms for the covariance function for the CAR(3) process is given by

$$\rho(t) = c_1 e^{-\lambda_1|t|} + c_2 e^{-\lambda_2|t|} + c_3 e^{-\lambda_3|t|}, \tag{4.1}$$

where $\lambda_1, \lambda_2, \lambda_3$ are the roots of the autoregressive polynomial $\tilde{a}(z) = z^3 + \tilde{a}_1 z^2 + \tilde{a}_2 z + \tilde{a}_3$,

$$c_j = \frac{k_j}{k_1 + k_2 + k_3}, \quad k_j = \frac{1}{\tilde{a}'(\lambda_j)\tilde{a}(-\lambda_j)},$$

$\lambda_i \neq \lambda_j$, $\lambda_i > 0$, $i, j = 1, \ldots, 3$, see Brockwell (2001). Specifically, we have

$$c_1 = \frac{\lambda_2\lambda_3(\lambda_2 + \lambda_3)}{(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3)(\lambda_1 + \lambda_2 + \lambda_3)},$$
$$c_2 = \frac{\lambda_1\lambda_3(\lambda_1 + \lambda_3)}{(\lambda_2 - \lambda_1)(\lambda_2 - \lambda_3)(\lambda_1 + \lambda_2 + \lambda_3)},$$
$$c_3 = \frac{\lambda_1\lambda_2(\lambda_1 + \lambda_2)}{(\lambda_3 - \lambda_1)(\lambda_3 - \lambda_2)(\lambda_1 + \lambda_2 + \lambda_3)}.$$

In this case, a BLUE is given in Proposition 4.2 with the following parameters:

$$\tau_0 = -\lambda_1^2\lambda_2^2\lambda_3^2,\ \tau_2 = \lambda_1^2\lambda_2^2 + \lambda_1^2\lambda_3^2 + \lambda_2^2\lambda_3^2,\ \tau_4 = -\lambda_1^2 - \lambda_2^2 - \lambda_3^2,$$

$$s_5 = \frac{2\lambda_1\lambda_2\lambda_3(\lambda_1+\lambda_2)(\lambda_1+\lambda_3)(\lambda_2+\lambda_3)}{\lambda_1+\lambda_2+\lambda_3} = 2\frac{\prod_i \lambda_i \prod_{i\neq j}(\lambda_i+\lambda_j)}{\sum_i \lambda_i},$$

$$\begin{aligned}
z_{0,A} &= f^{(5)}(A) - \textstyle\sum_i\lambda_i^2 f^{(3)}(A) - \prod_i\lambda_i f^{(2)}(A)\\
&\quad +[\textstyle\sum_{i\neq j}\lambda_i^2\lambda_j^2 + \prod_i\lambda_i\sum_i\lambda_i]f^{(1)}(A) - \prod_i\lambda_i\sum_{i\neq j}\lambda_i\lambda_j f(A)\\
z_{1,A} &= -f^{(4)}(A) + \textstyle\sum_{i,j}\lambda_i\lambda_j f^{(2)}(A) - \prod_{i\neq j}(\lambda_i+\lambda_j)f^{(1)}(A) + \prod_i\lambda_i\sum_i\lambda_i f(A)\\
z_{2,A} &= f^{(3)}(A) - \textstyle\sum_i\lambda_i f^{(2)}(A) + \sum_{i\neq j}\lambda_i\lambda_j f^{(1)}(A) - \prod_i\lambda_i f(A)
\end{aligned}$$

$$\begin{aligned}
-z_{0,B} &= f^{(5)}(B) - \textstyle\sum_i\lambda_i^2 f^{(3)}(B) - \prod_i\lambda_i f^{(2)}(B)\\
&\quad +[\textstyle\sum_{i\neq j}\lambda_i^2\lambda_j^2 + \prod_i\lambda_i\sum_i\lambda_i]f^{(1)}(B) - \prod_i\lambda_i\sum_{i\neq j}\lambda_i\lambda_j f(B)\\
-z_{1,B} &= -f^{(4)}(B) + \textstyle\sum_{i,j}\lambda_i\lambda_j f^{(2)}(B) - \prod_{i\neq j}(\lambda_i+\lambda_j)f^{(1)}(B) + \prod_i\lambda_i\sum_i\lambda_i f(B)\\
-z_{2,B} &= f^{(3)}(B) - \textstyle\sum_i\lambda_i f^{(2)}(B) + \sum_{i\neq j}\lambda_i\lambda_j f^{(1)}(B) - \prod_i\lambda_i f(B)
\end{aligned}$$

If we set $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$ then the above formulas give the explicit form of the BLUE for the Matérn kernel with parameter $5/2$; that is, the kernel defined by $\rho(t) = \left(1+\sqrt{5}t\lambda + 5t^2\lambda^2/3\right)\exp\left(-\sqrt{5}t\lambda\right)$.

5. Numerical study. In this section, we describe some numerical results on comparison of the accuracy of various estimators for the parameters in the regression models (1.1) with $[A,B]=[1,2]$ and the integrated Brownian motion as error process. The kernel $K(t,s)$ is given in (3.5) and the explicit form of the covariance matrix of the continuous BLUE can be found in Example 3.1. We denote this estimator by $\hat\theta_{cont.BLUE}$. We are interested in the efficiency of various estimators for this differentiable error process. For a given N (in the tables, we use $N = 3, 5, 10$), we consider the following four estimators that use $2N$ observations:

- $\hat\theta_{BLUE}(N,N)$: discrete BLUE based on observations $y(t_1),\dots,y(t_N)$, $y'(t_1),\dots,y'(t_N)$ with $t_i = 1+(i-1)/(N-1)$, $i=1,\dots,N$. This estimator uses N observations of the original process and its derivative (at equidistant points).

- $\hat{\theta}_{BLUE}(2N-2,2)$: discrete BLUE based on observations $y(t_1),\ldots,y(t_{2N-2}),y'(1),y'(2)$ with $t_i = 1+(i-1)/(2N-3)$, $i=1,\ldots,2N-3$. This estimator uses $2N-2$ observations of the original process (at equidistant points) and observations of its derivative at the boundary points of the design space.
- $\hat{\theta}_{BLUE}(2N,0)$: discrete BLUE based on observations $y(t_1),\ldots,y(t_{2N})$ with $t_i = 1+(i-1)/(2N-1)$, $i=1,\ldots,2N$. This estimator uses $2N$ observations of the original process (at equidistant points) and no observations from its derivative.
- $\hat{\theta}_{OLSE}(2N,0)$: ordinary least square estimator (OLSE) based on observations $y(t_1),\ldots,y(t_{2N})$ with $t_i = 1+(i-1)/(2N-1)$, $i=1,\ldots,2N$. This estimator uses $2N$ observations of the original process (at equidistant points) and no observations from its derivative.

In Table 1 we use the results derived in this paper to calculate the efficiencies

$$\mathrm{Eff}(\tilde{\theta}) = \frac{\mathrm{Var}(\hat{\theta}_{cont.BLUE})}{\mathrm{Var}(\tilde{\theta})}, \tag{5.1}$$

where $\tilde{\theta}$ is one of the four estimators under consideration. In particular we consider three different scenarios for the response function $f(t)$ in model (1.1):

$$m=1,\ \ f(t)=1, \tag{5.2}$$

$$m=3,\ \ f(t)=(1,\sin(3\pi t),\cos(3\pi t))^T, \tag{5.3}$$

$$m=5,\ \ f(t)=(1,t,t^2,1/t,1/t^2)^T. \tag{5.4}$$

The formulas provided in Example 3.1 give us expressions for a continuous BLUE. For the model (5.2) (recall that $[A,B]=[1,2]$) we obtain $\zeta_0(dt) = 12\delta_1(dt)$ and $\zeta_1(dt) = -6\delta_1(dt)$. Therefore, the estimator $\hat{\theta}_{cont.BLUE} = y(1)-0.5y'(1)$ is a BLUE. For the model (5.3) we obtain from Example 3.5 the vector-measures $\zeta_0(dt) = z_A\delta_A(dt)+z_B\delta_B(dt)+z(t)dt$ and $\zeta_1(dt) = z_{1,A}\delta_A(dt)+z_{1,B}\delta_B(dt)$, where $z(t) = 3^4\pi^4(0,\sin(3\pi t),\cos(3\pi t))^T$,

$$z_A = \begin{pmatrix} 12 \\ 27\pi^3+18\pi \\ -12 \end{pmatrix},\ z_B = \begin{pmatrix} 0 \\ 27\pi^3 \\ 0 \end{pmatrix},\ z_{1,A} = \begin{pmatrix} -6 \\ -12\pi \\ -9\pi^2+6 \end{pmatrix},\ z_{1,B} = \begin{pmatrix} 0 \\ 0 \\ -9\pi^2 \end{pmatrix}.$$

Similarly we get the BLUE measures for the model (5.4).

The results are very typical for many regression models with differentiable error processes (i.e. $q=1$) and can be summarized as follows. Any BLUE is far superior to the OLSE and any BLUE becomes highly efficient when N is large. Moreover, the use of information from the derivatives in constructing

Table 1
The efficiency defined by (5.1) for four different estimators based on $2N$ observations and the regression functions in (5.2) - (5.4)

model	(5.2)			(5.3)			(5.4)		
N	3	5	10	3	5	10	3	5	10
$\hat{\theta}_{BLUE}(N,N)$	1	1	1	0.412	0.929	0.997	0.696	0.960	0.998
$\hat{\theta}_{BLUE}(2N-2,2)$	1	1	1	0.456	0.987	0.999	0.869	0.994	0.999
$\hat{\theta}_{BLUE}(2N,0)$	0.859	0.915	0.957	0.478	0.772	0.896	0.100	0.333	0.625
$\hat{\theta}_{OLSE}(2N,0)$	0.073	0.073	0.073	0.001	0.001	0.002	0.089	0.141	0.119

BLUEs typically makes them more efficient than the BLUE which only uses values of $\{y(t)|t \in \mathcal{T}\}$; this is not true in general: see the case $N = 3$ for model (5.3) in Table 1. We also emphasize that the BLUEs which use more than two values of the derivative y' of the process have lower efficiency than the BLUE that uses exactly two values of derivatives, $y'(A)$ and $y'(B)$ (recall that the total number of observations is fixed). Therefore the best way of constructing the BLUE for N observations in the interval $[A, B]$ is to emulate the asymptotic BLUE: that is, to use $y'(A)$ and $y'(B)$ but for the other $N-2$ observations use values of the process $\{y(t)|t \in \mathcal{T}\}$. Similarly, for q times differentiable processes $y(t)$ with $q > 1$ and N large enough, the most efficient BLUE construction procedure would suggest observing values of the derivatives $y^{(i)}(A)$ and $y^{(i)}(B)$ for $i = 1, \ldots, q$ and using remaining $N - 2q$ observations for observing values of process $\{y(t)|t \in \mathcal{T}\}$.

6. Appendix.

6.1. *Proof of Lemma 2.1.* The mean of $\hat{\theta}_G^T$ is

$$\mathbb{E}[\hat{\theta}_G^T] = \theta^T \sum_{i=0}^{q} \int_{\mathcal{T}} f^{(i)}(t) G_i^T(dt) = \theta^T \int_{\mathcal{T}} F(t) G^T(dt)\,,$$

where $F(t) = (f(t), f^{(1)}(t), \ldots, f^{(q)}(t))$. This implies that the estimator $\hat{\theta}_G$ is unbiased if and only if

$$\int_{\mathcal{T}} F(t) G^T(dt) = I_m. \tag{6.1}$$

Since $G = (G_0, G_1, \ldots, G_q)$ with $G_i = C^{-1}\zeta_i$, we have

$$\int_{\mathcal{T}} F(t) G^T(dt) = \sum_{i=0}^{q} \int_{\mathcal{T}} f^{(i)}(t) \zeta_i^T(dt) {C^{-1}}^T = C^T {C^{-1}}^T = I_m\,,$$

which completes the proof.

6.2. *Proof of Theorem 2.1* .

I. We will call a signed matrix-measure G unbiased if the associated estimator $\widehat{\theta}_G$ defined in (2.1) is unbiased; that is, (6.1) holds. The set of all unbiased signed matrix-measures will be denoted by $\mathcal{S}$. This set is convex; moreover, if $G, H \in \mathcal{S}$ then $(1-\alpha)G + \alpha H \in \mathcal{S}$ for any real α.

The covariance matrix of any estimator $\widehat{\theta}_G$ is the matrix-valued function $\phi(G) = \mathrm{Var}(\hat{\theta}_G)$ defined in (2.3). The BLUE minimizes this matrix-valued function on the set $\mathcal{S}$.

Introduce the vector-function $d : \mathcal{T} \times \mathcal{S} \to \mathbb{R}^m$ by

$$d(s, G) = \sum_{j=0}^{q} \int_{\mathcal{T}} K^{(j)}(t, s) G_j(dt) - \phi(G) f(s) \,.$$

The validity of (2.4) for all $s \in \mathcal{T}$ is equivalent to the validity of $d(s, G) = 0_{m\times 1}$ for all $s \in \mathcal{T}$. Hence we are going to prove that $\widehat{\theta}_G$ is the BLUE if and only if $d(s, G) = 0_{m\times 1}$ for all $s \in \mathcal{T}$. For this purpose we will need the following auxiliary result.

LEMMA 6.1. *For any $G \in \mathcal{S}$ we have $\int_{\mathcal{T}} \mathbf{d}(s, G) G^T(ds) = 0_{m\times m}$, where $\mathbf{d}(s, G) = (d(s, G), d^{(1)}(s, G), \ldots, d^{(q)}(s, G))$ is an $m \times (q+1)$ matrix.*

Proof of Lemma 6.1 Using the unbiasedness condition (6.1), we have

$$\begin{aligned} \int_{\mathcal{T}} \mathbf{d}(s, G) G^T(ds) &= \int_{\mathcal{T}} \int_{\mathcal{T}} G(dt) \mathbf{K}(t, s) G^T(ds) - \phi(G) \int_{\mathcal{T}} F(s) G^T(ds) \\ &= \phi(G) - \phi(G) I_m = 0_{m\times m}. \end{aligned}$$

□

For any two matrix-measures G and H in $\mathcal{S}$, denote

$$\Phi(G, H) = \int_{\mathcal{T}} \int_{\mathcal{T}} G(dt) \mathbf{K}(t, s) H^T(ds)$$

which is a matrix of size $m \times m$. Note that for any $G \in \mathcal{S}$, the matrix $\phi(G) = \Phi(G, G)$ is exactly $\mathrm{Var}(\hat{\theta}_G)$, the covariance matrix of $\hat{\theta}_G$, see (2.3). For any two matrix-measures G and H in $\mathcal{S}$ and any real α, we have

$$\phi((1-\alpha)G + \alpha H) = (1-\alpha)^2 \phi(G) + \alpha^2 \phi(H) + \alpha(1-\alpha) \left[\Phi(G, H) + \Phi(H, G)\right] .$$

The directional derivative of $\phi((1-\alpha)G + \alpha H)$ as $\alpha \to 0$ is

$$\frac{\partial}{\partial \alpha} \phi((1-\alpha)G + \alpha H)\Big|_{\alpha=0} = \Phi(G, H) + \Phi(H, G) - 2\phi(G). \tag{6.2}$$

To rewrite (6.2), we note that $\int_{\mathcal{T}} \mathbf{d}(s,G)H^T(ds)$ can be written as

$$\begin{aligned} \int_{\mathcal{T}} \mathbf{d}(s,G)H^T(ds) &= \Phi(G,H) - \phi(G)\int_{\mathcal{T}} F(s)H^T(ds) \\ &= \Phi(G,H) - \phi(G), \end{aligned} \tag{6.3}$$

where in the last equality we have used the unbiasedness condition (6.1) for H. Using (6.2), (6.3) and the fact that the matrix $\Phi(H,G)-\phi(G)$ is the transpose of $\Phi(G,H)-\phi(G)$ we obtain

$$\frac{\partial}{\partial\alpha}\phi((1-\alpha)G+\alpha H)\Big|_{\alpha=0} = \int_{\mathcal{T}} \mathbf{d}(s,G)H^T(ds) + \left[\int_{\mathcal{T}} \mathbf{d}(s,G)H^T(ds)\right]^T . \tag{6.4}$$

This yields that if $d(s,G)=0_{m\times 1}$ for all $s\in\mathcal{T}$, then

$$\frac{\partial\phi((1-\alpha)G+\alpha H)}{\partial\alpha}\Big|_{\alpha=0} = 0_{m\times m}, \quad \forall\, H\in\mathcal{S}\,. \tag{6.5}$$

Also we have $\partial^2\phi((1-\alpha)G+\alpha H)/\partial\alpha^2 = 2\phi(G-H)$, which is a non-negative definite matrix for all $G,H\in\mathcal{S}$.

Let us assume that $G\in\mathcal{S}$ is such that $d(s,G)=0_{m\times 1}$ for all $s\in\mathcal{T}$, fix $H\in\mathcal{S}$ and a vector $c\in\mathbb{R}^m$. Consider a function $\psi_{c,H}(\alpha)=c^T\phi((1-\alpha)G+\alpha H)c$ as a function of $\alpha\in\mathbb{R}$. This is simply a quadratic and convex function of α, which, in view of (6.5), has zero derivative at $\alpha=0$. Therefore, for all $c\in\mathbb{R}^m$ and $H\in\mathcal{S}$ we have $\psi_{c,H}(0)=\min_\alpha\psi_{c,H}(\alpha)$, which is equivalent to the assertion that $\hat{\theta}_G$ is the BLUE.

II. Assume now that G gives the BLUE $\hat{\theta}_G$. This implies, first, that (6.5) holds and second, for all $c\in\mathbb{R}^m$ $c^T\phi(G)c\le c^T\phi(H)c$, for any $H\in\mathcal{S}$. Let us deduce that $d(s,G)=0_{m\times 1}$ for all $s\in\mathcal{T}$ (which is equivalent to validity of (2.4)). We are going to prove this by contradiction.

Assume that there exists $s_0\in\mathcal{T}$ such that $d(s_0,G)\neq 0$. Define the signed matrix-measure $\zeta=(\zeta_0,\zeta_1,\ldots,\zeta_q)$ with $\zeta_0(ds)=G_0(ds)+\kappa d(s_0,G)\delta_{s_0}(ds)$, $\kappa\neq 0$, and $\zeta_i(ds)=G_i(ds)$ for $i=1,\ldots,q$.

Since G is unbiased, $C_G=\int_{\mathcal{T}} G(dt)F^T(t)=I_m$. For any small positive or small negative κ, the matrix $C_\zeta=\int_{\mathcal{T}}\zeta(dt)F^T(t)=I_m+\kappa d(s_0,G)f^T(s_0)$ is non-degenerate and its eigenvalues are close to 1. In view of Lemma 2.1, $H(ds)=C_\zeta^{-1}\zeta(ds)$ is an unbiased matrix-measure. Using the identity (6.4) and Lemma 6.1 we obtain for the measure $G_\alpha=(1-\alpha)G+\alpha H$:

$$\frac{\partial\phi(G_\alpha)}{\partial\alpha}\Big|_{\alpha=0} = \kappa d(s_0,G)d^T(s_0,G)C_\zeta^{-1^T} + \kappa C_\zeta^{-1}d(s_0,G)d^T(s_0,G).$$

Write this as $\partial\phi(G_\alpha)/\partial\alpha\big|_{\alpha=0} = \kappa(X_0A^T + AX_0)$, where $A = C_\zeta^{-1}$ and $X_0 = d(s_0,G)d^T(s_0,G)$ is a symmetric matrix.
For any given A, the homogeneous Lyapunov matrix equation $XA^T+AX=0$ has only the trivial solution $X=0$ if and only if A and $-A$ have no common eigenvalues, see [§3, Ch. 8 in Gantmacher (1959)]; this is the case when $A = C_\zeta^{-1}$ and κ is small enough.
This yields that for $X = X_0$, the matrix $X_0A^T+AX_0$ is a non-zero symmetric matrix. Therefore, there exists a vector $c \in \mathbb{R}^m$ such that the directional derivative of $c^T\phi(G_\alpha)c$ is non-zero. For any such c, $c^T\phi(G_\alpha)c < c^T\phi(G)c$ for either small positive or small negative α and hence $\hat\theta_G$ is not the BLUE. Thus, the assumption of the existence of an $s_0 \in \mathcal{T}$ such that $d(s_0,G) \neq 0$ yields a contradiction to the fact that G gives the BLUE. This completes the proof that the equality (2.4) is necessary and sufficient for the estimator $\hat\theta_G$ to be the BLUE .

6.3. *Proof of Lemma 2.2.* We repeat i times the integration by parts

$$\int_{\mathcal{T}} \psi^{(i)}(t)\varphi(t)dt = \psi^{(i-1)}(t)\varphi(t)\Big|_A^B - \int_{\mathcal{T}} \psi^{(i-1)}(t)\varphi^{(1)}(t)dt$$

for differentiable functions $\psi(t)$ and $\varphi(t)$. This gives

$$\int_{\mathcal{T}} \psi^{(i)}(t)\varphi_i(t)dt = \sum_{j=1}^{i}(-1)^{j-1}\psi^{(i-j)}(t)\varphi_i{}^{(j-1)}(t)\Big|_A^B + (-1)^i\int_{\mathcal{T}} \psi(t)\varphi_i^{(i)}(t)dt.$$

Using the above equality with $\psi(t) = y^{(i)}(t)$ we obtain that the expectation of two estimators coincide. Also, using this equality with $\psi(t) = K^{(i)}(t,s)$ we obtain that the covariance matrices of the two estimators coincide.

6.4. *Proof of Proposition 2.2.* Straightforward calculus shows that

$$\begin{aligned}\int_{\mathcal{T}} K(t,s)\zeta(dt) &= K(A,s)z_A + K(B,s)z_B - \int_{\mathcal{T}} K(t,s)f^{(2)}(t)dt/(\lambda_1+\lambda_2)\\ &= K(A,s)z_A + K(B,s)z_B + \big[-K(t,s)f^{(1)}(t)|_A^s + K^{(1)}(t,s)f(t)|_A^{s-}\\ &\quad -K(t,s)f^{(1)}(t)|_s^B + K^{(1)}(t,s)f(t)|_{s+}^B\big]/(\lambda_1+\lambda_2) = (1+\lambda_1A-\lambda_2 s)z_A\\ &\quad +(1+\lambda_1 s-\lambda_2B)z_B + f(s) + \big[K(A,s)f^{(1)}(A) - K^{(1)}(A,s)f(A)\\ &\quad - K(B,s)f^{(1)}(B) + K^{(1)}(B,s)f(B)\big]/(\lambda_1+\lambda_2) = f(s).\end{aligned}$$

Therefore, the conditions of Corollary 2.1 are fulfilled.

6.5. *Proof of Theorem 3.1.* It is easy to see that $\hat{\theta}_{G_0,G_1}$ is unbiased. Further we are going to use Corollary 3.1 which gives the sufficient condition for an estimator to be the BLUE. We will show that the identity

$$LHS = \int_A^B K(t,s)\zeta_0(dt) + \int_A^B K^{(1)}(t,s)\zeta_1(dt) = f(s) \tag{6.6}$$

holds for all $s \in [A,B]$. By the definition of the measure ζ it follows that $LHS = z_A K(A,s) + z_B K(B,s) + I_A + I_B + z_{1,A}K^{(1)}(A,s) + z_{1,B}K^{(1)}(B,s)$, where $I_A = \int_A^s K(t,s)z(t)dt,\ I_B = \int_s^B K(t,s)z(t)dt$. Indeed, for the vector-function $z(t) = \tau_0 f(t) - \tau_2 f^{(2)}(t) + f^{(4)}(t)$, we have

$$\begin{aligned}
s_3 I_A &= \tau_0\int_A^s K(t,s)f(t)dt - \tau_2\int_A^s K(t,s)f^{(2)}(t)dt + \int_A^s K(t,s)f^{(4)}(t)dt \\
&= \tau_0\int_A^s K(t,s)f(t)dt - \tau_2 K(t,s)f^{(1)}(t)|_A^s + \tau_2 K^{(1)}(t,s)f(t)|_A^s \\
&\quad -\tau_2\int_A^s K^{(2)}(t,s)f(t)dt + K(t,s)f^{(3)}(t)|_A^s - K^{(1)}(t,s)f^{(2)}(t)|_A^{s-} \\
&\quad +K^{(2)}(t,s)f^{(1)}(t)|_A^{s-} - K^{(3)}(t,s)f(t)|_A^{s-} + \int_A^s K^{(4)}(t,s)f(t)dt.
\end{aligned}$$

By construction, the coefficients τ_0, τ_2, are chosen such that the equality (3.3) holds for all $t \in [A,B]$ and any s, implying that integrals in the expression for I_A are cancelled. Thus, we obtain

$$\begin{aligned}
s_3 I_A &= +\tau_2 K(A,s)f^{(1)}(A) - \tau_2 K^{(1)}(A,s)f(A) - K(A,s)f^{(3)}(A) \\
&\quad +K^{(1)}(A,s)f^{(2)}(A) - K^{(2)}(A,s)f^{(1)}(A) + K^{(3)}(A,s)f(A) \\
&\quad -\tau_2 K(s-,s)f^{(1)}(s) + \tau_2 K^{(1)}(s-,s)f(s) + K(s-,s)f^{(3)}(s) \\
&\quad -K^{(1)}(s-,s)f^{(2)}(s) + K^{(2)}(s-,s)f^{(1)}(s) - K^{(3)}(s-,s)f(s).
\end{aligned}$$

Similarly we have

$$\begin{aligned}
s_3 I_B &= -\tau_2 K(B,s)f^{(1)}(B) + \tau_2 K^{(1)}(B,s)f(B) + K(B,s)f^{(3)}(B) \\
&\quad -K^{(1)}(B,s)f^{(2)}(B) + K^{(2)}(B,s)f^{(1)}(B) - K^{(3)}(B,s)f(B) \\
&\quad +\tau_2 K(s+,s)f^{(1)}(s) - \tau_2 K^{(1)}(s+,s)f(s) - K(s+,s)f^{(3)}(s) \\
&\quad +K^{(1)}(s+,s)f^{(2)}(s) - K^{(2)}(s+,s)f^{(1)}(s) + K^{(3)}(s+,s)f(s).
\end{aligned}$$

Using the assumption on the derivatives of the kernel $K(t,s)$, we obtain

$$\begin{aligned}
s_3(I_A + I_B) = \tau_2 K(A,s)f^{(1)}(A) - \tau_2 K^{(1)}(A,s)f(A) - K(A,s)f^{(3)}(A) \\
+K^{(1)}(A,s)f^{(2)}(A) - K^{(2)}(A,s)f^{(1)}(A) + K^{(3)}(A,s)f(A) \\
-\tau_2 K(B,s)f^{(1)}(B) + \tau_2 K^{(1)}(B,s)f(B) + K(B,s)f^{(3)}(B) \\
-K^{(1)}(B,s)f^{(2)}(B) + K^{(2)}(B,s)f^{(1)}(B) - K^{(3)}(B,s)f(B) + s_3 f(s).
\end{aligned}$$

Also we have

$$\begin{aligned} s_3(z_A K(A,s) \;&+\; z_{1,A}K^{(1)}(A,s)) = \\ &= \big(f^{(3)}(A) - \gamma_{1,A}f^{(1)}(A) + \gamma_{0,A}f(A)\big)K(A,s) \\ &\quad +\big(-f^{(2)}(A) + \beta_{1,A}f^{(1)}(A) - \beta_{0,A}f(A)\big)K^{(1)}(A,s) \\ &= f^{(3)}(A)K(A,s) + (-\gamma_{1,A}K(A,s) + \beta_{1,A}K^{(1)}(A,s))f^{(1)}(A) \\ &\quad -K^{(1)}(A,s)f^{(2)}(A) + (\gamma_{0,A}K(A,s) - \beta_{0,A}K^{(1)}(A,s))f(A), \end{aligned}$$

and a similar result at the point $t = B$. Putting these expressions into (6.6) and using the assumption that constants $\gamma_{1,A}, \beta_{1,A}, \gamma_{0,A},\ \beta_{0,A}$ and $\gamma_{1,B}, \beta_{1,B}, \gamma_{0,B}, \beta_{0,B}$ are chosen such that the sum of the functions defined in (3.2) is identically equal to zero, we obtain

$$\int_A^B K(t,s)\zeta_0(dt) + \int_A^B K^{(1)}(t,s)\zeta_1(dt) = f(s);$$

this completes the proof.

6.6. *Proof of Theorem 3.2.* Observing (3.9), we write the vector c as

$$\begin{aligned} c \;&=\; \int_a^A \left[\int_A^B R(t,s)\zeta_0(dt) - f(s)\right] ds \\ &=\; \int_a^A \left[\int_A^B R(t,s')\zeta_0(dt) - f(s')\right] ds' + \int_A^s \left[\int_A^B R(t,s')\zeta_0(dt) - f(s')\right] ds' \\ &=\; \int_A^B \int_a^s R(t,s')ds'\zeta_0(dt) - \int_a^s f(s')ds' = \int_A^B K^{(1)}(t,s)\zeta_0(dt) - \tilde{f}(s). \end{aligned}$$

We now show that equation (2.5) in Corollary 2.1 holds for $q = 1, f = \tilde{f}$ and $\zeta_i = \tilde{\zeta}_i$. Observing (3.12) and the definition of $\tilde{\zeta}_i$ in Theorem 3.2 we obtain

$$\begin{aligned} &\int_{\mathcal{T}} K(t,s)\tilde{\zeta}_0(dt) + \int_{\mathcal{T}} K^{(1)}(t,s)\tilde{\zeta}_1(dt) \\ &= -c\left(\int_{\mathcal{T}} K(t,s)\eta_0(dt) + \int_{\mathcal{T}} K^{(1)}(t,s)\eta_1(dt)\right) + \int_{\mathcal{T}} K^{(1)}(t,s)\zeta_0(dt) \\ &= -c \cdot 1 + \widetilde{f}(s) + c = \widetilde{f}(s). \end{aligned}$$

6.7. *Proof of Propositions 3.3 and 3.4.* For the sake of brevity we only give a proof of Proposition 3.3, the other result follows by similar arguments. Direct calculus gives $s_3 = K^{(3)}(s+,s) - K^{(3)}(s-,s) = 2\lambda_1\lambda_2(\lambda_1+\lambda_2)$. Then we obtain that the identity (3.3) holds for $\tau_0 = \lambda_1^2\lambda_2^2$ and $\tau_2 = \lambda_1^2 + \lambda_2^2$. Straightforward calculations show that identities (3.2) hold with the specified values of constants $\gamma_1, \gamma_0, \beta_1, \beta_0$.

Acknowledgements. We are extremely grateful to the associate editor and also to two referees for their valuable and constructive comments, which led to substantial improvement of an earlier version of this manuscript. This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Project C2) of the German Research Foundation (DFG). The work of Pepelyshev was partially supported by the SPbSU project No. 6.38.435.2015 and the RFFI project No. 17-01-00161.

References.

Anderes, E. (2010). On the consistent separation of scale and variance for gaussian random fields. *The Annals of Statistics*, 38(2):870–893.

Anderson, T. W. (1970). Efficient estimation of regression coefficients in time series. Technical report, DTIC Document.

Berlinet, A. and Thomas-Agnan, C. (2011). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer.

Brockwell, P. (2001). Continuous-time ARMA processes. *Handbook of Statistics*, 19:249–276.

Brockwell, P., Davis, R., and Yang, Y. (2007). Continuous-time Gaussian autoregression. *Statistica Sinica*, 17(1):63.

Dette, H., Konstantinou, M., and Zhigljavsky, A. (2017). A new approach to optimal designs for correlated observations. *The Annals of Statistics*, 45(4):1579–1608.

Dette, H., Pepelyshev, A., and Zhigljavsky, A. (2013). Optimal design for linear models with correlated observations. *The Annals of Statistics*, 41(1):143–176.

Dette, H., Pepelyshev, A., and Zhigljavsky, A. (2016). Optimal designs in regression with correlated errors. *The Annals of Statistics*, 44(1):113–152.

Dette, H., Pepelyshev, A., and Zhigljavsky, A. (2018). Supplement to "the blue in continuous-time regression models with correlated errors". *Annals of Statistics*.

Fedorov, V. V. and Müller, W. G. (2007). Optimum design for correlated fields via covariance kernel expansions. In *mODa 8-Advances in Model-Oriented Design and Analysis*, pages 57–66. Springer.

Freeden, W. (1999). *Multiscale modelling of spaceborne geodata*. Teubner.

Gantmacher, F. R. (1959). *The Theory of Matrices*. Chelsea, New York.

Grenander, U. (1950). Stochastic processes and statistical inference. *Ark. Mat.*, 1:195–277.

Grenander, U. (1954). On the estimation of regression coefficients in the case of an autocorrelated disturbance. *The Annals of Mathematical Statistics*, 25(2):252–272.

Hájek, J. (1956). Linear estimation of the mean value of a stationary random process with convex correlation function. *Czechoslovak Mathematical Journal*, 6(81):94–117.

Hannan, E. (1975). Linear regression in continuous time. *Journal of the Australian Mathematical Society (Series A)*, 19(02):146–159.

He, S. and Wang, J. (1989). On embedding a discrete-parameter ARMA model in a continuous-parameter ARMA model. *Journal of Time Series Analysis*, 10(4):315–323.

Kanwal, R. (1997). *Linear Integral Equations*. Birkhauser, Boston.

Kholevo, A. (1969). On estimates of regression coefficients. *Theory of Probability & Its Applications*, 14(1):79–104.

Loh, W.-L., Lam, T.-K., et al. (2000). Estimating structured correlation matrices in smooth gaussian random field models. *The Annals of Statistics*, 28(3):880–904.

Morris, M. D., Mitchell, T. J., and Ylvisaker, D. (1993). Bayesian design and analysis of computer experiments: use of derivatives in surface prediction. *Technometrics*, 35(3):243–255.

Müller, W. G. and Pázman, A. (2003). Measures for designs in experiments with correlated errors. *Biometrika*, 90(2):423–434.

Näther, W. (1985). *Effective Observation of Random Fields.* Teubner Verlagsgesellschaft, Leipzig.

Osborne, M. A., Garnett, R., and Roberts, S. J. (2009). Gaussian processes for global optimization. In *3rd international conference on learning and intelligent optimization (LION3)*, pages 1–15. Citeseer.

Parzen, E. (1961). An approach to time series analysis. *The Annals of Mathematical Statistics*, 32(4):951–989.

Pronzato, L. and Müller, W. G. (2012). Design of computer experiments: space filling and beyond. *Statistics and Computing*, 22(3):681–701.

Ramm, A. G. and Charlot, J. (1980). *Theory and Applications of Some New Classes of Integral Equations.* Springer.

Rasmussen, C. and Williams, C. (2006). *Gaussian processes for machine learning.* MIT Press.

Ritter, K. (2000). *Average-Case Analysis of Numerical Problems.* Springer.

Ritter, K., Wasilkowski, G. W., and Woźniakowski, H. (1995). Multivariate integration and approximation for random fields satisfying Sacks-Ylvisaker conditions. *The Annals of Applied Probability*, 5(2):518–540.

Rosenblatt, M. (1956). Some regression problems in time series analysis. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pages 165–186. University of California.

Sacks, J. and Ylvisaker, D. (1970). Designs for regression problems with correlated errors III. *The Annals of Mathematical Statistics*, 41(6):2057–2074.

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

Sacks, J. and Ylvisaker, N. D. (1968). Designs for regression problems with correlated errors; many parameters. *Annals of Mathematical Statistics*, 39:49–69.

Sejdinovic, D., Sriperumbudur, B., Gretton, A., and Fukumizu, K. (2013). Equivalence of distance-based and rkhs-based statistics in hypothesis testing. *The Annals of Statistics*, 41(5):2263–2291.

Stein, M. L. (2012). *Interpolation of spatial data: some theory for kriging.* Springer Science & Business Media.

Székely, G. J. and Rizzo, M. L. (2013). Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference*, 143(8):1249–1272.

Yaglom, A. M. (1987). *Correlation Theory of Stationary and Related Random Functions. Volume I: Basic Results.* Springer Science & Business Media.

Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany
E-mail: holger.dette@rub.de

School of Mathematics
Cardiff University
Cardiff, CF24 4AG
UK
E-mail: pepelyshevan@cf.ac.uk
E-mail: zhigljavskyaa@cf.ac.uk