SFB 823

Discussion Paper

Best linear unbiased estimators in continuous time regression models

Holger Dette, Andrey Pepelyhev,
Anatoly Zhigljavsky

Nr. 58/2016

SFB 823

Submitted to the Annals of Statistics

BEST LINEAR UNBIASED ESTIMATORS IN CONTINUOUS TIME REGRESSION MODELS

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 by several examples.

1. Introduction. Consider a continuous-time linear regression model of the form

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

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 defined on some interval, say $[A, B]$, 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.
This paper is devoted to studying the best linear unbiased estimator (BLUE) of the parameter θ in the general setting and in many specific instances. Understanding of the explicit form of the BLUE has profound 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

Primary 62J05; secondary 31A10

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

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 (6.7) can be derived from the continuous BLUE, like the extended signed least squares estimator investigated in Dette et al. (2016a) and the estimators based on approximation of stochastic integrals proposed in Dette et al. (2017).

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 (6.8) with $q = 0$. In this situation, it is well understood that solving specific instances of an equation of Wiener-Hopf type

$$\int_A^B K(t,s)\zeta(dt) = f(s), \tag{1.3}$$

for an m-dimensional vector ζ of signed measures implies an explicit construction of the BLUE in the continuous-time model (6.8). 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 Pisarenko and Rozanov (1963); Kholevo (1969); 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 Hajek (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 ($q \ge 0$); 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 main reason why results on the BLUE in model (6.8) are so difficult to obtain consists in the fact that - except in the location-scale model - the functional to be minimized is not convex, so that the usual arguments are not applicable. The main 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 this book. In most of these examples either a Markovian structure of the error process is assumed or 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 theory of the RKHS. 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 2.4). 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 (6.8), 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 which can be used to determine its explicit form and the corresponding covariance matrix in numerous models. In particular, we do not have to restrict ourselves to one-dimensional regression models and to Markovian error processes. Thus our results require minimal assumptions regarding the regression function and the error process. Important new examples, where the BLUE can be determined explicitly, include the process with triangular covariance function (2.7), 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. We are able to circumvent the convexity problem 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 an online supplement [see Dette et al. (2016b)] we 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 discrete regression models (6.7) 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 (6.8) with covariance kernel $K(t,s) = \mathbb{E}[\epsilon(t)\epsilon(s)]$. Suppose that we can observe the process $\{y(t)|t \in \mathcal{T}\}$ along with its $q \geq 0$ mean square derivatives $\{y^{(i)}(t)|t \in \mathcal{T}\}$ for $i = 1,\dots,q$, where the design 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 $f(t) = (f_1(t),\dots,f_m(t))^T$ is q times differentiable on the interval $[A,B]$ with derivatives $f^{(1)},\dots f^{(q)}$ ($f^{(0)} = f$). Throughout this paper we will also assume that the functions $f_1,\dots,f_m$ are linearly independent on $\mathcal{T}$.
Let $Y(t) = \{(y^{(0)}(t),\dots,y^{(q)}(t))^T\}$ 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 (6.8) 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 $\zeta_0, \ldots, \zeta_q$ 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 $G = (G_0, \ldots, G_q)$, where G_i are the signed vector-measures and $G_i(dt) = C^{-1}\zeta_i(dt)$ for $i = 0, \ldots, q$. Then the estimator $\hat{\theta}_G$ is unbiased.

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

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

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.* If there exists a set of signed vector-measures, say $G = (G_0, \ldots, G_q)$, such that the estimator $\hat{\theta}_G = \int_{\mathcal{T}} G(dt)Y(t)$ is unbiased and $\mathrm{Var}(\hat{\theta}_H) \geq \mathrm{Var}(\hat{\theta}_G)$, 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 (6.8) using the set of observations $\mathbf{Y}_{\mathcal{T}}$. The BLUE depends on the kernel K, the vector-function f, the set $\mathcal{T}$ where the observations are taken and on the number q of available derivatives of the process $\{y(t) | t \in \mathcal{T}\}$.
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 (6.8), where the error process* $\{\epsilon(t) | t \in [A,B]\}$ *has a covariance kernel* $K(\cdot,\cdot) \in C^q([A,B] \times [A,B])$ *and* $f(\cdot) \in C^q([A,B])$ *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* $t \in \mathcal{T} \subseteq [A,B]$.
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).*

The next proposition is slightly weaker than Theorem 2.1 (here the covariance matrix of the BLUE is assumed to be non-degenerate) but will be very useful in further considerations.

PROPOSITION 2.1. *Let the assumptions of Theorem 2.1 be satisfied and let* $\zeta_0, \ldots, \zeta_q$ *be signed vector-measures defined on* $\mathcal{T}$ *such that the matrix* C *defined in (2.2) is non-degenerate. Define* $G = (G_0, \ldots, G_q)$, $G_i(dt) = C^{-1}\zeta_i(dt)$ *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* $\mathrm{Var}(\hat{\theta}_G) = C^{-1}$.

2.3. *Grenander's theorem and its generalizations.*
When $\mathcal{T} = [A,B]$, $q = 0$, $m = 1$ and the regression model (6.8) 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]. The reason why

Grenander's and Näther's theorems only deal with the location-scale model is caused by the fact that for this model the convexity of the functional to be minimized is easy to establish. For general regression models the convexity argument is not directly applicable and hence the problem is much harder. For the case of one-dimensional processes, Theorem 2.1 generalizes Grenander's theorem to arbitrary m-parameter regression models of the form (6.8) and the case of arbitrary $q \geq 0$. Another generalization of the Grenander's theorem is given below; it deals with a general m-parameter regression model (6.8) 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. Note that the conditions on the vector of regression functions $f(\cdot)$ in Theorem 2.2 are weaker (when $d = 1$) than the conditions on $f(\cdot)$ in Theorem 2.1 applied in the case $q = 0$.

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 $\hat{\theta}_G$ defined by (2.3).

The proof of this theorem is a simple extension of the proof of Theorem 2.1 with $q = 0$ to general $\mathcal{T} \subset \mathbb{R}^d$ 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 (6.8) 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 (6.8) 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 (6.8), 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 $H = (H_0, H_1, \dots, H_q)$ with vector-measures $H_1, \dots, H_q$ having no continuous parts.

(P5) Let $\zeta_0, \dots, \zeta_q$ 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 for all $s \in \mathcal{T}$, signed vector-measures $\eta_0, \dots, \eta_q$ satisfy the equation

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

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 $G = (G_0, \dots, G_q)$ by $G_i = C^{-1}(\zeta_i + \eta_i)$, $i = 0, \dots, q$, 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 Section 2.5. To prove (P5) we simply add the equations (2.5) and (2.6) and then use Proposition 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 (6.8) and especially for constructing a BLUE for new models from the cases when a BLUE is known for simpler models. As an example, assume that all functions in the vector f are not constant and set $g(t) = c$, where c is a constant vector. Then, if we know the BLUE for f and another BLUE for the location-scale model, we can use property (P5) to construct BLUE for $\theta^T(f(t) + c)$. This is an essential part of the proof of Theorem 3.2, which allows obtaining the explicit form of the BLUE for the integrated error processes from the explicit form of the BLUE for the corresponding non-integrated errors (which is a much easier problem).

2.5. *Non-uniqueness.* Let us show that if $\mathcal{T} = [A, B]$ then, under the additional smoothness conditions, for a given set of signed vector-measures

$G = (G_0, G_1, \dots, G_q)$ on $\mathcal{T}$ we can find another set of measures $H = (H_0, H_1, \dots, H_q)$ such that the signed vector-measures $H_1, \dots, H_q$ 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 $q > 0$, $G_0, \dots, G_q$ be some signed vector-measures and for some $i \in \{1, \dots, m\}$, the signed measure $G_i(dt)$ has the form

$$G_i(dt) = Q_i(dt) + \varphi(t)dt,$$

where $Q_i(dt)$ is a signed vector-measure and $\varphi \in C^i([A, B])$ (that is, φ is an i times differentiable vector-function on the interval $[A, B]$). Define the matrix $H = (H_0, \dots, H_q)$, where the columns of H are the following signed vector-measures:

$$H_0(dt) = G_0(dt) + (-1)^i \left[\varphi^{(i)}(t)dt - \varphi^{(i-1)}(A)\delta_A(dt) + \varphi^{(i-1)}(B)\delta_B(dt)\right],$$

$H_i(dt) = Q_i(dt)$, $H_j(dt) = G_j(dt)$, for $j = i+1, \dots, q$ and

$$H_j(dt) = G_j(dt) + (-1)^{i-j-1} \left[\varphi^{(i-j-1)}(A)\delta_A(dt) - \varphi^{(i-j-1)}(B)\delta_B(dt)\right]$$

for $j = 1, \dots, i-1$, where $\delta_A(dt)$ and $\delta_B(dt)$ are the Dirac delta-measures concentrated at the points A and B, respectively. 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.*

By Lemma 2.2 we can restrict the search of linear unbiased estimators to estimators $\hat{\theta}_G$ of the form (2.1), where the components $G_1, \dots, G_q$ of the signed matrix-measure $G = (G_0, \dots, G_q)$ have no continuous parts.

2.6. *Several examples of the BLUE for non-differentiable error processes.* For the sake of completeness we first consider the case when the errors in model (6.8) follow a Markov process; this includes the case of continuous autoregressive errors of order 1. In presenting these results we follow Näther (1985) and Dette et al. (2016a).

PROPOSITION 2.2. *Consider the regression model* (6.8) *with covariance kernel* $K(t, s) = u(t)v(s)$ *for* $t \le s$ *and* $K(t, s) = v(t)u(s)$ *for* $t > s$, *where* $u(\cdot)$ *and* $v(\cdot)$ *are 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

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

where 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 estimate $\hat\theta_G$ 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 in a special case of the covariance kernel $K(t,s)$ such that $K(t,s) \neq u(t)v(s)$. This statement provides the first example where an explicit form of the BLUE and its covariance matrix can be obtained for a non-Markovian error process. The proof is given in Section 6.

PROPOSITION 2.3. *Consider the regression model (6.8) on the interval $\mathcal{T} = [A,B]$ with errors having the covariance function $K(t,s) = 1 + \lambda_1 t - \lambda_2 s$, where $t \le s$, $\lambda_1 \ge \lambda_2$ 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 2.3 then we obtain the following case when the kernel is

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

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 (6.8) on the interval $\mathcal{T} = [0,1]$ with errors having the covariance kernel (2.7) with $\lambda \le 1$. Define the signed vector-measure

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

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}.

Consider now the case when the regression functions are linear combinations of eigenfunctions from the 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 $\mathcal{T} = [A, B]$ and ν be a measure on the corresponding Borel field with positive density. Consider the integral operator

$$T_K(h)(\cdot) = \int_A^B K(t,\cdot)h(t)\nu(dt) \tag{2.8}$$

on $L_2(\nu, [A, B])$, 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 (with respect to $\nu(dt)$) eigenfunctions $\phi_1, \phi_2, \dots$ with positive eigenvalues $\lambda_1, \lambda_2, \dots$ of the integral operator T_K. The next statement follows directly from Proposition 2.1.

PROPOSITION 2.4. *Let $\phi_1, \phi_2, \dots$ be the eigenfunctions of the integral operator (2.8) and $f(t) = \sum_{\ell=1}^\infty q_\ell \phi_\ell(t)$ with some sequence $\{q_\ell\}_{\ell\in\mathbb{N}}$ in $\mathbb{R}^m$ such that $f_1(t), \dots, f_m(x)$ are linearly independent. Then the estimator $\hat{\theta}_G$ with $G(dt) = C^{-1}\sum_{\ell=1}^\infty \lambda_\ell^{-1} q_\ell \phi_\ell(t)\nu(dt)$ and $C = \sum_{\ell=1}^\infty \lambda_\ell^{-1} q_\ell q_\ell^T$ is a BLUE with covariance matrix C^{-1}.*

Proposition 2.4 provides a way of constructing the 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 (6.8) 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.

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 (6.8) 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 Proposition 2.1 when $q = 1$.

COROLLARY 3.1. *Consider the regression model (6.8) 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. 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 $\tau_0, \tau_2, \gamma_{0,A}, \gamma_{1,A}, \beta_{0,A}, \beta_{1,A}, \gamma_{0,B}, \gamma_{1,B}, \beta_{0,B}, \beta_{1,B}, s_3$ are some constants and $s_3 = K^{(3)}(s-,s) - K^{(3)}(s+,s)$. 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* (6.8) *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* $K(t, s)$ *is also four times differentiable for* $t \neq s$ *such that*

$$\begin{aligned} K^{(i)}(s-, s) &- K^{(i)}(s+, s) = 0, \quad i = 0, 1, 2, \\ K^{(3)}(s-, s) &- K^{(3)}(s+, s) \neq 0. \end{aligned}$$

With the notation of the previous paragraph define the vector-measures

$$\begin{aligned} \zeta_0(dt) &= z_A \delta_A(dt) + z_B \delta_B(dt) + z(t)dt, \\ \zeta_1(dt) &= z_{1,A} \delta_A(dt) + z_{1,B} \delta_B(dt). \end{aligned}$$

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 $t, s \in [A, B]$, *(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 $0 \le a \le A$.

PROPOSITION 3.1. *Consider the regression model* (6.8) *with integrated covariance kernel given by* (3.4) *and suppose that* f *is four times differentiable on the interval* $[A, B]$. *Define the signed vector-measures*

$$\begin{aligned} \zeta_0(dt) &= z_A \delta_A(dt) + z_B \delta_B(dt) + z(t)dt, \\ \zeta_1(dt) &= z_{1,A} \delta_A(dt) + z_{1,B} \delta_B(dt), \end{aligned}$$

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 (6.8) on $\mathcal{T} = [A, B]$ with integrated covariance kernel

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

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 $z(t) = f^{(4)}(t)$,

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

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}.

Consider now the integrated triangular-shape kernel

$$\begin{aligned} K(t,s) &= \int_0^t \int_0^s \max\{0, 1 - \lambda|t' - s'|\} dt' ds' \\ &= ts - \lambda \min(t,s)\Big(3\max(t,s)^2 - 3ts + 2\min(t,s)^2\Big)/6. \end{aligned} \tag{3.6}$$

PROPOSITION 3.2. *Consider the regression model* (6.8) *on* $\mathcal{T} = [A, B]$ *with integrated covariance kernel* (3.6)*, where* $\lambda(B - A) < 1$*. Suppose that* f *is* *four times differentiable. Define the signed vector-measures*

$$\begin{aligned}\zeta_0(dt) &= z_A\delta_A(dt) + z_B\delta_B(dt) + z(t)dt,\\ \zeta_1(dt) &= z_{1,A}\delta_A(dt) + z_{1,B}\delta_B(dt),\end{aligned}$$

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 (3.9). Note that this extends the class of kernels considered in Sacks and Ylvisaker (1970) in a nontrivial way.

Consider the regression model (6.8) with a non-differentiable error process with covariance kernel $K(t,s)$ and BLUE

$$\hat\theta_{G_0} = \int_{\mathcal{T}} y(t)G_0(dt).$$

From Proposition 2.1 we have for the vector-measure $\zeta_0(dt)$ satisfying (2.5) and defining the BLUE

$$\int_A^B K(t,s)\zeta_0(dt) = f(s) \tag{3.7}$$

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

$$\tilde{y}(t)=\theta^T\tilde{f}(t)+\tilde{\varepsilon}(t), \tag{3.8}$$

which has the integrated covariance kernel

$$R(t,s)=\int_a^t\int_a^s K(u,v)dudv. \tag{3.9}$$

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

THEOREM 3.2. *Let the vector-measure ζ_0 satisfy the equality (3.7) and define the BLUE $\hat{\theta}_{G_0}$ with $G_0(dt)=C^{-1}\zeta_0(dt)$ in the regression model (6.8) with covariance kernel $K(\cdot,\cdot)$. Let the measures η_0,η_1 satisfy the equality*

$$\int_{\mathcal{T}} R(t,s)\eta_0(dt)+\int_{\mathcal{T}} R^{(1)}(t,s)\eta_1(dt)=1 \tag{3.10}$$

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 $c=\int_a^A[\int_A^B K(t,s)\zeta_0(dt)-f(s)]ds$. 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 (3.8) with integrated covariance kernel (3.9).

Repeated application of Theorem 3.2 extends the results to the case of several times integrated processes.

If $a=A$ in (3.9) we have $c=0$ in Theorem 3.2 and in this case, the statement of Theorem 3.2 can be proved easily. Moreover, in this case the class of kernels defined by (3.9) 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 of kernels considered in Sacks and Ylvisaker (1970).

3.4. *BLUE for AR(2) errors.* Consider the continuous-time regression model (6.8), 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

$$d\varepsilon^{(1)}(t) = \tilde{a}_1 \varepsilon^{(1)}(t) + \tilde{a}_2 \varepsilon(t) + \sigma_0^2 dW(t), \tag{3.11}$$

where $\mathrm{Var}(\varepsilon(t)) = \sigma^2$ 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 (3.11) 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 (3.11) 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

$$\rho_1(t) = \frac{\lambda_2}{\lambda_2 - \lambda_1} e^{-\lambda_1 |t|} - \frac{\lambda_1}{\lambda_2 - \lambda_1} e^{-\lambda_2 |t|}\,, \tag{3.12}$$

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

$$\rho_2(t) = e^{-\lambda |t|} \Big\{ \cos(\omega |t|) + \frac{\lambda}{\omega} \sin(\omega |t|) \Big\}\,, \tag{3.13}$$

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

$$\rho_3(t) = e^{-\lambda |t|} (1 + \lambda |t|)\,, \tag{3.14}$$

where $\lambda > 0$. Note that the kernel (6.13) 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 (6.8) with CAR(2) errors, where the covariance kernel $K(t, s) = \rho(t - s)$ has the form (6.11). 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* (6.8) *with CAR(2) errors, where the covariance kernel* $K(t,s) = \rho(t-s)$ *has the form* (6.12). *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 (6.13) 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. (2016b) we consider the regression model (6.7) 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* (6.8) *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

$$\begin{aligned}\zeta_0(dt) &= z_A\delta_A(dt) + z_B\delta_B(dt) + z(t)dt,\\ \zeta_1(dt) &= z_{1,A}\delta_A(dt) + z_{1,B}\delta_B(dt),\\ \zeta_2(dt) &= z_{2,A}\delta_A(dt) + z_{2,B}\delta_B(dt),\end{aligned}$$

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 (6.8), 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

$$\begin{aligned}z_{j,A} &= \sum\nolimits_{l=0}^{2p-j-1} \gamma_{l,j,A} f^{(j)}(A)/s_{2p-1},\\ z_{j,B} &= \sum\nolimits_{l=0}^{2p-j-1} \gamma_{l,j,B} f^{(j)}(B)/s_{2p-1}\end{aligned}$$

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 (6.8) with CAR(p) errors. Define the vector-measures*

$$\begin{aligned}\zeta_0(dt) &= z_{0,A}\delta_A(dt) + z_{0,B}\delta_B(dt) + z(t)dt, \\ \zeta_j(dt) &= z_{j,A}\delta_A(dt) + z_{j,B}\delta_B(dt),\ j = 1, \dots, p-1,\end{aligned}$$

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

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

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

Let us consider the construction of a BLUE for model (6.8) 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 $\tilde{a}(z) = z^3 + \tilde{a}_1 z^2 + \tilde{a}_2 z + \tilde{a}_3$,

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

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

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

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 (6.8) 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 $y(t_1), \ldots, y(t_N)$, $y'(t_1), \ldots, y'(t_N)$ with $t_i = 1 + (i-1)/(N-1)$, $i = 1, \ldots, N$. 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 $y(t_1), \dots, y(t_{2N})$ with $t_i = 1 + (i-1)/(2N-1)$, $i = 1, \dots, 2N$. This estimator uses $2N$ observations of the original process (at equidistant points) and no observations from its derivative.

In Table 1 – 3 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 drift in model (6.8) defined by

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

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

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

TABLE 1

The efficiency defined by (5.1) for four different estimators based on $2N$ observations. The drift function is given by (5.2)

$\tilde{\theta}$	$N = 3$	$N = 5$	$N = 10$
$\hat{\theta}_{BLUE}(N, N)$	1	1	1
$\hat{\theta}_{BLUE}(2N-2, 2)$	1	1	1
$\hat{\theta}_{BLUE}(2N, 0)$	0.8593	0.9147	0.9570
$\hat{\theta}_{OLSE}(2N, 0)$	0.0732	0.0733	0.0734

TABLE 2

The efficiency defined by (5.1) for four different estimators based on $2N$ observations. The drift function is given by (5.3)

$\tilde{\theta}$	$N = 3$	$N = 5$	$N = 10$
$\hat{\theta}_{BLUE}(N, N)$	0.41246	0.92907	0.99680
$\hat{\theta}_{BLUE}(2N-2, 2)$	0.45573	0.98706	0.99972
$\hat{\theta}_{BLUE}(2N, 0)$	0.47796	0.77195	0.89641
$\hat{\theta}_{OLSE}(2N, 0)$	0.00113	0.00137	0.00218

The results of Table 1 – 3 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 very efficient

TABLE 3
The efficiency defined by (5.1) for four different estimators based on $2N$ observations. The drift function is given by (5.4)

$\tilde{\theta}$	$N = 3$	$N = 5$	$N = 10$
$\hat{\theta}_{BLUE}(N, N)$	0.69608	0.95988	0.99791
$\hat{\theta}_{BLUE}(2N - 2, 2)$	0.86903	0.99379	0.99981
$\hat{\theta}_{BLUE}(2N, 0)$	0.10040	0.33338	0.62529
$\hat{\theta}_{OLSE}(2N, 0)$	0.08873	0.14103	0.11890

when N is large. Moreover, the use of information from the derivatives in constructing BLUEs makes them more efficient than the BLUE which only uses values of $\{y(t)|t \in \mathcal{T}\}$. 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 $y'(B)$. 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 $\{y(t)|t \in \mathcal{T}\}$.

6. Appendix.

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

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

This implies that the estimator $\hat{\theta}_G$ is unbiased if and only if

$$\sum_{i=0}^{q} \int_{\mathcal{T}} f^{(i)}(t) G_i^T(dt) = I_m. \tag{6.1}$$

Since $G_i = C^{-1}\zeta_i$, we have

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

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. The set of all unbiased signed matrix-measures will be denoted by $\mathcal{S}$. This set is convex.

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 a $m \times (q + 1)$ matrix.

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

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

where $F(s) = (f(s), f^{(1)}(s), \ldots, f^{(q)}(s))$. □

For any two 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 $\phi(G) = \Phi(G, G)$ for any $G \in \mathcal{S}$. For any two matrix-measures G and H in $\mathcal{S}$ and any real α, the matrix-valued function

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

is quadratic in α. Also we have $\partial^2 \phi((1-\alpha)G + \alpha H)/\partial\alpha^2 = 2\phi(G - H) \geq 0$ and hence $\phi(\cdot)$ is a matrix-convex function on the space $\mathcal{S}$ (see e.g. Hansen and Tomiyama (2007) for properties of matrix-convex functions).

Since the matrix-function $\phi((1-\alpha)G+\alpha H)$ is quadratic and convex in $\alpha \in \mathbb{R}$, the assertion that G is the optimal matrix measure minimizing $\phi(\cdot)$ on $\mathcal{S}$, is equivalent to

$$\frac{\partial \phi((1-\alpha)G+\alpha H)}{\partial \alpha}\Big|_{\alpha=0} = 0, \;\; \forall\, H \in \mathcal{S}\,. \tag{6.2}$$

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

$$\frac{\partial}{\partial \alpha}\phi((1-\alpha)G+\alpha H)\Big|_{\alpha=0} = \Phi(G,H)+\Phi(H,G)-2\phi(G). \tag{6.3}$$

To rewrite (6.3), we note that $\int_{\mathcal{T}} \mathbf{d}(s,G)H^T(ds)$ can be written as

$$\begin{aligned} \int_{\mathcal{T}} \mathbf{d}(s,G)H^T(ds) &= \Phi(G,H) - \phi(G)\int_{\mathcal{T}} F(s)H^T(ds) \\ &= \Phi(G,H)-\phi(G), \end{aligned} \tag{6.4}$$

where in the last equality we have used the unbiasedness condition (6.1) for H. Using (6.3), (6.4) and the fact that the matrix $\Phi(H,G)-\phi(G)$ is the transpose of $\Phi(G,H)-\phi(G)$ we obtain

$$\frac{\partial}{\partial \alpha}\phi((1-\alpha)G+\alpha H)\Big|_{\alpha=0} = \int_{\mathcal{T}} \mathbf{d}(s,G)H^T(ds) + \left[\int_{\mathcal{T}} \mathbf{d}(s,G)H^T(ds)\right]^T . \tag{6.5}$$

Consequently, if $d(s,G)=0_{m\times 1}$ for all $s \in \mathcal{T}$, then (6.2) holds and hence G gives the BLUE.

II. Assume now that G gives the BLUE $\hat{\theta}_G$. This implies, first, that (6.2) 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 (6.5) 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 $\partial\phi(G_\alpha)/\partial\alpha\big|_{\alpha=0} = \kappa(X_0 A^T + AX_0)$, 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 $X_0 A^T + AX_0$ 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 formula

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

for any differentiable function $\psi(t)$. This gives

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

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

6.4. *Proof of Proposition 2.3.* Straightforward calculus shows that

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

Therefore, the conditions of Proposition 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 covariance kernel $K(t,s)$, we obtain

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

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 (3.7) the vector c is can be written as

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

We now show that equation (2.5) in Proposition 2.1 holds for $K = R, q = 1, f = \tilde{f}$ and $\zeta_i = \tilde{\zeta}_i$. Observing (3.10) and the definition of $\tilde{\zeta}_i$ in Theorem 3.2

we obtain

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

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. 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) and the National Institute Of General Medical Sciences of the National Institutes of Health under Award Number R01GM107639. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The work of Andrey Pepelyshev was partly supported by the project "Actual problems of design and analysis for regression models" (6.38.435.2015) of St. Petersburg State University. The authors would like to thank Martina Stein, who typed parts of this manuscript with considerable technical expertise.

References.

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 simple approach to optimal designs for correlated observations. *The Annals of Statistics*. To appear.

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. (2016a). Optimal designs in regression with correlated errors. *The Annals of Statistics*, 44(1):113–152.

Dette, H., Pepelyshev, A., and Zhigljavsky, A. (2016b). Supplement to "best linear unbiased estimators in continuous time regression models". *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.

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.

Hajek, 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.

Hansen, F. and Tomiyama, J. (2007). Differential analysis of matrix convex functions. *Linear Algebra and its Applications*, 420(1):102–116.

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.

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.

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

Pisarenko, V. F. and Rozanov, J. A. (1963). On certain problems for stationary processes leading to integral equations related to Wiener-Hopf equations. *Problemy Peredaci Informacii*, 14:113–135.

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.

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, 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.

Sacks, J. and Ylvisaker, N. D. (1970). Statistical designs and integral approximation. *Proc. 12th Bienn. Semin. Can. Math. Congr.(R. Pyke, ed.)*, pages 115–136.

Submitted to the Annals of Statistics

SUPPLEMENT TO "BEST LINEAR UNBIASED ESTIMATORS IN CONTINUOUS TIME REGRESSION MODELS"

By Holger Dette[‡], Andrey Pepelyshev[§] and Anatoly Zhigljavsky[§]

Ruhr-Universität Bochum[‡] and Cardiff University[§]

We demonstrate that the covariance matrix of the BLUE in the continuous-time regression model model with a CAR(2) error process can be obtained as limit of the covariance matrix of a BLUE in the discrete regression model with observations at equidistant points and a discrete AR(2) error process.

Here we investigate the approximation of the BLUE for continuous-time regression time models with a CAR(2) error process (see Section 4 of the paper) by the BLUE in the model

$$y(t_i) = \theta^T f(t_i) + \epsilon(t_i)\,, \quad A \le t_1 < t_2 \ldots < t_{N-1} < t_N \le B\,, \tag{6.7}$$

where the errors follow a discrete AR(2) process. This model will be abbreviated as DAR(2) throughout this section . The main difficulty to establish the connection between the discrete and continuous AR(2) cased lies in the fact that the discretised CAR(2) process follows an ARMA(2, 1) model rather than the AR(2), see He and Wang (1989). To be precise, assume that the observations in the continuous-time regression model

$$y(t) = \theta^T f(t) + \epsilon(t)\,, \quad t \in [A, B], \tag{6.8}$$

are taken at N equidistant points of the form

$$t_j = A + (j-1)\Delta\ ,\ (j = 1, \ldots, N) \tag{6.9}$$

on the interval $[A, B]$, where $\Delta = (B - A)/(N - 1)$, and that the errors $\epsilon_1, \ldots, \epsilon_N$ satisfy the discrete AR(2) equation

$$\epsilon_j - a_1\epsilon_{j-1} - a_2\epsilon_{j-2} = \varsigma_j, \tag{6.10}$$

Primary 62K05; secondary 31A10

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

where ς_j are Gaussian independent identically distributed random variables with mean 0 and variance $\sigma_\varsigma^2 = \sigma^2(1+a_2)((1-a^2)-a_1^2)/(1-a_2)$. Here we make a usual assumption that the equation (6.10) defines the AR(2) process for $j \in \{\ldots,-2,-1,0,1,2,\ldots\}$ but we only take the values such that $j \in \{1,2,\ldots,N\}$. Let $r_k = \mathbb{E}[\epsilon_j \epsilon_{j+k}]$ denote the autocovariance function of the AR(2) process $\{\epsilon_1,\ldots,\epsilon_N\}$ and assume without loss of generality that $\sigma^2 = 1$.

There are in fact three different forms of the autocovariance function (note that we assume throughout $\sigma^2 = 1$) of CAR(2) processes [see e.g. formulas (14)–(16) in He and Wang (1989)], which are given by

$$\rho_1(t) = \frac{\lambda_2}{\lambda_2-\lambda_1} e^{-\lambda_1|t|} - \frac{\lambda_1}{\lambda_2-\lambda_1} e^{-\lambda_2|t|}\,, \tag{6.11}$$

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

$$\rho_2(t) = e^{-\lambda|t|}\Big\{ \cos(q|t|) + \frac{\lambda}{q}\sin(q|t|)\Big\}\,, \tag{6.12}$$

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

$$\rho_3(t) = e^{-\lambda|t|}(1+\lambda|t|)\,, \tag{6.13}$$

where $\lambda > 0$.

Also, there are three forms of autocovariance functions of the discrete AR(2) process of the form (6.10) [see formulas (11)–(13) in He and Wang (1989)], which are given by

$$r_k^{(1)} = Cp_1^k + (1-C)p_2^k, \quad C = \frac{(1-p_2^2)p_1}{(1-p_2^2)p_1 - (1-p_1^2)p_2}\,, \tag{6.14}$$

where $j \geq 0$, $p_1 \neq p_2$, $0 < |p_1|, |p_2| < 1$; by

$$r_k^{(2)} = p^k\big(\cos(bk) + C\sin(bk)\big), \quad C = \cot(b)\frac{1-p^2}{1+p^2}\,, \tag{6.15}$$

where $0 < p < 1$, $0 < b < 2\pi$ and $b \neq \pi$, and finally by

$$r_k^{(3)} = p^k\,(1+kC), \quad C = \frac{1-p^2}{1+p^2}\,, \tag{6.16}$$

where $0 < |p| < 1$. Each form of the autocovariance function should be considered individually. However, we can formulate the following general statement.

THEOREM 6.1. *Consider the multi-parameter model (6.8) such that the errors follow the* AR(2) *model. Assume that $f(\cdot)$ is four times continuously differentiable. Define the following constants depending on the form of the autocovariance function r_k. If r_k is of the form (6.14), set*

$$\begin{aligned}
\lambda_1 &= -\frac{\ln(p_1)}{\Delta},\ \lambda_2 = -\frac{\ln(p_2)}{\Delta},\\
\tau_0 &= \lambda_1^2\lambda_2^2,\ \tau_2 = \lambda_1^2+\lambda_2^2,\ \beta_1 = \lambda_1+\lambda_2,\ \beta_0 = \lambda_1\lambda_2,\\
\gamma_1 &= \lambda_1^2+\lambda_1\lambda_2+\lambda_2^2\ ,\gamma_0 = \lambda_1\lambda_2(\lambda_1+\lambda_2),\ s_3 = 2\lambda_1\lambda_2(\lambda_1+\lambda_2).
\end{aligned}$$

If r_k is of the form (6.15), set

$$\begin{aligned}
\lambda &= -\frac{\ln(p)}{\Delta},\ q = -\frac{b}{\Delta},\\
\tau_0 &= (\lambda^2+q^2)^2,\ \tau_2 = 2(\lambda^2-q^2),\ \beta_1 = 2\lambda,\ \beta_0 = \lambda^2+q^2,\\
\gamma_1 &= 3\lambda^2-q^2\ ,\gamma_0 = 2\lambda(\lambda^2+q^2),\ s_3 = 4\lambda(\lambda^2+q^2).
\end{aligned}$$

If r_k is of the form (6.16), set

$$\begin{aligned}
\lambda &= -\frac{\ln(p)}{\Delta},\ \tau_0 = \lambda^4,\ \tau_2 = 2\lambda^2,\ \beta_1 = 2\lambda,\ \beta_0 = \lambda^2,\\
\gamma_1 &= 3\lambda^2\ ,\gamma_0 = 2\lambda^3,\ s_3 = 4\lambda^3.
\end{aligned}$$

For large N, the discrete BLUE $\hat{\theta}_{\mathrm{BLUE},N}$ based on N observations at the points (6.9) can be approximated by the continuous estimator

$$\hat{\theta} = D^*\big(z_{1,B}y'(B)+z_{1,A}y'(A)+z_Ay(A)+z_By(B)+\int_{\mathcal{T}} z(t)y(t)dt\big)$$

where

$$D^* = \Big(f^{(1)}(B)z_{1,B}^T+f^{(1)}(A)z_{1,A}^T+f(A)z_A^T+f(B)z_B^T+\int_{\mathcal{T}} f(t)z^T(t)dt\Big)^{-1}.$$

Moreover, for this approximation, we have $D^* = \lim_{N\to\infty}\mathrm{Var}(\hat{\theta}_{\mathrm{BLUE},N})$, i.e. D^* is the limit of the variance of the discrete BLUE as $N\to\infty$. Here the quantities $z(t)$, z_A, z_B, $z_{1,A}$ and $z_{1,B}$ in the continuous estimator are defined by

$$\begin{aligned}
z(t) &= -\big(\tau_2 f^{(2)}(t)-\tau_0 f(t)-f^{(4)}(t)\big)/s_3,\\
z_A &= \big(f^{(3)}(A)-\gamma_1 f^{(1)}(A)+\gamma_0 f(A)\big)/s_3,\\
z_B &= \big(-f^{(3)}(B)+\gamma_1 f^{(1)}(B)+\gamma_0 f(B)\big)/s_3,\\
z_{1,A} &= \big(-f^{(2)}(A)+\beta_1 f^{(1)}(A)-\beta_0 f(A)\big)/s_3,\\
z_{1,B} &= \big(f^{(2)}(B)+\beta_1 f^{(1)}(B)+\beta_0 f(B)\big)/s_3.
\end{aligned}$$

Proof. It is well known that the inverse of the covariance matrix $\Sigma = (\mathbb{E}[\epsilon_j \epsilon_k])_{j,k}$ of the discrete AR(2) process is a five-diagonal matrix, i.e.

$$\Sigma^{-1} = \frac{1}{S} \begin{pmatrix} k_{11} & k_{12} & k_2 & 0 & 0 & 0 & \dots & \\ k_{21} & k_{22} & k_1 & k_2 & 0 & 0 & \dots & \\ k_2 & k_1 & k_0 & k_1 & k_2 & 0 & \dots & \\ 0 & k_2 & k_1 & k_0 & k_1 & k_2 & & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & 0 & k_2 & k_1 & k_0 & k_1 & k_2 \\ & & 0 & 0 & k_2 & k_1 & k_{22} & k_{12} \\ & & 0 & 0 & 0 & k_2 & k_{21} & k_{11} \end{pmatrix}, \tag{6.17}$$

where the non-vanishing elements are given by $k_0 = 1 + a_1^2 + a_2^2$, $k_1 = -a_1 + a_1 a_2$, $k_2 = -a_2$, $k_{11} = 1$, $k_{12} = k_{21} = -a_1$, $k_{22} = 1 + a_1^2$ and $S = (1+a_1-a_2)(1-a_1-a_2)(1+a_2)/(1-a_2)$. Using the explicit form (6.17) for Σ^{-1} we immediately obtain the following result.

COROLLARY 6.1. *Consider the linear regression model (6.8) with observations at N equidistant points (6.9) and errors that follow the discrete $AR(2)$ model (6.10). Let h_i be the i-th column of matrix $H = X^T \Sigma$ and $f_i = f(t_i)$. Then the vectors $h_1, \dots, h_N$ can be represented explicitly as follows:*

$$\begin{aligned} h_1 &= \frac{1}{S}\left(k_{11} f_1 + k_{12} f_2 + k_2 f_3\right), \\ h_2 &= \frac{1}{S}\left(k_{21} f_1 + k_{22} f_2 + k_1 f_3 + k_2 f_4\right), \\ h_N &= \frac{1}{S}\left(k_{11} f_N + k_{21} f_{N-1} + k_2 f_{N-2}\right), \\ h_{N-1} &= \frac{1}{S}\left(k_{12} f_N + k_{22} f_{N-1} + k_1 f_{N-2} + k_4 f_{N-3}\right), \\ h_i &= \frac{1}{S}\left(k_2 f_{i-2} + k_1 f_{i-1} + k_0 f_i + k_1 f_{i+1} + k_2 f_{i+2}\right) \end{aligned}$$

for $i = 3, \dots, N-2$.

For the approximation of h_i, we have to study the behavior of the coefficients which depend on the autocovariance function r_k of the AR(2) process (6.10). In the following subsections we will consider the different types of autocovariance functions separately and prove Theorem 6.1 by deriving approximations for the vectors h_i.

Autocovariances of the form (6.14)**.** From Corollary 6.1 we obtain that

$$\begin{aligned}
Sh_i &= -a_2 f_{i-2} + (a_1a_2 - a_1)f_{i-1} + (1 + a_1^2 + a_2^2)f_i + (a_1a_2 - a_1)f_{i+1} - a_2 f_{i+2} \\
&= a_2(2f_i - f_{i-2} - f_{i+2}) - (a_1a_2 - a_1)(2f_i - f_{i-1} - f_{i+1}) \\
&\quad +(1 + a_1^2 + a_2^2 - 2a_2 + 2a_1a_2 - 2a_1)f_i \\
&= a_2(2f_i - f_{i-2} - f_{i+2}) - (a_1a_2 - a_1)(2f_i - f_{i-1} - f_{i+1}) \\
&\quad +(a_1 + a_2 - 1)^2 f_i
\end{aligned}$$

for $i = 3, 4, \ldots, N-2$. Now consider the case when the autocovariance structure of the errors has the form (6.14) for fixed N. Suppose that the parameters of the autocovariance function (6.14) satisfy $p_1 \neq p_2$, $0 < p_1, p_2 < 1$. We do not discuss the case with negative p_1 or negative p_2 because discrete AR(2) processes with such parameters do not have continuous real-valued analogues. From the Yule-Walker equations we obtain that the coefficients a_1 and a_2 in (6.10) are given by

$$a_1 = r_1 \frac{1 - r_2}{1 - r_1^2}, \quad a_2 = \frac{r_2 - r_1^2}{1 - r_1^2}\,, \tag{6.18}$$

where $r_1 = r_1^{(1)}$ and $r_2 = r_2^{(1)}$ are defined by (6.14). With the notation $\lambda_1 = -\log(p_1)/\Delta$ and $\lambda_2 = -\log(p_2)/\Delta$ with $\Delta = (B - A)/N$ we obtain

$$p_1 = e^{-\lambda_1 \Delta}, \quad p_2 = e^{-\lambda_2 \Delta}. \tag{6.19}$$

We will assume that λ_1 and λ_2 are fixed but Δ is small and consider the properties of different quantities as $\Delta \to 0$. By a straightforward Taylor expansion we obtain the approximations

$$\begin{aligned}
a_1 = a_1(\Delta) &= 2 - (\lambda_1 + \lambda_2)\Delta + (\lambda_1^2 + \lambda_2^2)\Delta^2/2 + O(\Delta^3), \\
a_2 = a_2(\Delta) &= -1 + (\lambda_1 + \lambda_2)\Delta - (\lambda_1 + \lambda_2)^2\Delta^2/2 + O(\Delta^3), \\
S = S(\Delta) &= 2\lambda_1\lambda_2(\lambda_1 + \lambda_2)\Delta^3 + O(\Delta^4), \\
C = C(\Delta) &= \frac{\lambda_2}{\lambda_2 - \lambda_1} + \frac{1}{6}\lambda_1\lambda_2 \frac{\lambda_1 + \lambda_2}{\lambda_1 - \lambda_2}\Delta^2 + O(\Delta^4). \qquad (6.20)
\end{aligned}$$

Consequently (observing (6.19) and (6.20)), for large N the continuous AR(2) process with autocovariances (6.11) can be considered as an approximation to the discrete AR(2) process with autocovariances (6.14).
Since $S = O(\Delta^3)$, $a_1 = 2 + O(\Delta)$ and $a_2 = -1 + O(\Delta)$, it follows

$$\begin{aligned}
S\frac{h_i}{\Delta^4} &= f^{(4)}(t_i) - 4a_2 \frac{1}{\Delta^2} f^{(2)}(t_i) + (a_1a_2 - a_1)\frac{1}{\Delta^2} f^{(2)}(t_i) + \frac{1}{\Delta^4}(a_1 + a_2 - 1)^2 f_i + O(\Delta) \\
&= f^{(4)}(t_i) + \frac{1}{\Delta^2}(a_1a_2 - a_1 - 4a_2) f^{(2)}(t_i) + \frac{1}{\Delta^4}(a_1 + a_2 - 1)^2 f_i + O(\Delta) \\
&= f^{(4)}(t_i) - (\lambda_1^2 + \lambda_2^2) f^{(2)}(t_i) + \lambda_1^2\lambda_2^2 f_i + O(\Delta).
\end{aligned}$$

Thus, the vectors h_i, $i = 3, \dots, N-2$, are approximated by the vector-function

$$z(t) = -\frac{1}{s_3}\big((\lambda_1^2+\lambda_2^2)f^{(2)}(t) - \lambda_1^2\lambda_2^2 f(t) - f^{(4)}(t)\big),$$

where $s_3 = 2\lambda_1\lambda_2(\lambda_1+\lambda_2)$. For the boundary points we obtain

$$\begin{aligned} Sh_1 &= f_1 - a_1 f_2 - a_2 f_3 \\ &= (-2f_2 + f_3 + f_1) + (\lambda_1+\lambda_2)(f_2 - f_3)\Delta \\ &\quad +((-1/2f_2 + 1/2f_3)\lambda_1^2 + f_3\lambda_1\lambda_2 + (-1/2f_2+1/2f_3)\lambda_2^2)\Delta^2 \\ &\quad +((1/6f_2 - 1/6f_3)\lambda_1^3 - 1/2f_3\lambda_1^2\lambda_2 - 1/2f_3\lambda_1\lambda_2^2 + (1/6f_2 - 1/6f_3)\lambda_2^3)\Delta^3 + O(\Delta^4) \\ &= \big(f^{(2)}(t_2) - (\lambda_1+\lambda_2)f^{(1)}(t_2) + f_3\lambda_1\lambda_2\big)\Delta^2 + O(\Delta^3) \end{aligned}$$

and

$$\begin{aligned} Sh_2 &= -a_1 f_1 + (1+a_1^2)f_2 + (a_1a_2 - a_1)f_3 - a_2 f_4 \\ &= (-2f_1 + f_4 + 5f_2 - 4f_3) + (\lambda_1+\lambda_2)(f_1 - 4f_2 + 4f_3 - f_4)\Delta \\ &\quad +((-1/2f_1 + 1/2f_4 - 3f_3 + 3f_2)\lambda_1^2 + (2f_2 - 4f_3 + f_4)\lambda_2\lambda_1 \\ &\qquad + (-1/2f_1 + 1/2f_4 - 3f_3 + 3f_2)\lambda_2^2)\Delta^2 \\ &\quad +((1/6f_1 - 5/3f_2 + 5/3f_3 - 1/6f_4)\lambda_1^3 + (-f_2 + 3f_3 - 1/2f_4)\lambda_2\lambda_1^2 \\ &\qquad + (-f_2 + 3f_3 - 1/2f_4)\lambda_2^2\lambda_1 + (1/6f_1 - 5/3f_2 + 5/3f_3 - 1/6f_4)\lambda_2^3)\Delta^3 + O(\Delta^4) \\ &= \big(f^{(2)}(t_3) - 2f^{(2)}(t_2) + (\lambda_1+\lambda_2)(3f^{(1)}(t_2) - f^{(1)}(t_1) - f^{(1)}(t_3)) - f_3\lambda_1\lambda_2\big)\Delta^2 + O(\Delta^3) \\ &= \big(-f^{(2)}(t_2) + (\lambda_1+\lambda_2)f^{(1)}(t_2) - f_3\lambda_1\lambda_2\big)\Delta^2 + O(\Delta^3) \end{aligned}$$

Thus, we can see that

$$h_1 = -h_2 + O(1) = z_{1,A}\frac{1}{\Delta} + O(1),$$

where

$$z_{1,A} = \frac{1}{s_3}\big(-f^{(2)}(A) + (\lambda_1+\lambda_2)f^{(1)}(A) - \lambda_1\lambda_2 f(A)\big).$$

This means that the vectors h_1 and h_2 at t_1 and t_2 are large in absolute value and have different signs. Similarly, we have

$$h_N = -h_{N-1} + O(1) = z_{1,B}\frac{1}{\Delta} + O(1)$$

where

$$z_{1,B} = \frac{1}{s_3}\big(f^{(2)}(B) + (\lambda_1+\lambda_2)f^{(1)}(B) + \lambda_1\lambda_2 f(B)\big).$$

To do a finer approximation, it is necessary to investigate the quantity

$$g := Sh_1 + Sh_2,$$

which is of order $O(1)$. Indeed, we have

$$\begin{aligned} g &= (3f_2 - 3f_3 - f_1 + f_4) + (\lambda_2 + \lambda_1)(f_1 - 3f_2 + 3f_3 - f_4)\Delta \\ &\quad + ((-f_1 + f_4 - 5f_3 + 5f_2)/2(\lambda_1^2 + \lambda_2^2) + (2f_2 - 3f_3 + f_4)\lambda_2\lambda_1)\Delta^2 \\ &\quad + ((f_1 - 9f_2 + 9f_3 - f_4)/6(\lambda_1^3 + \lambda_2^3) \\ &\quad + (-2f_2 + 5f_3 - f_4)/2(\lambda_1^2\lambda_2 + \lambda_1\lambda_2^2))\Delta^3 + O(\Delta^4) \\ &= f^{(3)}(t_1)\Delta^3 + O(\Delta^4) + (-f^{(1)}(t_1)(\lambda_1^2 + \lambda_2^2) - f^{(1)}(t_1)\lambda_2\lambda_1)\Delta^3 \\ &\quad + f(t_1)(\lambda_1^2\lambda_2 + \lambda_1\lambda_2^2)\Delta^3 + O(\Delta^4) \\ &= \big(f^{(3)}(t_1) - (\lambda_1^2 + \lambda_1\lambda_2 + \lambda_2^2)f^{(1)}(t_1) + \lambda_1\lambda_2(\lambda_1 + \lambda_2)f(t_1)\big)\Delta^3 + O(\Delta^4) \end{aligned}$$

and, consequently,

$$h_1 + h_2 = \frac{1}{s_3}\big(f^{(3)}(t_1) - (\lambda_1^2 + \lambda_1\lambda_2 + \lambda_2^2)f^{(1)}(t_1) + \lambda_1\lambda_2(\lambda_1 + \lambda_2)f(t_1)\big) + O(\Delta),$$

where $s_3 = 2\lambda_1\lambda_2(\lambda_1+\lambda_2)$. Therefore, if $\Delta \to 0$, it follows that $h_1+h_2 \approx z_A$, where

$$z_A = \frac{1}{s_3}\big(f^{(3)}(A) - (\lambda_1^2 + \lambda_1\lambda_2 + \lambda_2^2)f^{(1)}(A) + \lambda_1\lambda_2(\lambda_1 + \lambda_2)f(A)\big).$$

Similarly, we obtain $h_N + h_{N-1} \approx z_B$ if $\Delta \to 0$, where

$$z_B = \frac{1}{s_3}\big(- f^{(3)}(B) + (\lambda_1^2 + \lambda_1\lambda_2 + \lambda_2^2)f^{(1)}(B) + \lambda_1\lambda_2(\lambda_1 + \lambda_2)f(B)\big).$$

Autocovariances of the form (6.15). Consider the autocovariance function of the form (6.15), then the coefficients a_1 and a_2 are given by (6.18) where $r_1 = r_1^{(2)}$ and $r_2 = r_2^{(2)}$ are defined by (6.15). With the notations $\lambda = -\log p/\Delta$ and $q = b/\Delta$ (or equivalently $p = e^{-\lambda\Delta}$ and $b = q\Delta$) we obtain by a Taylor expansion

$$\begin{aligned} a_1 &= 2 - 2\lambda\Delta + (\lambda^2 - q^2)\Delta^2 + O(\Delta^3), \\ a_2 &= -1 + 2\lambda\Delta - 2\lambda^2\Delta^2 + O(\Delta^3), \\ S &= 4\lambda(\lambda^2 + q^2)\Delta^3 + O(\Delta^4) \end{aligned}$$

and

$$C = \frac{\lambda}{q} - \frac{\lambda(\lambda^2 + q^2)}{3q}\Delta^2 + O(\Delta^4)$$

as $\Delta \to 0$. Similarly, we have

$$\begin{aligned} S\frac{h_i}{\Delta^4} &= f^{(4)}(t_i) + \frac{1}{\Delta^2}(a_1a_2 - a_1 - 4a_2)f^{(2)}(t_i) + \frac{1}{\Delta^4}(a_1 + a_2 - 1)^2 f_i + O(\Delta) \\ &= -2(\lambda^2 - q^2)f^{(2)}(t_i) + (\lambda^2 + q^2)^2 f_i + O(\Delta). \end{aligned}$$

Thus, the optimal weights h_i, $i = 3, \ldots, N-2$, are approximated by the signed density

$$z(t) = -\frac{1}{s_3}\big(2(\lambda^2 - q^2)f^{(2)}(t) - (\lambda^2 + q^2)^2 f(t) - f^{(4)}(t)\big),$$

where $s_3 = 4\lambda(\lambda^2 + q^2)$. Similarly, we obtain that

$$\begin{aligned} h_1 &= -h_2 + O(1) = z_{1,A}\frac{1}{\Delta} + O(1), \\ h_N &= -h_{N-1} + O(1) = z_{1,B}\frac{1}{\Delta} + O(1), \end{aligned}$$

where

$$\begin{aligned} z_{1,A} &= \frac{1}{s_3}\big(-f^{(2)}(A) + 2\lambda f^{(1)}(A) - (\lambda^2 + q^2)f(A)\big), \\ z_{1,B} &= \frac{1}{s_3}\big(f^{(2)}(B) + 2\lambda f^{(1)}(B) + (\lambda^2 + q^2)f(B)\big). \end{aligned}$$

Calculating $g := Sh_1 + Sh_2$ we have

$$\begin{aligned} g &= (3f_2 - f_1 - 3f_3 + f_4) + 2\lambda(f_1 - 3f_2 + 3f_3 - f_4)\Delta \\ &\quad + ((-f_1 + 7f_2 - 8f_3 + 2f_4)\lambda^2 + q^2(f_1 - 3f_2 + 2f_3))\Delta^2 \\ &\quad + ((-f_1 + 7f_2 - 4f_3)\lambda q^2 + (f_1 - 15f_2 + 24f_3 - 4f_4)/3\lambda^3)\Delta^3 + O(\Delta^4) \\ &= f^{(3)}(t_1)\Delta^3 - (3\lambda^2 - q^2)f^{(1)}(t_1)\Delta^3 + 2\lambda(\lambda^2 + q^2)f(t_1)\Delta^3 + O(\Delta^4). \end{aligned}$$

Therefore, it follows that $h_1 + h_2 \approx P_A$ if $\Delta \to 0$, where

$$z_A = \frac{1}{s_3}\big(f^{(3)}(A) - (3\lambda^2 - q^2)f^{(1)}(A) + 2\lambda(\lambda^2 + q^2)f(A)\big),$$

and $s_3 = 4\lambda(\lambda^2 + q^2)$. Similarly, we obtain the approximation $h_N + h_{N-1} \approx P_B$ if $\Delta \to 0$, where

$$z_B = \frac{1}{s_3}\big(-f^{(3)}(B) + (3\lambda^2 - q^2)f^{(1)}(B) + 2\lambda(\lambda^2 + q^2)f(B)\big).$$

Autocovariances of the form (6.16)**.** For the autocovariance function (6.16) the coefficients a_1 and a_2 in the AR(2) process are given by (6.18) where $r_1 = r_1^{(3)}$ and $r_2 = r_2^{(3)}$ are defined by (6.16). With the notation $\lambda = -\log p/\Delta$ (or equivalently $p = e^{-\lambda\Delta}$) we obtain the Taylor expansions

$$\begin{aligned} a_1 &= 2 - 2\lambda\Delta + \lambda^2\Delta^2 + O(\Delta^3), \\ a_2 &= -1 + 2\lambda\Delta - 2\lambda^2\Delta^2 + O(\Delta^3), \\ S &= 4\lambda^3\Delta^3 + O(\Delta^4), \\ C &= \lambda\Delta - \frac{\lambda^3}{3}\Delta^3 + O(\Delta^5) \end{aligned}$$

as $\Delta \to 0$. Similar calculations as given in the previous paragraphs give

$$\begin{aligned} S\frac{h_i}{\Delta^4} &= f^{(4)}(t_i) + \frac{1}{\Delta^2}(a_1a_2 - a_1 - 4a_2)f^{(2)}(t_i) + \frac{1}{\Delta^4}(a_1 + a_2 - 1)^2 f_i + O(\Delta) \\ &= f^{(4)}(t_i) - 2\lambda^2 f^{(2)}(t_i) + \lambda^4 f_i + O(\Delta). \end{aligned}$$

Thus, the vectors h_i, $i = 3, \ldots, N-2$, are approximated by the signed density

$$z(t) = -\frac{1}{s_3}\big(2\lambda^2 f^{(2)}(t) - \lambda^4 f(t) - f^{(4)}(t)\big),$$

where $s_3 = 4\lambda^3$. For the remaining vectors h_1, h_2, h_{N-1} and h_N we obtain

$$\begin{aligned} h_1 &= -h_2 + O(1) = z_{1,A}\frac{1}{\Delta} + O(1), \\ h_N &= -h_{N-1} + O(1) = z_{1,B}\frac{1}{\Delta} + O(1), \end{aligned}$$

with

$$\begin{aligned} z_{1,A} &= \frac{1}{s_3}\big(- f^{(2)}(A) + 2\lambda f^{(1)}(A) - \lambda^2 f(A)\big), \\ z_{1,B} &= \frac{1}{s_3 f}\big(f^{(2)}(B) + 2\lambda f^{(1)}(B) + \lambda^2 f(B)\big). \end{aligned}$$

Calculating $g := Sh_1 + Sh_2$ we have

$$\begin{aligned} g &= (3f_2 - 3f_3 - f_1 + f_4) + 2\lambda(f_1 - 3f_2 + 3f_3 - f_4)\Delta \\ &\quad -\lambda^2(f_1 - 7f_2 + 8f_3 - 2f_4)\Delta^2 \\ &\quad +1/3\lambda^3(f_1 - 15f_2 + 24f_3 - 4f_4)\Delta^3 + O(\Delta^4) \\ &= f^{(3)}(t_1)\Delta^3 - 3\lambda^2 f^{(1)}(t_1)\Delta^3 + 2\lambda^3 f(t_1)\Delta^3 + O(\Delta^4). \end{aligned}$$

Therefore, if $\Delta \to 0$, it follows that $h_1 + h_2 \approx z_A$, where

$$z_A = \frac{1}{s_3}\big(f^{(3)}(A) - 3\lambda^2 f^{(1)}(A) + 2\lambda^3 f(A)\big),$$

and $s_3 = 4\lambda^3$. Similarly, we obtain the approximation $h_N + h_{N-1} \approx z_B$ if $\Delta \to 0$, where

$$z_B = \frac{1}{s_3}\big(- f^{(3)}(B) + 3\lambda^2 f^{(1)}(B) + 2\lambda^3 f(B)\big).$$

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