Optimal designs for comparing regression curves - dependence within and between groups

Kirsten Schorning

Holger Dette

Technische Universität Dortmund

Ruhr-Universität Bochum

Fakultät Statistik

Fakultät für Mathematik

44221 Dortmund, Germany

44780 Bochum, Germany

Abstract

We consider the problem of designing experiments for the comparison of two regression curves describing the relation between a predictor and a response in two groups, where the data between and within the group may be dependent. In order to derive efficient designs we use results from stochastic analysis to identify the best linear unbiased estimator (BLUE) in a corresponding continuous time model. It is demonstrated that in general simultaneous estimation using the data from both groups yields more precise results than estimation of the parameters separately in the two groups. Using the BLUE from simultaneous estimation, we then construct an efficient linear estimator for finite sample size by minimizing the mean squared error between the optimal solution in the continuous time model and its discrete approximation with respect to the weights (of the linear estimator). Finally, the optimal design points are determined by minimizing the maximal width of a simultaneous confidence band for the difference of the two regression functions. The advantages of the new approach are illustrated by means of a simulation study, where it is shown that the use of the optimal designs yields substantially narrower confidence bands than the application of uniform designs.

Keywords and Phrases: optimal design, correlated observations, Gaussian white noise model, comparison of curves

AMS Subject classification: Primary 62K05; Secondary: 62M05

1 Introduction

The application of optimal or efficient designs can improve the accuracy of statistical analysis substantially and meanwhile there exists a well established and powerful theory for the construction of (approximate) optimal designs for independent observations, see for example the monographs of Pukelsheim (2006) or Fedorov and Leonov (2013). In contrast, the determination of optimal designs for efficient statistical analysis from dependent data is more challenging because the corresponding optimization problems are in general not convex and therefore the powerful tools of convex analysis are not applicable. Although design problems for correlated data have been discussed for a long time (see, for example Sacks and Ylvisaker, 1966, 1968; Bickel and Herzberg, 1979; Näther, 1985, who use asymptotic arguments to develop continuous but in general non-convex optimzation problems in this context) a large part of the discussion is restricted to models with a small number of parameters and we refer Pázman and Müller (2001), Müller and Pázman (2003), Dette et al. (2008), Kiselak and Stehlík (2008), Harman and Štulajter (2010), Rodriguez-Diaz (2017), Campos-Barreiro and López-Fidalgo (2015) and Attia and Constantinescu (2020) among others.

Recently, Dette et al. (2013) suggest a more systematic approach to the problem and determine (asymptotic) optimal designs for least squares estimation, under the additional assumption that the regression functions are eigenfunctions of an integral operator associated with the covariance kernel of the error process. This approach refers to models, where the regression functions are eigenfunctions of the integral operator corresponding to the covariance kernel, which is used to describe the dependencies. For more general models Dette et al. (2016) propose to construct the optimal design and estimator simultaneously. More precisely, they construct a class of estimators and corresponding optimal designs with a variance converging (as the sample size increases) to the optimal variance in the continuous time model. Dette et al. (2017a) propose an alternative strategy for this purpose. They start with the construction of the best linear unbiased estimator (BLUE) in the continuous time model using stochastic calculus and determine in a second step an implementable design, which is "close" to the solution in the continuous time model. By this approach these authors are able to provide an easily implementable estimator with a corresponding design which is practically non distinguishable from the weighted least squares estimate (WLSE) with corresponding optimal design. Their results are applicable for a broad class of linear regression models with various covariance kernels and have recently been extended to the situation, where also derivatives of the process can be observed (see Dette et al., 2019).

Dette and Schorning (2016) and Dette et al. (2017b) propose designs for the comparison of regression curves from two independent samples, where the latter reference also allows for dependencies within the samples. Their work is motivated by applications in drug development,

where a comparison between two regression models that describe the relation between a common response and the same covariates for two groups is used to establish the non-superiority of one model to the other or to check whether the difference between two regression models can be neglected. For example, if the similarity between two regression functions describing the dose response relationships in the groups individually has been established subsequent inference in drug development could be based on the combined samples such that a more efficient statistical analysis is possible on the basis of the larger population. Because of its importance several procedures for the comparison of curves have been investigated in linear and nonlinear models (see Liu et al., 2009; Gsteiger et al., 2011; Liu et al., 2011; Dette et al., 2018; Bretz et al., 2018; Möllenhoff et al., 2018, 2020, among others). Designs minimizing the maximal width of a (simultaneous) confidence band for the difference between the regression curves calculated from two independent groups are determined by Dette and Schorning (2016) and Dette et al. (2017b), who also demonstrate that the use of these designs yields to substantially narrower confidence bands.

While these results refer to independent groups it is the purpose of the present paper to investigate designs for the comparison of regression curves corresponding to two groups, where the data within the groups and between the groups may be dependent. It will be demonstrated that in most cases simultaneous estimation of the parameters in the regression models using the data from both groups yields to more efficient inference than estimating the parameters in the models corresponding to the different groups separately. Moreover, the simultaneous estimation procedure can never be worse. While this property holds independently of the design under consideration, we subsequently construct efficient designs for the comparison of curves corresponding to not necessarily independent groups and demonstrate its superiority by means of a simulation study.

The remaining part of this paper is organized as follows. In Section 2 we introduce the basics and the design problem. Section 3 is devoted to a continuous time model, which could be interpreted as a limiting experiment of the discrete model if the sample size converges to infinity. In this model we derive an explicit representation of the BLUE if estimation is performed simultaneously is both groups. In Section 4 we develop a discrete approximation of the continuous BLUE by determining the optimal weights for the linear estimator. Finally, the optimal design points are determined such that the maximum width of the confidence band for the difference of the two regression functions is minimal. Section 5 is devoted to a small numerical comparison of the performance of the optimal designs with uniform designs. In particular, it is demonstrated that optimal designs yield substantially narrower confidence bands. In many cases the maximal width of a confidence band from the uniform design is by a factor between 2 and 10 larger than the width of a confidence band from the optimal design.

2 Simultaneous estimation of two regression models

Throughout this paper we consider the situation of two groups of observations $Y_{1,1}, \ldots, Y_{1,n}$ and $Y_{2,1}, \ldots, Y_{2,n}$ at time points $t_1, \ldots, t_n$ ($i = 1, 2$) where there may exist dependencies within and between the groups. We assume the relation between the response and the covariate t in each group is described by a linear regression models given by

$$Y_{ij} = Y_i(t_j) = f_i^\top(t_j)\theta^{(i)} + \eta_i(t_j)\,,\ j = 1, \ldots, n\,,\ i = 1, 2\,. \tag{2.1}$$

Thus in each group n observations are taken at the same time points $t_1, \ldots, t_n$ which can be chosen in a compact interval, say $[a, b]$, and observations at different time points and in different groups might be dependent. The vectors of the unknown parameters $\theta^{(1)}$ and $\theta^{(2)}$ are assumed to be p_1- and p_2-dimensional, respectively, and the corresponding vectors of regression functions $f_i(t) = (f_{i,1}(t), \ldots, f_{i,p_i}(t))^\top$, $i = 1, 2$, have continuously differentiable linearly independent components.

To address the situation of correlation between the groups, we start with a very simple covariance structure for each group, but we emphasize that all results presented in this paper are correct for more general covariance structures corresponding to Markov processes, see Remark 3.3 for more details. To be precise, let $\{\varepsilon_1(t)|\ t \in [a, b]\}$ and $\{\varepsilon_2(t)|\ t \in [a, b]\}$ denote two independent Brownian motions, such that

$$\mathbb{E}[\varepsilon_i(t_j)] = 0,\quad K_i(t_j, t_k) = \mathbb{E}[\varepsilon_i(t_j)\varepsilon_i(t_k)] = \min(t_j, t_k) \tag{2.2}$$

denotes the mean value and the covariance of the individual process ε_i at the points t_j and t_k, respectively. Let $\sigma_1, \sigma_2 > 0$, $\varrho \in (-1, 1)$, denote by $\mathbf{\Sigma}^{1/2}$ the square root of the covariance matrix

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\varrho \\ \sigma_1\sigma_2\varrho & \sigma_2^2 \end{pmatrix}, \tag{2.3}$$

and define for $t \in [a, b]$ the two-dimensional process $\{\boldsymbol{\eta}(t)|\ t \in [a, b]\}$ by

$$\boldsymbol{\eta}(t) = \begin{pmatrix} \eta_1(t) \\ \eta_2(t) \end{pmatrix} = \mathbf{\Sigma}^{1/2}\boldsymbol{\varepsilon}(t), \tag{2.4}$$

where $\boldsymbol{\varepsilon}(t) = (\varepsilon_1(t), \varepsilon_2(t))^\top$. Note that $\varrho \in (-1, 1)$ denotes the correlation between the observations $Y_1(t_j)$ and $Y_2(t_j)$ ($j = 1, \ldots, n$), and that in general the correlation between $Y_1(t_j)$ and $Y_2(t_k)$ is given by

$$\mathrm{Corr}(Y_1(t_j), Y_2(t_k)) = \varrho \left\{ \sqrt{\frac{t_j}{t_k}} \wedge \sqrt{\frac{t_k}{t_j}} \right\}$$

if $t_j, t_k \in (0,1)$.

Considering the two groups individually results in proper (for example weighted least squares) estimators of the parameters $\theta^{(1)}$ and $\theta^{(2)}$. However, this procedure ignores the correlation between the two groups and estimating the parameters $\theta^{(1)}$ and $\theta^{(2)}$ simultaneously from the data of both groups might result in more precise estimates. In order to define estimators for the parameters $\theta^{(1)}$ and $\theta^{(2)}$ using the information from both groups we now consider a more general two-dimensional regression model, which on the one hand contains the situation described in the previous paragraph as special case, but on the other hand also allows us to consider the case, where some of the components in θ_1 and θ_2 coincide, see Example 2.2 and Section 3.3 for details. To be precise we define the regression model

$$\mathbf{Y}(t_j) = \begin{pmatrix} Y_1(t_j) \\ Y_2(t_j) \end{pmatrix} = \mathbf{F}^\top(t_j)\theta + \boldsymbol{\eta}(t_j) = \mathbf{F}^\top(t_j)\theta + \mathbf{\Sigma}^{1/2}\boldsymbol{\varepsilon}(t_j), \qquad j = 1, \ldots, n, \qquad (2.5)$$

where two-dimensional observations

$$\mathbf{Y}(t_1) = (Y_1(t_1), Y_2(t_1))^\top, \ldots, \mathbf{Y}(t_n) = (Y_1(t_n), Y_2(t_n))^\top$$

are available at time points $t_1, \ldots, t_n \in [a, b]$. In model (2.5) the vector $\theta = (\vartheta_1, \ldots, \vartheta_p)^\top$ is a p-dimensional parameter and

$$\mathbf{F}^\top(t) = \begin{pmatrix} F_1^\top(t) \\ F_2^\top(t) \end{pmatrix} = \begin{pmatrix} F_{1,1}(t) & \ldots & F_{1,p}(t) \\ F_{2,1}(t) & \ldots & F_{2,p}(t) \end{pmatrix} \qquad (2.6)$$

denotes a $(p \times 2)$ matrix containing continuously differentiable regression functions, where the two-dimensional functions $(F_{1,1}(t), F_{2,1}(t))^\top, \ldots, (F_{1,p}(t), F_{2,p}(t))^\top$ are assumed to be linearly independent.

Example 2.1 The individual models defined in (2.1) are contained in this two-dimensional model. More precisely, defining the $p = (p_1 + p_2)$-dimensional vector of parameters θ by $\theta = ((\theta^{(1)})^\top, (\theta^{(2)})^\top)^\top$ and the regression function $\mathbf{F}^\top(t)$ in (2.6) by the rows

$$F_1^\top(t) = (f_1^\top(t), 0_{p_2}^\top), \quad F_2^\top(t) = (0_{p_1}^\top, f_2^\top(t)),$$

it follows that model (2.5) coincides with model (2.1). Moreover, this composite model takes the correlation between the groups into account. In this case the models describing the relation between the variable t and the responses $Y_1(t)$ and $Y_2(t)$ do not share any parameters.

Example 2.2 In this example we consider the situation where some of the parameters of the individual models in (2.1) coincide. This situation occurs, for example, if $Y_1(t)$ and $Y_2(t)$ represent clinical parameters (depending on time) before and after treatment, where it can be assumed that the effect at time a coincides before and after the treatment. In this case a reasonable model for average effect in the two groups is given by

$$\mathbb{E}[Y_\ell(t)] = \theta^{(0)} + (\tilde{\theta}^{(\ell)})^\top \tilde{f}_\ell(t) \, , \quad \ell = 1, 2 \, .$$

More generally, we consider the situation where the vectors of the parameters are given by

$$\theta^{(1)} = (\theta^{(0)^\top}, \tilde{\theta}^{(1)^\top})^\top \quad , \quad \theta^{(2)} = (\theta^{(0)^\top}, \tilde{\theta}^{(2)^\top})^\top \, ,$$

where $\theta^{(0)} \in \mathbb{R}^{p_0}$ denotes the vector of common parameters in both models and vectors $\tilde{\theta}^{(1)} \in \mathbb{R}^{p_1-p_0}$ and $\tilde{\theta}^{(2)} \in \mathbb{R}^{p_2-p_0}$ contain the different parameters in the two individual models. The corresponding regression functions are given by

$$f_1^\top(t) = (f_0^\top(t), \tilde{f}_1^\top(t)) \quad , \quad f_2^\top(t) = (f_0^\top(t), \tilde{f}_2^\top(t)) \, , \tag{2.7}$$

where the vector $f_0^\top(t)$ contains the regression functions corresponding to the common parameters in the two models, and $\tilde{f}_1^\top(t)$ and $\tilde{f}_2^\top(t)$ denote the vectors of regression functions corresponding to the different parameters $\tilde{\theta}^{(1)}$ and $\tilde{\theta}^{(2)}$, respectively.

Defining the $p = (p_1 + p_2 - p_0)$-dimensional vector of parameters θ by $\theta = (\theta^{(0)}, \tilde{\theta}^{(1)}, \tilde{\theta}^{(2)})$ and the regression function $\mathbf{F}^\top(t)$ in (2.6) by the rows

$$F_1^\top(t) = (f_0^\top(t), \tilde{f}_1^\top(t), 0_{p_2-p_0}^\top), \quad F_2^\top(t) = (f_0^\top(t), 0_{p_1-p_0}^\top, \tilde{f}_2^\top(t)),$$

it follows that the model (2.5) contains the individual models in (2.1), where the regression functions are given by (2.7) and the parameters $\theta^{(1)}$ and $\theta^{(2)}$ share the parameter $\theta^{(0)}$. Moreover, this composite model takes the potential correlation between the groups into account.

3   Continuous time models

It was demonstrated by Dette et al. (2017a) that efficient designs for dependent data in regression problems can be derived by first considering the estimation problem in a continuous time model. In this model there is no optimal design problem as the data can be observed over the full interval $[a, b]$. However, efficient designs can be determined in two steps. First, one derives the best linear unbiased estimator (BLUE) in the continuous time model and, secondly, one determines design points (and an estimator) such that the resulting estimator from the discrete data provides a good approximation of the optimal solution in the continuous time model. In this paper we will use this strategy to develop optimal designs for the comparison of regression

curves from two (possible) dependent groups. In the present section we discuss a continuous time model corresponding to the discrete model (2.5), while the second step, the determination of an "optimal" approximation will be postponed to the following Section 4 .

3.1 Best linear unbiased estimation

To be precise, we consider the continuous time version of the linear regression model in (2.5), that is,

$$\mathbf{Y}(t) = \begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix} = \mathbf{F}^\top(t)\theta + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\varepsilon}(t)\,, \quad t \in [a,b]\,, \tag{3.1}$$

where we assume $0 < a < b$ and the full trajectory of the process $\{\mathbf{Y}(t) \mid t \in [a,b]\}$ is observed, $\{\boldsymbol{\varepsilon}(t) = (\varepsilon_1(t), \varepsilon_2(t))^\top \mid t \in [a,b]\}$ is a vector of independent Brownian motions as defined in (2.2) and the matrix $\boldsymbol{\Sigma}^{1/2}$ is the square root of the covariance matrix $\boldsymbol{\Sigma}$ defined in (2.3). Note that we restrict ourselves to an interval on the positive line, because in this case the notation is slightly simpler. But we emphasize that the theory developed in this section can also be applied for $a = 0$, see Remark 3.1 for more details. We further assume that the $(p \times p)$-matrix

$$\mathbf{M} = \int_a^b \dot{\mathbf{F}}(t)\boldsymbol{\Sigma}^{-1}\dot{\mathbf{F}}^\top(t)\, dt + \frac{1}{a}\mathbf{F}(a)\boldsymbol{\Sigma}^{-1}\mathbf{F}^\top(a) \tag{3.2}$$

is non-singular.

Theorem 3.1 *Consider the continuous time linear regression model* (3.1) *on the interval* $[a,b]$, $a > 0$, *with a continuously differentiable matrix of regression functions* $\mathbf{F}$, *a vector* $\{\boldsymbol{\varepsilon}(t) = (\varepsilon_1(t), \varepsilon_2(t))^\top \mid t \in [a,b]\}$ *of independent Brownian motions and a covariance matrix* $\boldsymbol{\Sigma}$ *defined by* (2.3). *The best linear unbiased estimator of the parameter* θ *is given by*

$$\hat{\theta}_{\mathrm{BLUE}} = \mathbf{M}^{-1}\Big(\int_a^b \dot{\mathbf{F}}(t)\boldsymbol{\Sigma}^{-1}\, d\mathbf{Y}(t) + \frac{1}{a}\mathbf{F}(a)\boldsymbol{\Sigma}^{-1}\mathbf{Y}(a) \Big). \tag{3.3}$$

Moreover, the minimum variance is given by

$$\mathrm{Cov}(\hat{\theta}_{\mathrm{BLUE}}) = \mathbf{M}^{-1} = \Big(\int_a^b \dot{\mathbf{F}}(t)\boldsymbol{\Sigma}^{-1}\dot{\mathbf{F}}^\top(t)\, dt + \frac{1}{a}\mathbf{F}(a)\boldsymbol{\Sigma}^{-1}\mathbf{F}^\top(a) \Big)^{-1}. \tag{3.4}$$

Proof. Multiplying $\mathbf{Y}$ by the matrix $\boldsymbol{\Sigma}^{-1/2}$ yields a transformed regression model

$$\tilde{\mathbf{Y}}(t) = \begin{pmatrix} \tilde{Y}_1(t) \\ \tilde{Y}_2(t) \end{pmatrix} = \boldsymbol{\Sigma}^{-1/2} \begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix} = \boldsymbol{\Sigma}^{-1/2}\mathbf{F}^\top(t)\theta + \boldsymbol{\varepsilon}(t)\,, \tag{3.5}$$

where $\boldsymbol{\Sigma}^{-1/2}$ is the inverse of $\boldsymbol{\Sigma}^{1/2}$, the square root of the covariance matrix $\boldsymbol{\Sigma}$ defined in (2.3). Note that the components of the vector $\tilde{\boldsymbol{Y}}$ are independent, and consequently, the joint likelihood function can be obtained as the product of the individual components. Next we rewrite the components of the continuous time model (3.5) in terms of two stochastic differential equations, that is

$$d\tilde{Y}_i(t) \;=\; \mathbf{1}_{[a,b]}(t)\boldsymbol{\Sigma}_i^{-1/2}\dot{\mathbf{F}}^\top(t)\theta dt + d\varepsilon_i(t)\,, t \in [0,b]\,, \tag{3.6}$$

$$\tilde{Y}_i(a) \;=\; \boldsymbol{\Sigma}_i^{-1/2}\mathbf{F}^\top(a)\theta + \varepsilon_i(a)\,, \tag{3.7}$$

where $\mathbf{1}_A$ is the indicator function of the set A and $\boldsymbol{\Sigma}_i^{-1/2}$ denotes the i-th row of the matrix $\boldsymbol{\Sigma}^{-1/2}$ ($i = 1, 2$). Since $\{\varepsilon_i(t)|\ t \in [a,b]\}$ is a Brownian motion its increments are independent. Consequently, the processes $\{\tilde{Y}_i(t)|\ t \in [0,b]\}$ and the random variable $\tilde{Y}_i(a)$ are independent. To obtain the joint density of the processes defined by (3.6) and (3.7) it is therefore sufficient to derive the individual densities.

Let $\mathbb{P}_\theta^{(i)}$ and $\mathbb{P}_0^{(i)}$ denote the measures on $C([0,b])$ associated with the process $\tilde{Y}_i = \{Y_i(t)|\ t \in [0,b]\}$ and $\{\varepsilon_i(t)|\ t \in [0,b]\}$, respectively. It follows from Theorem 1 in Appendix II of Ibragimov and Has'minskii (1981) that $\mathbb{P}_\theta^{(i)}$ is absolute continuous with respect to $\mathbb{P}_0^{(i)}$ with Radon-Nikodym-density

$$\frac{d\mathbb{P}_\theta^{(i)}}{d\mathbb{P}_0^{(i)}}(\tilde{Y}_i) = \exp\left\{ \int_a^b \boldsymbol{\Sigma}_i^{-1/2}\dot{\mathbf{F}}^\top(t)\theta d\tilde{Y}_i(t) - \frac{1}{2}\int_a^b (\boldsymbol{\Sigma}_i^{-1/2}\dot{\mathbf{F}}^\top(t)\theta)^2 dt \right\}\,.$$

Similarly, if $\mathbb{Q}_\theta$ denotes the distribution of the random variable $\tilde{Y}_i(a) \sim \mathcal{N}(\boldsymbol{\Sigma}_i^{-1/2}\mathbf{F}^\top(a)\theta, a)$ in (3.7), then the Radon-Nikodym-density of $\mathbb{Q}_\theta^{(i)}$ with respct to $\mathbb{Q}_0^{(i)}$ is given by

$$\frac{d\mathbb{Q}_\theta^{(i)}}{d\mathbb{Q}_0^{(i)}}(\tilde{Y}_i(a)) = \exp\left\{ \frac{\tilde{Y}_i(a)\boldsymbol{\Sigma}_i^{-1/2}\mathbf{F}^\top(a)\theta}{a} - \frac{1}{2}\frac{(\boldsymbol{\Sigma}_i^{-1/2}\mathbf{F}^\top(a)\theta)^2}{a} \right\}\,.$$

Consequently, because of independence, the joint density of $(\mathbb{P}_\theta^{(i)}, \mathbb{Q}_\theta^{(i)})$ with respect to $(\mathbb{P}_0^{(i)}, \mathbb{Q}_0^{(i)})$ is obtained as

$$\frac{d\mathbb{P}_\theta^{(i)}}{d\mathbb{P}_0^{(i)}}(\tilde{Y}_i) \times \frac{d\mathbb{Q}_\theta^{(i)}}{d\mathbb{Q}_0^{(i)}}(\tilde{Y}_i(a)) = \exp\left\{ \left(\int_a^b \boldsymbol{\Sigma}_i^{-1/2}\dot{\mathbf{F}}^\top(t)\theta d\tilde{Y}_i(t) + \frac{\tilde{Y}_i(a)\boldsymbol{\Sigma}_i^{-1/2}\mathbf{F}^\top(a)\theta}{a} \right) \right.$$
$$\left. - \frac{1}{2}\left(\int_a^b (\boldsymbol{\Sigma}_i^{-1/2}\dot{\mathbf{F}}^\top(t)\theta)^2 dt + \frac{(\boldsymbol{\Sigma}_i^{-1/2}\mathbf{F}(a)\theta)^2}{a} \right) \right\}\,.$$

As the processes $\tilde{Y}_1$ and $\tilde{Y}_2$ are independent by construction the maximum likelihood estimator in the model (3.1) can be determined by solving the equation

$$\frac{\partial}{\partial\theta}\log\left\{ \prod_{i=1}^2 \frac{d\mathbb{P}_\theta^{(i)}}{d\mathbb{P}_0^{(i)}}(\tilde{Y}_i) \times \frac{d\mathbb{Q}_\theta^{(i)}}{d\mathbb{Q}_0^{(i)}}(\tilde{Y}_i(a)) \right\} = \sum_{i=1}^2 \left\{ \int_a^b \dot{\mathbf{F}}(t)\boldsymbol{\Sigma}_i^{-1/2}d\tilde{Y}_i(t) + \frac{\mathbf{F}(a)\boldsymbol{\Sigma}_i^{-1/2}\tilde{Y}_i(a)}{a} \right.$$
$$\left. - \left(\int_a^b \dot{\mathbf{F}}(t)\boldsymbol{\Sigma}_i^{-1/2}\boldsymbol{\Sigma}_i^{-1/2}\dot{\mathbf{F}}^\top(t)\,dt + \dot{\mathbf{F}}(a)\boldsymbol{\Sigma}_i^{-1/2}\boldsymbol{\Sigma}_i^{-1/2}\dot{\mathbf{F}}^\top(a) \right)\theta \right\} = 0$$

with respect to θ. The solution coincides with the linear estimate defined in (3.3), and a straight-forward calculation, using Ito's formula and the fact that the random variables $\int_a^b \dot{\boldsymbol{F}}(t)d\varepsilon_t$ and ε_a are independent, gives

$$
\begin{aligned}
\mathrm{Cov}(\hat{\theta}_{\mathrm{BLUE}}) &= \mathbf{M}^{-1}\mathbb{E}_\theta\Big[\Big(\int_a^b \dot{\mathbf{F}}(t)\boldsymbol{\Sigma}^{-1}\,d\mathbf{Y}(t) + \frac{1}{a}\mathbf{F}(a)\boldsymbol{\Sigma}^{-1}\mathbf{Y}(a)\Big) \\
&\quad \times \Big(\int_a^b \dot{\mathbf{F}}(t)\boldsymbol{\Sigma}^{-1}\,\mathbf{Y}(t) + \frac{1}{a}\mathbf{F}(a)\boldsymbol{\Sigma}^{-1}\mathbf{Y}(a)\Big)^\top\Big]\mathbf{M}^{-1} \\
&= \mathbf{M}^{-1}\Big(\int_a^b \dot{\mathbf{F}}(t)\boldsymbol{\Sigma}^{-1}\dot{\mathbf{F}}^\top(t)\,dt + \frac{1}{a}\mathbf{F}(a)\boldsymbol{\Sigma}^{-1}\mathbf{F}^\top(a)\Big)\mathbf{M}^{-1} = \mathbf{M}^{-1},
\end{aligned}
$$

where the matrix $\mathbf{M}$ is defined in (3.2). Since the covariance matrix $\mathbf{M}^{-1}$ is the inverse of the information matrix in the continuous time regression model in (3.1) (see Ibragimov and Has'minskii, 1981, p. 81), the linear estimator (3.3) is the BLUE, which completes the proof of Theorem 3.1. □

Remark 3.1 The proof of Theorem 3.1 can easily be modified to obtain the BLUE for the continuous time model on the interval $a = [0,b]$. More precisely, for $a = 0$ equation (3.7) becomes a deterministic equation equivalent to

$$\mathbf{Y}(0) = \mathbf{F}^\top(0)\theta , \tag{3.8}$$

and we have to distinguish three cases.

(1) If the regression function $\mathbf{F}$ satisfies $\mathbf{F}(0) = \mathbf{0}_{p\times 2}$ (that is $\mathrm{rank}(\mathbf{F}(0) = 0)$), the deterministic equation (3.8) does not contain any further information about the parameter θ and the maximum likelihood estimator in model (3.1) is given by

$$\hat{\theta}_{\mathrm{BLUE}} = \mathbf{M}_0^{-1}\Big(\int_0^b \dot{\mathbf{F}}(t)\boldsymbol{\Sigma}^{-1}\,d\mathbf{Y}(t)\Big),$$

where the minimum variance is given by

$$\mathrm{Cov}(\hat{\theta}_{\mathrm{BLUE}}) = \mathbf{M}_0^{-1} = \Big(\int_0^b \dot{\mathbf{F}}(t)\boldsymbol{\Sigma}^{-1}\dot{\mathbf{F}}^\top(t)\,dt\Big)^{-1}.$$

(2) If the rank of the matrix $\mathbf{F}(0)$ satisfies $\mathrm{rank}(\mathbf{F}(0)) = 1$, the deterministic equation (3.8) contains one informative equation about θ. In that case, we assume without loss of generality that $F_{1,1}(0) \neq 0$ and it follows by (3.8) that θ_1 can be reformulated by $\theta_2, \ldots, \theta_p$ through

$$\theta_1 = \frac{Y_1(0) - \sum_{i=j}^p \theta_j F_{1,j}(0)}{F_{1,1}(0)} . \tag{3.9}$$

9

Using (3.9) in combination with model (3.1), we obtain a reduced model by

$$\mathbf{Z}(t) = \mathbf{Y}(t) - \frac{Y_1(0)}{F_{1,1}(0)} \begin{pmatrix} F_{1,1}(0) \\ F_{2,1}(0) \end{pmatrix} = \tilde{\mathbf{F}}(t)\tilde{\theta} + \mathbf{\Sigma}^{1/2}\boldsymbol{\varepsilon}(t)\,, \tag{3.10}$$

where the matrix valued function $\tilde{\mathbf{F}}(t)$ is defined by

$$\tilde{\mathbf{F}}^T(t) = \left(F_{i,j}(t) - \frac{F_{i,1}(0)}{F_{1,1}(0)} F_{1,j}(0) \right)_{i=1,2, j=2,\dots p} \tag{3.11}$$

and the reduced $(p-1)$-dimensional parameter $\tilde{\theta}$ is given by $\tilde{\theta} = (\theta_2, \dots, \theta_p)$. It follows by $\mathrm{rank}(\mathbf{F}(0)) = 1$, that the matrix valued function $\tilde{\mathbf{F}}(t)$ defined in (3.11) satisfies $\tilde{\mathbf{F}}^T(\mathbf{0}) = 0_{2 \times p}$. Consequently, the modified model given by (3.10) satisfies the condition of the case given in (1) and the best linear unbiased estimator for the reduced parameter $\tilde{\theta}$ is obtained by

$$\hat{\tilde{\theta}}_{\mathrm{BLUE}} = \mathbf{M}_0^{-1}\left(\int_0^b \dot{\tilde{\mathbf{F}}}(t)\mathbf{\Sigma}^{-1}\, d\mathbf{Z}(t) \right), \tag{3.12}$$

where the process $\{\mathbf{Z}(\mathbf{t}); t \in [0, b]\}$ is defined by (3.10), the matrix $\tilde{\mathbf{F}}(t)$ is given by (3.11) and the minimum variance is given by

$$\mathrm{Cov}(\hat{\tilde{\theta}}_{\mathrm{BLUE}}) = \mathbf{M}_0^{-1} = \left(\int_0^b \dot{\tilde{\mathbf{F}}}(t)\mathbf{\Sigma}^{-1}\dot{\tilde{\mathbf{F}}}^\top(t)\, dt \right)^{-1}.$$

The best linear unbiased estimator for the remaining parameter θ_1 is then obtained by

$$\hat{\theta}_1 = \frac{Y_1(0) - \sum_{i=j}^p \hat{\tilde{\theta}}_{\mathrm{BLUE},j} F_{1,j}(0)}{F_{1,1}(0)}.$$

(3) If the rank of the matrix $\mathbf{F}(0)$ satisfies $\mathrm{rank}(\mathbf{F}(0)) = 2$, equation (3.8) contains two informative equations about θ.

Let

$$\mathbf{A}(t) = \begin{pmatrix} F_{1,1}(t) & F_{1,2}(t) \\ F_{2,1}(t) & F_{2,2}(t) \end{pmatrix} \tag{3.13}$$

be the submatrix of $\mathbf{F}$ which contains the first two columns of $\mathbf{F}^T(t)$. Without loss of generality, we assume that $\mathbf{A}(0)$ is non-singular (as $\mathrm{rank}(\mathbf{F}(0) = 2)$.

Then it follows by (3.8) that

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \mathbf{A}^{-1}(0)\left(\mathbf{Y}(0) - \left(\sum_{j=3}^p F_{i,j}(0)\theta_j \right)_{i=1,2} \right). \tag{3.14}$$

10

Using (3.14) in combination with (3.1) we obtain a reduced model given by

$$\mathbf{Z}(t) = \mathbf{Y}(t) - \mathbf{A}(t)\mathbf{A}^{-1}(0)\mathbf{Y}(0) = \tilde{\mathbf{F}}(t)\tilde{\theta} + \mathbf{\Sigma}^{1/2}\boldsymbol{\varepsilon}(t) \tag{3.15}$$

where the matrix valued function $\mathbf{A}(t)$ is given by (3.13), the matrix valued function $\tilde{\mathbf{F}}^T(t)$ is of the form

$$\tilde{\mathbf{F}}^T(t) = \mathbf{A}(t)\mathbf{A}^{-1}(0)\big(F_{i,j}(0)\big)_{i=1,2;j=3,\ldots p} + \big(F_{i,j}(t)\big)_{i=1,2;j=3,\ldots p} \tag{3.16}$$

and the reduced $(p-2)$-dimensional parameter $\tilde{\theta}$ is given by $\tilde{\theta} = (\theta_3, \ldots, \theta_p)$.

The matrix valued function $\tilde{\mathbf{F}}(t)$ defined in (3.16) satisfies $\tilde{\mathbf{F}}^{\mathbf{T}}(0) = \mathbf{0}_{2\times p}$. Consequently, the modified model given by (3.15) satisfies the condition of the case given in (1) and the best linear unbiased estimator $\hat{\tilde{\theta}}_{\mathrm{BLUE}}$ for the reduced $(p-2)$-dimensional parameter $\tilde{\theta}$ is obtained by (3.12) using the process $\{\mathbf{Z}(\mathbf{t}); t \in [0,b]\}$ defined by (3.15) and the matrix valued function $\tilde{\mathbf{F}}(t)$ given by (3.16). The best linear unbiased estimator for the remaining parameter $(\theta_1, \theta_2)^T$ is then obtained by

$$\begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} = \mathbf{A}^{-1}(0)\Big(\mathbf{Y}(0) - \big(\sum_{j=3}^{p} F_{i,j}(0)\hat{\tilde{\theta}}_{\mathrm{BLUE,j}}\big)_{i=1,2}\Big) .$$

3.2   Model with no common parameters

Recall the definition of model (2.1) in Section 1. It was demonstrated in Example 2.1 that this case is a special case of model (2.5), where the matrix $\mathbf{F}^\top$ is given by

$$\mathbf{F}^\top(t) = \begin{pmatrix} f_1^\top(t) & 0_{p_2}^\top \\ & \\ 0_{p_1}^\top & f_2^\top(t) \end{pmatrix} \tag{3.17}$$

and $\theta = (\theta^{(1)\top}, \theta^{(2)\top})^\top$. Considering both components in the vector $\mathbf{Y}$ separately, we obtain continuous versions of the two models introduced in (2.1), that is,

$$Y_i(t) = f_i^\top(t)\theta^{(i)} + \eta_i(t), \; i = 1, 2 \,, \tag{3.18}$$

where the error processes $\{\eta(t) \mid t \in [a,b]\}$ is defined by (2.4). An application of Theorem 3.1 yields the following BLUE.

Corollary 3.1 *Consider the continuous time linear regression model (2.5) on the interval $[a,b]$, with continuously differentiable matrix (3.17), a vector $\{\boldsymbol{\varepsilon}(t) = (\varepsilon_1(t), \varepsilon_2(t))^\top \mid t \in [a,b]\}$ of*

11

independent Brownian motions and a matrix $\mathbf{\Sigma}$ defined by (2.3). The best linear unbiased estimator for the parameter θ is given by

$$\hat{\theta}_{\text{BLUE}} = \begin{pmatrix} \hat{\theta}^{(1)}_{\text{BLUE}} \\ \hat{\theta}^{(2)}_{\text{BLUE}} \end{pmatrix} = \frac{1}{\sigma_1^2 \sigma_2^2 (1 - \varrho^2)} \mathbf{M}^{-1} \left\{ \int_a^b \begin{pmatrix} \sigma_2^2 \dot{f}_1(t) & -\sigma_1 \sigma_2 \varrho \dot{f}_1(t) \\ -\sigma_1 \sigma_2 \varrho \dot{f}_2(t) & \sigma_1^2 \dot{f}_2(t) \end{pmatrix} d \begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix} \right.$$

$$\left. + \frac{1}{a} \begin{pmatrix} \sigma_2^2 f_1(a) & -\sigma_1 \sigma_2 \varrho f_1(a) \\ -\sigma_1 \sigma_2 \varrho f_2(a) & \sigma_1^2 f_2(a) \end{pmatrix} \begin{pmatrix} Y_1(a) \\ Y_2(a) \end{pmatrix} \right\}.$$

(3.19)

The minimum variance is given by $\mathbf{M}^{-1}$, where

$$\mathbf{M} = \frac{1}{\sigma_1^2 \sigma_2^2 (1 - \varrho^2)} \begin{pmatrix} \sigma_2^2 \mathbf{M}_{11} & -\sigma_1 \sigma_2 \varrho \mathbf{M}_{12} \\ -\sigma_1 \sigma_2 \varrho \mathbf{M}_{21} & \sigma_1^2 \mathbf{M}_{22} \end{pmatrix}$$

and

$$\mathbf{M}_{ij} = \int_a^b \dot{f}_i(t) \dot{f}_j^\top(t) dt + \frac{1}{a} f_i(a) f_j^T(a) \qquad i,j = 1,2.$$

(3.20)

It is of interest to compare the estimator (3.19) with the estimator $\hat{\theta}_{\text{mar}} = ((\hat{\theta}^{(1)}_{\text{mar}})^\top, (\hat{\theta}^{(2)}_{\text{mar}})^\top)^\top$, which is obtained by estimating the parameter in both models (3.18) separately. It follows from Theorem 2.1 in Dette et al. (2017a) that the best linear unbiased estimators in these models are given by

$$\hat{\theta}^{(\ell)}_{\text{mar}} = \mathbf{M}_{\ell\ell}^{-1} \left(\int_a^b \dot{f}_\ell(t) dY_\ell(t) + \frac{1}{a} f_\ell(a) Y_\ell(a) \right), \qquad \ell = 1,2,$$

(3.21)

where the matrices are defined by

$$\mathbf{M}_{\ell\ell} = \int_a^b \dot{f}_\ell(t) \dot{f}_\ell^\top(t) dt + \frac{1}{a} f_\ell(a) f_\ell^\top(a), \quad \ell = 1,2.$$

Moreover, the covariance matrices of the estimators $\hat{\theta}^{(1)}_{\text{mar}}$ and $\hat{\theta}^{(2)}_{\text{mar}}$ are the inverses of the Fisher information matrices in the individual models, that is

$$\text{Cov}(\hat{\theta}^{(\ell)}_{\text{mar}}) = \sigma_\ell^2 \mathbf{M}_{\ell\ell}^{-1} \quad \ell = 1,2.$$

(3.22)

The following result compares the variance of the two estimators (3.19) and (3.21).

Theorem 3.2 *If the assumptions of Corollary 3.1 are satisfied, we have (with respect to the Loewner ordering)*

$$\text{Cov}(\hat{\theta}^{(\ell)}_{\text{BLUE}}) \leq \text{Cov}(\hat{\theta}^{(\ell)}_{\text{mar}}), \quad \ell = 1,2,$$

for all $\varrho \in (-1, 1)$, where the $\hat{\theta}^{(\ell)}_{\mathrm{BLUE}}$ and $\hat{\theta}^{(\ell)}_{\mathrm{mar}}$ are the best linear unbiased estimators of the parameter $\theta^{(\ell)}$ obtained by simultaneous estimation (see (3.19)) and separate estimation in the two groups (see (3.21)), respectively.

Proof. Without loss of generality we consider the case $\ell = 1$, the proof for the index $\ell = 2$ is obtained by the same arguments. Let $\mathbf{K_1}^\top = (\mathbf{I}_{p_1}, \mathbf{0}_{p_1 \times p_2})$ be a $p_1 \times (p_1 + p_2)$- matrix, where $\mathbf{I}_{p_1}$ and $\mathbf{0}_{p_1 \times p_2}$ denote the p_1-identity matrix and a $(p_1 \times p_2)$-matrix filled with zeros. Then,

$$\mathrm{Cov}(\hat{\theta}^{(\ell)}_{\mathrm{BLUE}}) = (\mathbf{C}_{\mathbf{K}_1}(\mathbf{M}))^{-1},$$

where

$$\mathbf{C}_{\mathbf{K}_1}(\mathbf{M}) = (\mathbf{K}_1^\top \mathbf{M}^{-1} \mathbf{K}_1)^{-1} = \frac{1}{\sigma_1^2(1 - \varrho^2)} \left(\mathbf{M}_{11} - \varrho^2 \mathbf{M}_{12} \mathbf{M}_{22}^{-1} \mathbf{M}_{12}^\top \right) \qquad (3.23)$$

is the Schur complement of the block $\mathbf{M}_{22}$ of the information matrix $\mathbf{M}$ (see p. 74 in Pukelsheim, 2006). Observing (3.22) we now compare $\mathbf{C}_{\mathbf{K}_1}(\mathbf{M})$ and $\frac{1}{\sigma^2} \mathbf{M}_{11}$ and obtain

$$\begin{aligned}
\mathbf{C}_{\mathbf{K}_1}(\mathbf{M}) - \frac{1}{\sigma_1^2} \mathbf{M}_{11} &= \frac{1}{\sigma_1^2(1 - \varrho^2)} \left(\mathbf{M}_{11} - \varrho^2 \mathbf{M}_{12} \mathbf{M}_{22}^{-1} \mathbf{M}_{12}^\top \right) - \frac{1}{\sigma_1^2} \mathbf{M}_{11} \\
&= \frac{\varrho^2}{\sigma_1^2(1 - \varrho^2)} \left(\mathbf{M}_{11} - \mathbf{M}_{12} \mathbf{M}_{22}^{-1} \mathbf{M}_{12}^\top \right) \qquad (3.24) \\
&:= \frac{\varrho^2}{\sigma_1^2(1 - \varrho^2)} \mathbf{C}_{\mathbf{K}_1}(\tilde{\mathbf{M}}),
\end{aligned}$$

where $\mathbf{C}_{\mathbf{K}_1}(\tilde{\mathbf{M}})$ is the Schur complement of the block $\mathbf{M}_{22}$ of the matrix

$$\tilde{\mathbf{M}} = \begin{pmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{pmatrix}.$$

Note that the matrix $\tilde{\mathbf{M}}$ is nonnegative definite. An application of Lemma 3.12 of Pukelsheim (2006) shows that the Schur complement $\mathbf{C}_{\mathbf{K}_1}(\tilde{\mathbf{M}})$ is also nonnegative definite, that is $\mathbf{C}_{\mathbf{K}_1}(\tilde{\mathbf{M}}) \geq 0$ with respect to the Loewner ordering. Observing (3.24) we have

$$\left(\mathrm{Cov}(\hat{\theta}^{(1)}_{\mathrm{BLUE}}) \right)^{-1} = \mathbf{C}_{\mathbf{K}_1}(\mathbf{M}) \geq \frac{1}{\sigma_1^2} \mathbf{M}_{11} = \left(\mathrm{Cov}(\hat{\theta}^{(1)}_{\mathrm{mar}}) \right)^{-1}$$

and the statement of the theorem follows. $\qquad \square$

Remark 3.2 If $\varrho = 0$ we have $\mathbf{C}_{K_1}(\mathbf{M}) = \mathbf{M}_{11}$, and it follows from (3.23) that separate estimation in the individual groups does not yield less precise estimates, that is $\mathrm{Cov}(\hat{\theta}^{(l)}_{\mathrm{mar}}) = \mathrm{Cov}(\hat{\theta}^{(1)}_{\mathrm{BLUE}})$ ($\ell = 1, 2$). However, in general we have $\mathrm{Cov}(\hat{\theta}^{(l)}_{\mathrm{mar}}) \geq \mathrm{Cov}(\hat{\theta}^{(1)}_{\mathrm{BLUE}})$. Moreover, the

inequality is strict in most cases, which means that simultaneous estimation of the parameters $\theta^{(1)}$ and $\theta^{(2)}$ yields more precise estimators. A necessary condition for strict inequality (i.e the matrix $\mathrm{Cov}(\hat{\theta}^{(l)}_{\mathrm{mar}}) - \mathrm{Cov}(\hat{\theta}^{(1)}_{\mathrm{BLUE}})$ is positive definite) is the condition $\varrho \neq 0$. The following result shows that this condition is not sufficient. It considers the important case where the regression functions f_1 and f_2 in (3.17) are the same and shows that in this case the two estimators $\hat{\theta}_{\mathrm{BLUE}}$ and $\hat{\theta}_{\mathrm{mar}}$ coincide.

Corollary 3.2 *If the assumptions of Corollary 3.1 hold and additionally the regression functions in model* (2.5) *satisfy $f_1 = f_2$, the best linear unbiased estimator for the parameter θ is given by*

$$\hat{\theta}_{\mathrm{BLUE}} = \begin{pmatrix} \hat{\theta}^{(1)}_{\mathrm{BLUE}} \\ \hat{\theta}^{(2)}_{\mathrm{BLUE}} \end{pmatrix} = \int_a^b \left(\mathbf{I}_2 \otimes \mathbf{M}_{11}^{-1} \dot{f}_1(t) \right) d\mathbf{Y}(t) + \frac{1}{a} \left(\mathbf{I}_2 \otimes \mathbf{M}_{11}^{-1} f_1(a) \right) \mathbf{Y}(a),$$

where $\mathbf{I}_2$ denotes the 2×2-identity matrix and the matrix $\mathbf{M}_{11}$ is defined by (3.27). *Moreover, the minimum variance is given by $\mathrm{Cov}(\hat{\theta}_{\mathrm{BLUE}}) = \mathbf{\Sigma} \otimes \mathbf{M}_{11}^{-1}$ and*

$$\mathrm{Cov}(\hat{\theta}^{(l)}_{\mathrm{mar}}) = \mathrm{Cov}(\hat{\theta}^{(1)}_{\mathrm{BLUE}}) \quad (\ell = 1, 2) .$$

3.3 Models with common parameters

Recall the definition of model (2.1) in Section 1. It was demonstrated in Example 2.2 that this case is a special case of model (2.5), where the matrix of regression functions is given by

$$\mathbf{F}^\top(t) = \begin{pmatrix} f_0^\top(t), \tilde{f}_1^\top(t), 0_{p_2 - p_0}^\top \\ f_0^\top(t), 0_{p_1 - p_0}^\top, \tilde{f}_2^\top(t) \end{pmatrix} \tag{3.25}$$

and the vector of parameters is defined by

$$\theta = (\theta^{(0)\top}, \tilde{\theta}^{(1)\top}, \tilde{\theta}^{(2)\top})^\top .$$

An application of Theorem 3.1 yields the BLUE in model (2.5) with the matrix $\mathbf{F}^\top$ defined by (3.25).

Corollary 3.3 *Consider the continuous time linear regression model* (2.5) *on the interval $[a, b]$, where the the matrix of regression functions $\mathbf{F}^\top$ is continuously differentiable. The best linear unbiased estimator for the parameter θ is given by*

14

$$\hat{\theta}_{\text{BLUE}} = \begin{pmatrix} \hat{\theta}_{\text{BLUE}}^{(0)} \\ \hat{\theta}_{\text{BLUE}}^{(1)} \\ \hat{\theta}_{\text{BLUE}}^{(2)} \end{pmatrix} = \frac{1}{\sigma_1^2 \sigma_2^2 (1-\varrho^2)} \mathbf{M}^{-1} \left\{ \int_a^b \begin{pmatrix} (\sigma_2^2 - \sigma_1\sigma_2\varrho)\dot{f}_0(t) & (\sigma_1^2 - \sigma_1\sigma_2\varrho)\dot{f}_0(t) \\ \sigma_2^2 \dot{\tilde{f}}_1(t) & -\sigma_1\sigma_2\varrho \dot{\tilde{f}}_1(t) \\ -\sigma_1\sigma_2\varrho \dot{\tilde{f}}_2(t) & \sigma_1^2 \dot{\tilde{f}}_2(t) \end{pmatrix} d \begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix} \right.$$

$$\left. + \frac{1}{a} \begin{pmatrix} (\sigma_2^2 - \sigma_1\sigma_2\varrho)f_0(a) & (\sigma_1^2 - \sigma_1\sigma_2\varrho)f_0(a) \\ \sigma_2^2 \tilde{f}_1(a) & -\sigma_1\sigma_2\varrho \tilde{f}_1(a) \\ -\sigma_1\sigma_2\varrho \tilde{f}_2(a) & \sigma_1^2 \tilde{f}_2(a) \end{pmatrix} \begin{pmatrix} Y_1(a) \\ Y_2(a) \end{pmatrix} \right\}. \tag{3.26}$$

The minimum variance is

$$\text{Cov}(\hat{\theta}_{\text{BLUE}}) = \mathbf{M}^{-1} \,,$$

where

$$\mathbf{M} = \frac{1}{\sigma_1^2 \sigma_2^2 (1-\varrho^2)} \begin{pmatrix} (\sigma_1^2 + \sigma_2^2 - \sigma_1\sigma_2\varrho)\mathbf{M}_{00} & (\sigma_2^2 - \sigma_1\sigma_2\varrho)\mathbf{M}_{01} & (\sigma_1^2 - \sigma_1\sigma_2\varrho)\mathbf{M}_{02} \\ (\sigma_2^2 - \sigma_1\sigma_2\varrho)\mathbf{M}_{10} & \sigma_2^2 \mathbf{M}_{11} & -\sigma_1\sigma_2\varrho \mathbf{M}_{12} \\ (\sigma_1^2 - \sigma_1\sigma_2\varrho)\mathbf{M}_{20} & -\sigma_1\sigma_2\varrho \mathbf{M}_{21} & \sigma_1^2 \mathbf{M}_{22} \end{pmatrix}$$

and individual blocks in this matrix are given by

$$\mathbf{M}_{ij} = \int_a^b \dot{g}_i(t)\dot{g}_j^\top(t)dt + \frac{1}{a}g_i(a)g_j^\top(a) \,, \tag{3.27}$$

for $i,j = 0,1,2$, where $g_0(t) = f_0(t)$ and $g_i(t) = \tilde{f}_i(t)$ for $i = 1,2$.

It is again of interest to compare the estimate (3.26) with the estimate $\hat{\theta}_{\text{mar}} = ((\hat{\theta}_{\text{mar}}^{(1)})^\top, (\hat{\theta}_{\text{mar}}^{(2)})^\top)^\top$, which is obtained by estimating the parameter $\theta^{(1)} = ((\theta^{(0)})^\top, (\tilde{\theta}^{(1)})^\top)^\top$ in both models (3.18) separately by using (3.21). The corresponding covariances of the estimators $\hat{\theta}_{\text{mar}}^{(1)}$ and $\hat{\theta}_{\text{mar}}^{(2)}$ are given by (3.22). The following result compares the variance of the two estimators (3.26) and (3.21). Its proof is similar to the proof of Theorem 3.2 and therefore omitted.

Theorem 3.3 *If the assumptions of Corollary 3.3 are satisfied, we have (with respect to the Loewner ordering)*

$$\text{Cov}(\hat{\theta}_{\text{BLUE}}^{(\ell)}) \leq \text{Cov}(\hat{\theta}_{\text{mar}}^{(\ell)}) \,, \quad \ell = 1,2 \,,$$

for all $\varrho \in (-1,1)$, where the $\hat{\theta}_{\text{BLUE}}^{(\ell)}$ and $\hat{\theta}_{\text{mar}}^{(\ell)}$ are the best linear unbiased estimators of the parameter $\theta^{(\ell)}$ obtained by simultaneous and separate estimation, respectively.

Remark 3.3 The results presented so far have been derived for the case where the error process $\{\boldsymbol{\varepsilon}(t) = (\varepsilon_1(t), \varepsilon_2(t))^\top \,|\, t \in [a, b]\}$ in (2.5) consist of two independent Brownian motions. This assumption has been made to simplify the notation. Similar results can be obtained for Markov processes and in this remark we indicate the essential arguments.

To be precise, assume that the error processes $\{\boldsymbol{\varepsilon}(t) = (\varepsilon_1(t), \varepsilon_2(t))^\top \,|\, t \in [a, b]\}$ in model (2.5) consist of two independent centered Gaussian processes with continuous covariance kernel given by

$$K(s, t) = \mathbb{E}[\varepsilon_i(s)\varepsilon_i(t)] = u(s)v(t)\min\{q(s), q(t)\} \quad s, t \in [a, b]\,, \tag{3.28}$$

where $u(\cdot)$ and $v(\cdot)$ are functions defined on the interval $[a, b]$ such that the function $q(\cdot) = u(\cdot)/v(\cdot)$ is positive and strictly increasing. Kernels of the form (3.28) are called *triangular* kernels and a famous result in Doob (1949) essentially shows that a Gaussian process is a Markov process if and only if its covariance kernel is triangular (see also Mehr and McFadden, 1965). In this case model (2.5) can be transformed into a model with an error process consisting of two independent Brownian motions using the arguments given in Appendix B of Dette et al. (2016). More precisely, define

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

and consider the stochastic process

$$\boldsymbol{\varepsilon}(t) = v(t)\tilde{\boldsymbol{\varepsilon}}(q(t)),$$

where $\{\tilde{\boldsymbol{\varepsilon}}(\tilde{t}) = (\tilde{\varepsilon}_1(\tilde{t})^\top, \tilde{\varepsilon}_2(\tilde{t}))\,|\, \tilde{t} \in [\tilde{a}, \tilde{b}]\}$ consists of two independent Brownian motions on the interval $[\tilde{a}, \tilde{b}] = [q(a), q(b)]$. It now follows from Doob (1949) that the process $\{\boldsymbol{\varepsilon}(t) = (\varepsilon_1(t), \varepsilon_2(t))^\top \,|\, t \in [a, b]\}$ consists of two independent centered Gaussian process on the interval $[a, b]$ with covariance kernel (3.28). Consequently, if we consider the model

$$\tilde{\boldsymbol{Y}}(\tilde{t}) = \begin{pmatrix} \tilde{Y}_1(\tilde{t}) \\ \tilde{Y}_2(\tilde{t}) \end{pmatrix} = \tilde{\mathbf{F}}^\top(\tilde{t})\theta + \boldsymbol{\Sigma}^{1/2}\tilde{\boldsymbol{\varepsilon}}(\tilde{t})\,, \ \tilde{t} \in [q(a), q(b)], \tag{3.29}$$

and

$$\tilde{\mathbf{F}}(\tilde{t}) = \frac{\mathbf{F}(q^{-1}(\tilde{t}))}{v(q^{-1}(\tilde{t}))} \ , \ \tilde{\boldsymbol{\varepsilon}}(\tilde{t}) = \frac{\boldsymbol{\varepsilon}(q^{-1}(\tilde{t}))}{v(q^{-1}(\tilde{t}))} \ , \ \tilde{\boldsymbol{Y}}(\tilde{t}) = \frac{\boldsymbol{Y}(q^{-1}(\tilde{t}))}{v(q^{-1}(\tilde{t}))} \ ,$$

the results obtained so far are applicable. Thus, a "good" estimator obtained for the parameter θ in model (3.29) is also a "good estimator" for the parameter θ in model (3.1) with error process consisting of two Gaussian processes with covariance kernel (3.28). Consequently, we can derive the optimal estimator for the parameter θ in the continuous time model (3.1) with covariance kernel (3.28) from the best linear unbiased estimator in the model given in (3.29)

16

with Brownian motions by an application of Theorem 3.1. The resulting best linear unbiased estimator for θ in model (3.1) with triangular kernel (3.28) is of the form

$$\hat{\theta}_{\text{BLUE}} = \mathbf{M}^{-1}\bigg\{ \int_a^b \frac{\dot{\mathbf{F}}(t)v(t) - \mathbf{F}(t)v(t)}{\dot{u}(t)v(t) - u(t)\dot{v}(t)} \mathbf{\Sigma}^{-1} d\left(\frac{\mathbf{Y}(t)}{v(t)} \right) + \frac{\mathbf{F}(a)\mathbf{\Sigma}^{-1}\mathbf{Y}(a)}{u(a)v(a)} \bigg\},$$

where the minimum variance is given by

$$\mathbf{M}^{-1} = \left(\int_a^b \frac{\left(\dot{\mathbf{F}}(t)v(t) - \mathbf{F}(t)\dot{v}(t) \right) \mathbf{\Sigma}^{-1} \left(\dot{\mathbf{F}}(t)v(t) - \mathbf{F}(t)\dot{v}(t) \right)^{\top}}{v^2(t)[\dot{u}(t)v(t) - u(t)\dot{v}(t)]} dt + \frac{\mathbf{F}(a)\mathbf{\Sigma}^{-1}\mathbf{F}^{\top}(a)}{u(a)v(a)} \right)^{-1}.$$

4 Optimal designs for comparing curves

In this section we will derive optimal designs for comparing curves. The first part is devoted to a discretization of the BLUE in the continuous time model. In the second part we develop an optimality criterion to obtain efficient designs for the comparison of curves based on the discretized estimators.

4.1 From the continuous to the discrete model

To obtain a discrete design for n observations at the points $a = t_1, \ldots, t_n$ from the continuous design derived in Section 3, we use a similar approach as in Dette et al. (2017a) and construct a discrete approximation of the stochastic integral in (3.3). For this purpose we consider the linear estimator

$$\begin{aligned} \hat{\theta}_n &= \mathbf{M}^{-1}\bigg\{ \sum_{i=2}^{n} \mathbf{\Omega}_i \dot{\mathbf{F}}(t_{i-1}) \mathbf{\Sigma}^{-1}(Y(t_i) - Y(t_{i-1})) + \frac{\mathbf{F}(a)}{a} \mathbf{\Sigma}^{-1} Y_a \bigg\} \qquad (4.1) \\ &= \mathbf{M}^{-1}\bigg\{ \sum_{i=2}^{n} \mathbf{\Phi}_i \mathbf{\Sigma}^{-1}(Y(t_i) - Y(t_{i-1})) + \frac{\mathbf{F}(a)}{a} \mathbf{\Sigma}^{-1} Y_a \bigg\}, \end{aligned}$$

were $a = t_1 < t_2 < \ldots < t_{n-1} < t_n = b$, $\mathbf{\Omega}_2, \ldots, \mathbf{\Omega}_n$ are $p \times p$ weight matrices and $\mathbf{\Phi}_2 = \mathbf{\Omega}_2 \dot{\mathbf{F}}(t_1), \ldots, \mathbf{\Phi}_n = \mathbf{\Omega}_n \dot{\mathbf{F}}(t_{n-1})$ are $p \times 2$ matrices, which have to be chosen in a reasonable way. The matrix $\mathbf{M}^{-1}$ is given in (3.4). To determine these weights in an "optimal" way we first derive a representation of the mean squared error between the best linear estimate (3.3) in the continuous time model and its discrete approximation (4.1). The following result is a direct consequence of Ito's formula.

Lemma 4.1 *Consider the continuous time model (3.1). If the assumptions of Theorem 3.1 are satisfied, we have*

$$\mathbb{E}_\theta[(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)^\top] = \mathbf{M}^{-1}\Big\{ \sum_{i=2}^n \int_{t_{i-1}}^{t_i} \big[\dot{\mathbf{F}}(s) - \mathbf{\Phi}_i\big]\mathbf{\Sigma}^{-1}\big[\dot{\mathbf{F}}(s) - \mathbf{\Phi}_i\big]^\top ds$$

$$+ \sum_{i,j=2}^n \int_{t_{i-1}}^{t_i} \big[\dot{\mathbf{F}}(s) - \mathbf{\Phi}_i\big]\mathbf{\Sigma}^{-1}\dot{\mathbf{F}}^\top(s)\, ds\, \theta\, \theta^\top \int_{t_{j-1}}^{t_j} \dot{\mathbf{F}}(s)\mathbf{\Sigma}^{-1}\big[\dot{\mathbf{F}}(s) - \mathbf{\Phi}_i\big]^\top ds \Big\}\mathbf{M}^{-1}. \quad (4.2)$$

In the following we choose optimal $p \times 2$ matrices $\mathbf{\Phi}_i = \mathbf{\Omega}_i\dot{\mathbf{F}}(t_{i-1})$ and design points $t_2, \ldots, t_{n-1}$ ($t_1 = a, t_n = b$), such that the linear estimate (4.1) is unbiased and the mean squared error matrix in (4.2) "becomes small". An alternative criterion is to replace the mean squared error $\mathbb{E}_\theta[(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)^\top]$ by the mean squared error

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

between the estimate $\hat{\theta}_n$ defined in (4.1) and the "true" vector of parameters. The following result shows that in the class of unbiased estimators both optimization problems yield the same solution. The proof is similar to the proof of Theorem 3.1 in Dette et al. (2017a).

Theorem 4.1 *The estimator $\hat{\theta}_n$ defined in (4.1) is unbiased if and only if the identity*

$$\mathbf{M}_0 = \int_a^b \dot{\mathbf{F}}(s)\mathbf{\Sigma}^{-1}\dot{\mathbf{F}}^\top(s)\, ds = \sum_{i=2}^n \mathbf{\Phi}_i\mathbf{\Sigma}^{-1} \int_{t_{i-1}}^{t_i} \dot{\mathbf{F}}^\top(s)\, ds = \sum_{i=2}^n \mathbf{\Phi}_i\mathbf{\Sigma}^{-1}(\mathbf{F}(t_i) - \mathbf{F}(t_{i-1}))^\top, \quad (4.3)$$

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

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

In order to describe a solution in terms of optimal "weights" $\mathbf{\Phi}_i^*$ and design points t_i^* we recall that the condition of unbiasedness of the estimate $\hat{\theta}_n$ in (4.1) is given by (4.3) and introduce the notation

$$\mathbf{B}_i = [\mathbf{F}(t_i) - \mathbf{F}(t_{i-1})]\mathbf{\Sigma}^{-1/2}/\sqrt{t_i - t_{i-1}}, \quad (4.4)$$
$$\mathbf{A}_i = \mathbf{\Phi}_i\mathbf{\Sigma}^{-1/2}\sqrt{t_i - t_{i-1}}.$$

It follows from Lemma 4.1 and Theorem 4.1 that for an unbiased estimator $\hat{\theta}_n$ of the form (4.1) the mean squared error has the representation

$$\mathbb{E}_\theta\big[(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)^\top(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)\big] = -\mathbf{M}^{-1}\mathbf{M}_0\mathbf{M}^{-1} + \sum_{i=2}^n \mathbf{M}^{-1}\mathbf{A}_i\mathbf{A}_i^\top\mathbf{M}^{-1}, \quad (4.5)$$

18

which has to be "minimized" subject to the constraint

$$\mathbf{M}_0 = \int_a^b \dot{\mathbf{F}}(s)\mathbf{\Sigma}^{-1}\dot{\mathbf{F}}^\top(s)ds = \sum_{i=2}^n \mathbf{A}_i\mathbf{B}_i^\top. \tag{4.6}$$

The following result shows that a minimization with respect to the weights $\mathbf{\Phi}_i$ (or equivalently $\mathbf{A}_i$) can actually be carried out with respect to the Loewner ordering.

Theorem 4.2 *Assume that the assumptions of Theorem 3.1 are satisfied and that the matrix*

$$\mathbf{B} = \sum_{i=2}^n \mathbf{B}_i\mathbf{B}_i^\top = \sum_{i=2}^n \frac{[\mathbf{F}(t_i) - \mathbf{F}(t_{i-1})]\mathbf{\Sigma}^{-1}[\mathbf{F}(t_i) - \mathbf{F}(t_{i-1})]^\top}{t_i - t_{i-1}}, \tag{4.7}$$

is non-singular. Let $\mathbf{\Phi}_2^*, \ldots, \mathbf{\Phi}_n^*$ denote $(p \times 2)$-matrices satisfying the equations

$$\mathbf{\Phi}_i^* = \mathbf{M}_0\mathbf{B}^{-1}\frac{\mathbf{F}(t_i) - \mathbf{F}(t_{i-1})}{t_i - t_{i-1}} \qquad i = 2, \ldots, n, \tag{4.8}$$

then $\mathbf{\Phi}_2^*, \ldots, \mathbf{\Phi}_n^*$ are optimal weight matrices minimizing $\mathbb{E}_\theta[(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)^\top]$ with respect to the Loewner ordering among all unbiased estimators of the form (4.1). *Moreover, the variance of the resulting estimator $\hat{\theta}_n^*$ is given by*

$$\mathrm{Cov}(\hat{\theta}_n^*) = \mathbf{M}^{-1}\left\{\mathbf{M}_0\mathbf{B}^{-1}\mathbf{M}_0 + \frac{1}{a}\mathbf{F}(a)\mathbf{\Sigma}^{-1}\mathbf{F}^\top(a)\right\}\mathbf{M}^{-1}$$

Proof. Let v denote a p-dimensional vector and consider the problem of minimizing the criterion

$$v^\top\mathbb{E}_\theta[(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)^\top]v \tag{4.9}$$

subject to the constraint (4.6). Observing (4.5) this yields the Lagrange function

$$G_v(\mathbf{A}_1 \ldots, \mathbf{A}_n) = -v^\top\mathbf{M}^{-1}\mathbf{M}_0\mathbf{M}^{-1}v + \sum_{i=2}^n(v^\top\mathbf{M}^{-1}\mathbf{A}_i\mathbf{A}_i^\top\mathbf{M}^{-1}v) - \mathrm{tr}\Big\{\mathbf{\Lambda}(\mathbf{M}_0 - \sum_{i=2}^n\mathbf{A}_i\mathbf{B}_i^\top)\Big\}, \tag{4.10}$$

where $\mathbf{A}_2, \ldots, \mathbf{A}_n$ are $(p \times 2)$-matrices and $\mathbf{\Lambda} = (\lambda_{k,\ell})_{k,\ell=1}^p$ is a $(p \times p)$-matrix of Lagrange multipliers. The function G_v is convex with respect to $\mathbf{A}_2, \ldots, \mathbf{A}_n$. Therefore, taking derivatives with respect to $\mathbf{A}_j$ yields as necessary and sufficient for the extremum (here we use matrix differential calculus)

$$2(\mathbf{M}^{-1}v)^\top\mathbf{A}_i \otimes (\mathbf{M}^{-1}v)^\top + \mathrm{vec}\{\mathbf{\Lambda}\mathbf{B}_i\} = 0_{2p}^\top, \qquad i = 2, \ldots, n.$$

Rewriting this system of linear equations in a $(p \times 2)$-matrix form gives

$$2\mathbf{M}^{-1}vv^\top\mathbf{M}^{-1}\mathbf{A}_i = -\mathbf{\Lambda}\mathbf{B}_i \qquad i = 2, \ldots, n.$$

19

Substituting the expression in (4.6) and using the non-singularity of the matrices $\mathbf{M}$ and $\mathbf{B}$ yields for the matrix of Lagrangian multipliers

$$\mathbf{\Lambda} = -2\mathbf{M}^{-1}vv^{\top}\mathbf{M}^{-1}\mathbf{M}_0\mathbf{B}^{-1} \; ,$$

which gives

$$2\mathbf{M}^{-1}vv^{\top}\mathbf{M}^{-1}\mathbf{A}_i = 2\mathbf{M}^{-1}vv^{\top}\mathbf{M}^{-1}\mathbf{M}_0\mathbf{B}^{-1}\mathbf{B}_i \qquad i = 2,\ldots,n\,. \tag{4.11}$$

Note that one solution of (4.11) is given by

$$\mathbf{A}_i^* = \mathbf{M}_0\mathbf{B}^{-1}\mathbf{B}_i, \qquad i = 2,\ldots,n$$

which does not depend on the vectors v. Therefore, the tuple of matrices $(\mathbf{A}_2^*,\ldots,\mathbf{A}_n^*)$ minimizes the convex function G_v in (4.10) for all $v \in \mathbb{R}^p$.

Observing the notations in (4.4) shows that the optimal matrix weights are given by (4.8). Moreover, these weights in (4.8) do not depend on the vector v either and provide a simultaneous minimizer of the criterion defined in (4.9) for all $v \in \mathbb{R}^p$. Consequently, the weights defined in (4.8) minimize $\mathbb{E}_\theta[(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)(\hat{\theta}_{\mathrm{BLUE}} - \hat{\theta}_n)^{\top}]$ under the unbiasedness constraint (4.6) with respect to the Loewner ordering. $\qquad\square$

Remark 4.1 If the matrix $\mathbf{B}$ in Theorem 4.2 is singular, the optimal weights are not uniquely determined and we propose to replace the inverse B by its Moore-Penrose inverse.

Note that for fixed design points $t_1,\ldots,t_n$ Theorem 4.2 yields universally optimal weights $\mathbf{\Phi}_2^*,\ldots,\mathbf{\Phi}_n^*$ (with respect to the Loewner ordering) for estimators of the form (4.1) satisfying (4.3). On the other hand, a further optimization with respect to the Loewner ordering with respect to the choice of the points $t_2,\ldots,t_{n-1}$ ($t_1 = a, t_n = b$) is not possible, and we have to apply a real valued optimality criterion for this purpose. In the following section, we will derive such a criterion which explicitly addresses the comparison of the regression curves from the two groups introduced in Section 2.

4.2 Confidence bands

We return to the practical scenario of the two groups introduced in (2.1), where we now focus on the comparison of these groups on the interval $[a, b]$.

More precisely, consider the model introduced in (2.5) and let $\hat{\theta}_n^*$ be the estimator (4.1) with optimal weights defined by (4.8) from n observations taken at the time points $a = t_1 < t_2 < \ldots < t_{n-1} < t_n = b$. Then this estimator is normally distributed with mean $\mathbb{E}[\hat{\theta}_n^*] = \theta$ and covariance matrix

$$\mathrm{Cov}(\hat{\theta}_n^*) = \mathbf{M}^{-1}\left\{\mathbf{M}_0\mathbf{B}^{-1}\mathbf{M}_0 + \frac{1}{a}\mathbf{F}(a)\mathbf{\Sigma}^{-1}\mathbf{F}^{\top}(a)\right\}\mathbf{M}^{-1}$$

where the matrices $\mathbf{M}, \mathbf{M}_0$ and $\mathbf{B}$ are given by (3.2), (4.6) and (4.7), respectively. Note that the covariance matrix depends on the time points $t_1, \ldots, t_n$ through the matrix $\mathbf{B}^{-1}$. Moreover, using the estimator $\hat{\theta}_n^*$ the prediction of the difference of a fixed time point $t \in [a, b]$ satisfies

$$(1, -1)\mathbf{F}^\top(t)\hat{\theta}_n^* - (1, -1)\mathbf{F}^\top(t)\theta \sim \mathcal{N}_p(0, h(t; t_1, \ldots, t_n)) \ ,$$

where

$$h(t; t_1, \ldots, t_n) = (1, -1)\mathbf{F}^\top \mathbf{M}^{-1} \left\{ \mathbf{M}_0 \mathbf{B}^{-1} \mathbf{M}_0 + \frac{1}{a}\mathbf{F}(a)\mathbf{\Sigma}^{-1}\mathbf{F}^\top(a) \right\} \mathbf{M}^{-1}\mathbf{F}(t)(1, -1)^T .$$

We now use this result and the results of Gsteiger et al. (2011) to obtain a simultaneous confidence band for the difference of the two curves. More precisely, if the interval $[a, b]$ is the range where the two curves should be compared, the simultaneous confidence band is defined as follows. Consider the statistic

$$\hat{T} = \sup_{t \in [a,b]} \frac{|(1, -1)\mathbf{F}^\top(t)\hat{\theta}_n^* - (1, -1)\mathbf{F}^\top(t)\theta|}{\{h(t; t_1, \ldots, t_n)\}^{1/2}},$$

and define D as the $(1 - \alpha)$-quantile of the corresponding distribution, that is

$$\mathbb{P}(\hat{T} \leq D) = 1 - \alpha.$$

Note that Gsteiger et al. (2011) propose the parametric bootstrap for choosing the critical value D. Define

$$
\begin{aligned}
u(t; t_1, \ldots, t_n) &= (1, -1)\mathbf{F}^\top(t)\hat{\theta}_n^* + D \cdot \{h(t; t_1, \ldots, t_n)\}^{1/2}, \\
l(t; t_1, \ldots, t_n) &= (1, -1)\mathbf{F}^\top(t)\hat{\theta}_n^* - D \cdot \{h(t; t_1, \ldots, t_n)\}^{1/2},
\end{aligned}
$$

then the confidence band for the difference of the two regression functions is defined by

$$\mathcal{C}_{1-\alpha} = \left\{ g : [a, b] \to \mathbb{R} \mid l(t; t_1, \ldots, t_n) \leq g(t) \leq u(t; t_1, \ldots, t_n) \text{ for all } t \in [a, b] \right\} . \quad (4.12)$$

Consequently, good time points $t_1 = a < t_2 < \ldots < t_{n-1}, t_n = b$ should minimize the width

$$u(t; t_1, \ldots, t_n) - l(t; t_1, \ldots, t_n) = 2 \cdot D \cdot \{h(t; t_1, \ldots, t_n)\}^{1/2}$$

of this band at each $t \in [a, b]$. As this is only possible in rare circumstances, we propose to minimize an L_p-norm of the function $h(\cdot; t_1 \ldots, t_n)$ as a design criterion, that is

$$\Phi_p(t_1, \ldots, t_n) = \|h(\cdot; t_1 \ldots, t_n)\|_p := \left(\int_a^b [h(t; t_1 \ldots, t_n)]^p \right)^{1/p} dt, \quad 1 \leq p \leq \infty, \quad (4.13)$$

where the case $p = \infty$ corresponds to the maximal deviation

$$\|h(\cdot; t_1 \ldots, t_n)\|_\infty = \sup_{t \in [a,b]} |h(t; t_1 \ldots, t_n)|.$$

Finally, the optimal points $a = t_1^* < t_2^* < \ldots < t_n^* = b$ (minimizing (4.13)) and the corresponding weights derived in Theorem 4.2 provide the optimal linear unbiased estimator of the form (4.1) (with the corresponding optimal design).

Example 4.1 We now conclude this section by considering the cases of no common and common parameters, respectively.

(a) If we are in the situation of Example 2.1 (no common parameters), the regression function $\mathbf{F}^\top(t)$ is of the form in (3.17) and the variance of the prediction of the difference at a fixed point $t \in [a, b]$ reduces to

$$h(t; t_1, \ldots, t_n) = (f_1^\top(t), -f_2^\top(t))\mathbf{M}^{-1}\left\{\mathbf{M}_0\mathbf{B}^{-1}\mathbf{M}_0 + \frac{1}{a}\mathbf{F}(a)\mathbf{\Sigma}^{-1}\mathbf{F}^\top(a)\right\}\mathbf{M}^{-1}(f_1^\top(t), -f_2^\top(t))^\top.$$

The corresponding design criterion is given by

$$\Phi_p(t_1, \ldots, t_n) = \|(f_1^\top, -f_2^\top)\mathbf{M}^{-1}\left\{\mathbf{M}_0\mathbf{B}^{-1}\mathbf{M}_0 + \frac{1}{a}\mathbf{F}(a)\mathbf{\Sigma}^{-1}\mathbf{F}^\top(a)\right\}\mathbf{M}^{-1}(f_1^\top, -f_2^\top)^\top\|_p.$$
$$(4.14)$$

(b) If we are in the situation of Example 2.2 (common parameters), the regression function $\mathbf{F}^\top(t)$ is given by (3.25) and the variance of the prediction of the difference at a fixed point $t \in [a, b]$ reduces to

$$h(t; t_1, \ldots, t_n) = (0, \tilde{f}_1^\top(t), -\tilde{f}_2^\top(t))\mathbf{M}^{-1}\left\{\mathbf{M}_0\mathbf{B}^{-1}\mathbf{M}_0 + \frac{1}{a}\mathbf{F}(a)\mathbf{\Sigma}^{-1}\mathbf{F}^\top(a)\right\}\mathbf{M}^{-1}(0, \tilde{f}_1^\top(t), -\tilde{f}_2^\top(t))^\top.$$

The corresponding design criterion is given by

$$\Phi_p(t_1, \ldots, t_n) = \|(0_{p_0}^\top, \tilde{f}_1^\top, -\tilde{f}_2^\top)\mathbf{M}^{-1}\left\{\mathbf{M}_0\mathbf{B}^{-1}\mathbf{M}_0 + \frac{1}{a}\mathbf{F}(a)\mathbf{\Sigma}^{-1}\mathbf{F}^\top(a)\right\}\mathbf{M}^{-1}(0_{p_0}^\top, \tilde{f}_1^\top, -\tilde{f}_2^\top)^\top\|_p.$$

5   Numerical Examples

In this section the methodology is illustrated in examples by means of a simulation study. To be precise, we consider the regression model (2.5), where the matrix $\mathbf{F}(t)$ is given by (3.17) corresponding to the case that the regression function do not share common parameters, see Section 3.2 for more details. In this case the corresponding bounds for the confidence band is given by (4.12), where

$$u(t; t_1, \ldots, t_n) = (\hat{\theta}_n^{*(1)})^\top f^{(1)}(t) - (\hat{\theta}_n^{*(2)})^\top f^{(2)}(t) + D \cdot \{h(t; t_1, \ldots, t_n)\}^{1/2},$$
$$l(t; t_1, \ldots, t_n) = (\hat{\theta}_n^{*(1)})^\top f^{(1)}(t) - (\hat{\theta}_n^{*(2)})^\top f^{(2)}(t) - D \cdot \{h(t; t_1, \ldots, t_n)\}^{1/2},$$

and $\hat{\theta}_n^* = ((\hat{\theta}_n^{*(1)})^\top, (\hat{\theta}_n^{*(2)})^\top)^\top$ is the estimator (4.1) with optimal weights defined in (4.8). The design space is given by the interval $[a, b] = [1, 10]$, and we consider three choices for the functions f_1 and f_2 in the matrix (3.17), that is

$$
\begin{aligned}
f_A(t) &= (t, \sin(t), \cos(t))^\top, \\
f_B(t) &= (t^2, \cos(t), \cos(2t))^\top, \\
f_C(t) &= \left(t, \log(t), \frac{1}{t}\right)^\top.
\end{aligned}
\tag{4.15}
$$

To model the dependence between the two groups we use the covariance matrix

$$
\Sigma = \begin{pmatrix} 1 & \varrho \\ \varrho & 1 \end{pmatrix},
$$

in (2.5), where the correlations are chosen as $\varrho = 0.2, 0.5, 0.7$. Following the discussion in Section 4.1 we focus on the comparison of the regression curves for the two groups and derive optimal designs, minimizing the criterion Φ_∞ defined in (4.14). As result, we obtain simultaneous confidence bands with a smaller maximal width for the difference of the curves describing the relation in the two groups. We can obtain similar results different values $p \in (0, \infty)$ in (4.14) but for the sake of brevity we concentrate on the criterion Φ_∞ which is probably also the easiest to interpret for practitioners.

We denote by $\hat{\theta}_n^*$ the linear unbiased estimator derived in Section 4. For each of the combinations of regression functions containing two different functions defined in (4.15), the optimal weights have been found by Theorem 4.2 and the optimal design points t_i^* are determined minimising the criterion Φ_∞ defined in (4.14). For the numerical optimisation the Particle Swarm Optimisation (PSO) algorithm is used (see, for example, Clerc, 2006) assuming a sample size of four observations in each group, that is, $n = 4$. Furthermore, the uniform design used in the following calculations is the design which has four equally spaced design points in the interval $[1, 10]$. The Φ_∞-optimal design points minimizing the criterion criterion in (4.14) are given in Table 1 for all combinations of models and correlations under consideration. Note that for each model the corresponding optimal design points change for different values of correlation ϱ.

In order to investigate the impact of the optimal design on the structure of the confidence bands we have performed a small simulation study simulating confidence bands for the difference of the regression functions. The vector of parameter values used for each model is $\theta = (\theta^{(1)\top}, \theta^{(2)\top})^\top = (1, 1, 1, 1, 1, 1)^\top$. In Figure 1 we display the averages of uniform confidence bands defined in (4.12) under the uniform and optimal design calculated by 100 simulation runs.

The left, middle and right columns show the results for the correlation $\varrho = 0.2$, $\varrho = 0.5$ and $\varrho = 0.7$, respectively, while the rows correposnd to different combinations for the functions f_1

Figure 1: *Confidence bands for the difference of the regression functions (solid grey line) on the basis of an optimal (solid lines) and uniform design (dashed lines). Left panel: $\varrho = 0.2$. Middle panel: $\varrho = 0.5$. Right panel: $\varrho = 0.7$. First row: model with $f_1 = f_A$ and $f_2 = f_B$. Second row: model with $f_1 = f_A$ and $f_2 = f_C$. Third row: model with $f_1 = f_B$ and $f_2 = f_C$.*
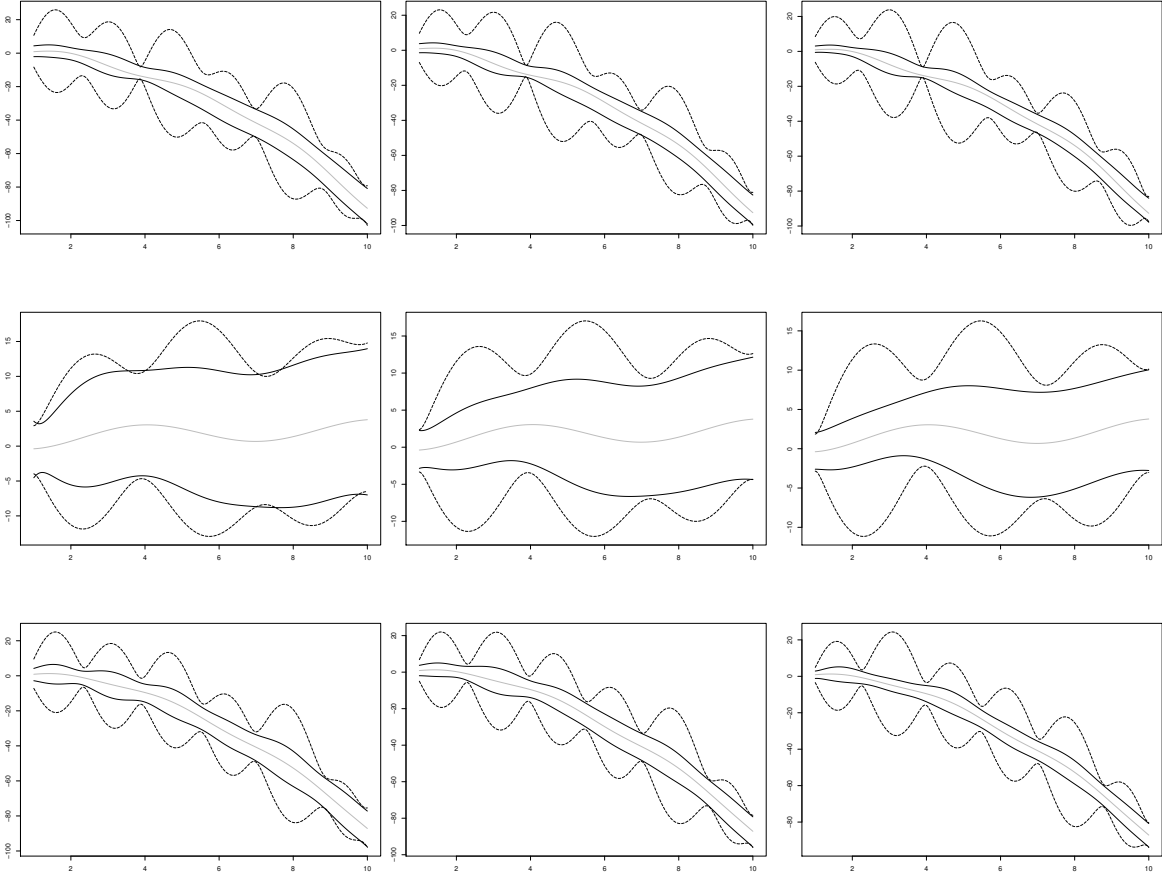
Table 1: *Optimal designs points on the interval* $[1, 10]$ *for the estimator* $\hat{\theta}_n^*$ *in* (4.1) *minimizing the criterion* Φ_∞ *in* (4.14). *Different correlations* $\varrho = 0.2, 0.5, 0.7$ *and different regression functions defined in* (4.15) *are considered.*

models	correlation		
	$\varrho = 0.20$	$\varrho = 0.50$	$\varrho = 0.70$
$f_1 = f_A$ & $f_2 = f_B$	[1, 1.59, 3.93, 10]	[1, 1.62, 3.91, 10]	[1, 1.74, 7.99, 10]
$f_1 = f_A$ & $f_2 = f_C$	[1, 3.46, 9.60, 10]	[1, 2.86, 8.83, 10]	[1, 2.61, 3.52, 10]
$f_1 = f_B$ & $f_2 = f_C$	[1, 2.20 , 6.25, 10]	[1, 1.62, 3.98 , 10]	[1, 2.85, 6.29 , 10]

and f_2 (first row: $f_1 = f_A$, $f_2 = f_B$, middle row: $f_1 = f_A$, $f_2 = f_C$ and last row $f_1 = f_B$, $f_2 = f_C$). In each graph, the confidence bands from the Φ_∞-optimal or the uniform design are plotted separately using the solid and dashed lines respectively, along with the plot for the true difference $f_1^\top(t)\theta^{(1)} - f_2^\top(t)\theta^{(2)}$ (solid grey lines).

We observe, that in all cases under considerations the use of Φ_∞-optimal designs yields a clearly visible improvement compared to the uniform design. The maximal width of the confidence band is reduced substantially. Moreover, the bands from the Φ_∞-optimal designs are nearly uniformly more narrow than the bands based on the uniform design. Even more importantly, the confidence bands based on the Φ_∞-optimal design show a similar structure as the true differences, while the confidence bands from the uniform design oscillate.

A comparison of the left, middle and right columns in Figure 1 shows that the maximum width for the confidence bands based on the optimal design decreases with increasing (absolute) correlation ϱ. This effect is not visible for the confidence bands based on the uniform design. For example, for the middle row of Figure 1, which corresponds to the case $f_1 = f_A$ and $f_2 = f_C$, the maximum width of the confidence bands based on the equally spaced design points even seem to increase.

Table 2 presents the values of the criterion Φ_∞ in (4.14) for the different scenarios and confirms the conclusions drawn from the visual inspection of the confidence bands plots. We observe that the use of the optimal design points reduces the maximum width of the confidence bands substantially. Moreover, for the optimal design the maximum width becomes smaller with increasing (absolute) correlation. On the other hand this monotonicity cannot be observed in all cases for the uniform designs.

Summarizing, the use of the proposed Φ_∞-optimal design improves statistically inference substantially reducing the maximum variance of the difference of the two estimated regression curves. Moreover, simultaneous estimation in combination with a Φ_∞-optimal design yields a further reduction of the maximum width of the confidence bands, thus providing a more precise inference for the difference of the curves describing the relation between t and the responses in the two groups.

Table 2: *Values of the criterion Φ_∞ for the optimal and uniform design with four observations in each group in the interval $[1, 10]$. The error process is given by a two independent Brownian motions with correlation $\varrho = 0.2, 0.5, 0.7$ between the groups, respectively.*

		Correlation		
Models	Design	$\varrho = 0.2$	$\varrho = 0.5$	$\varrho = 0.7$
$f_1 = f_A$ & $f_2 = f_B$	optimal	14.79	9.44	6.09
	uniform	141.87	142.59	148.74
$f_1 = f_A$ & $f_2 = f_C$	optimal	16.00	10.00	6.60
	uniform	33.32	29.10	25.66
$f_1 = f_B$ & $f_2 = f_C$	optimal	14.71	9.53	5.99
	uniform	147.27	127.19	115.07

References

Attia, A. and Constantinescu, E. (2020). Optimal experimental design for inverse problems in the presence of observation correlations. *arxiv:2007.14476*.

Bickel, P. J. and Herzberg, A. M. (1979). Robustness of design against autocorrelation in time I: Asymptotic theory, optimality for location and linear regression. *Annals of Statistics*, 7(1):77–95.

Bretz, F., Möllenhoff, K., Dette, H., Liu, W., and Trampisch, M. (2018). Assessing the similarity of dose response and target doses in two non-overlapping subgroups. *Statistics in Medicine*, 37:722–738.

Campos-Barreiro, S. and López-Fidalgo, J. (2015). *D*-optimal experimental designs for a growth model applied to a Holstein-Friesian dairy farm. *Statistical Methods & Applications*, 24:491–505.

Clerc, M. (2006). *Particle Swarm Optimization*. Iste Publishing Company, London.

Dette, H., Konstantinou, M., and Zhigljavsky, A. (2017a). A new approach to optimal designs for correlated observations. *Ann. Statist.*, 45(4):1579–1608.

Dette, H., Kunert, J., and Pepelyshev, A. (2008). Exact optimal designs for weighted least squares analysis with correlated errors. *Statistica Sinica*, 18(1):135–154.

Dette, H., Möllenhoff, K., Volgushev, S., and Bretz, F. (2018). Equivalence of regression curves. *Journal of the American Statistical Association*, 113:711–729.

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

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

Dette, H., Pepelyshev, A., and Zhigljavsky, A. (2019). The blue in continuous-time regression models with correlated errors. *Ann. Statist.*, 47(4):1928–1959.

Dette, H. and Schorning, K. (2016). Optimal designs for comparing curves. *Ann. Statist.*, 44(3):1103–1130.

Dette, H., Schorning, K., and Konstantinou, M. (2017b). Optimal designs for comparing regression models with correlated observations. *Computational Statistics & Data Analysis*, 113:273 – 286.

Doob, J. L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. *The Annals of Mathematical Statistics*, 20(3):393–403.

Fedorov, V. V. and Leonov, S. L. (2013). *Optimal Design for Nonlinear Response Models*. CRC Press, Boca Raton, FL, USA.

Gsteiger, S., Bretz, F., and Liu, W. (2011). Simultaneous confidence bands for non- linear regression models with application to population pharmacokinetic analyses. *Journal of Biopharmaceutical Statistics*, 21(4):708–725.

Harman, R. and Štulajter, F. (2010). Optimal prediction designs in finite discrete spectrum linear regression models. *Metrika*, 72(2):281–294.

Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation*. Springer-Verlag, New York-Berlin.

Kiselak, J. and Stehlík, M. (2008). Equidistant D-optimal designs for parameters of Ornstein-Uhlenbeck process. *Statistics and Probability Letters*, 78:1388–1396.

Liu, W., Bretz, F., Hayter, A. J., and Wynn, H. P. (2009). Assessing non-superiority, non-inferiority or equivalence when comparing two regression models over a restricted covariate region. *Biometrics*, 65:1279–1287.

Liu, W., Jamshidian, M., and Zhang, Y. (2011). Multiple comparison of several linear regression models. *Journal of the American Statistical Association*, 99(466):395–403.

Mehr, C. B. and McFadden, J. (1965). Certain properties of Gaussian processes and their first-passage times. *Journal of the Royal Statistical Society, Ser. B.*, 27(3):505–522.

Möllenhoff, K., Bretz, F., and Dette, H. (2020). Equivalence of regression curves sharing common parameters. *Biometrics*, 76:518–529.

Möllenhoff, K., Dette, H., Kotzagiorgis, E., Volgushev, S., and Collignon, O. (2018). Regulatory assessment of drug dissolution profiles comparability via maximum deviation. *Statistics in Medicine*, 37:2968–2981.

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

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

Pázman, A. and Müller, W. G. (2001). Optimal design of experiments subject to correlated errors. *Statist. Probab. Lett.*, 52(1):29–34.

Pukelsheim, F. (2006). *Optimal Design of Experiments*. SIAM, Philadelphia.

Rodriguez-Diaz, J. M. (2017). Computation of c-optimal designs for models with correlated observations. *Computational Statistics & Data Analysis*, 113:287 – 296.

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.