Ann. Inst. Statist. Math.
Vol. 52, No. 4, 658–679 (2000)

ASYMPTOTICALLY OPTIMAL TESTS AND OPTIMAL DESIGNS FOR TESTING THE MEAN IN REGRESSION MODELS WITH APPLICATIONS TO CHANGE-POINT PROBLEMS

WOLFGANG BISCHOFF* AND FRANK MILLER

Institute of Mathematical Stochastics, Department of Mathematics, University of Karlsruhe, D-76128 Karlsruhe, Germany

(Received July 27, 1998; revised July 2, 1999)

Abstract. Let a linear regression model be given with an experimental region $[a, b] \subseteq \mathbb{R}$ and regression functions $f_1, \dots, f_{d+1} : [a, b] \to \mathbb{R}$. In practice it is an important question whether a certain regression function f_{d+1}, say, does or does not belong to the model. Therefore, we investigate the test problem H_0 : "f_{d+1} does not belong to the model" against K : "f_{d+1} belongs to the model" based on the least-squares residuals of the observations made at design points of the experimental region $[a, b]$. By a new functional central limit theorem given in Bischoff (1998, *Ann. Statist.*, **26**, 1398–1410), we are able to determine optimal tests in an asymptotic way. Moreover, we introduce the problem of experimental design for the optimal test statistics. Further, we compare the asymptotically optimal test with the likelihood ratio test (F-test) under the assumption that the error is normally distributed. Finally, we consider real change-point problems as examples and investigate by simulations the behavior of the asymptotic test for finite sample sizes. We determine optimal designs for these examples.

Key words and phrases: Asymptotically optimal tests, linear regression, F-test, likelihood ratio test, Gaussian processes, optimal designs, change-point problem, quality control.

1. Introduction

In practice, it is an important question whether the linear regression model

$$H_0: \ Y(t) = f(t)^\top \boldsymbol{\beta} + \epsilon, \qquad t \in [a, b]$$

holds true, where $f(t)^\top = (f_1(t), \dots, f_d(t))$ is the vector of known regression functions, $d \geq 1$, or whether we additionally need a known regression function f_{d+1} for describing $Y(t)$. That means the regression model

$$K : Y(t) = f(t)^T \boldsymbol{\beta} + f_{d+1}(t)\beta_{d+1} + \epsilon, \qquad t \in [a, b]$$

holds true. For the models, we assume that $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)^\top \in \mathbb{R}^d$ is an unknown parameter vector, $\beta_{d+1} \in \mathbb{R}$ is an unknown parameter, and ϵ is a real random variable with

$$E(\epsilon) = 0, \qquad \text{Var}(\epsilon) = \sigma^2 \in (0, \infty) \text{ known or unknown.}$$

* The work of the first author was partly and the work of the second author fully supported by the Deutsche Forschungsgemeinschaft Grant Bi655/2-1.

The set $\mathcal{E} := [a, b] \subseteq \mathbb{R}$ is called the experimental region.

We investigate the problem of testing

$$H_0 : f_{d+1} \text{ does not belong to the model, that is } \beta_{d+1} = 0,$$

against

$$K : f_{d+1} \text{ belongs to the model, that is } \beta_{d+1} \neq 0,$$

given n independent observations $Y(t_1), \ldots, Y(t_n)$, where $t_1, \ldots, t_n \in \mathcal{E}$. If $Y(t_i)$, $i = 1, \ldots, n$, is normally distributed, then it is well-known that the uniformly most powerful size α test is a suitable likelihood ratio test (F-test). But under the weaker assumption that the error variables $\epsilon_1, \ldots, \epsilon_n$ corresponding to the observations $Y(t_1), \ldots, Y(t_n)$ are independent and identically distributed, nothing is known about this optimality to our knowledge.

In the present paper, the test problem described above is treated asymptotically. For this we consider the so-called residual partial sums limit processes; these are the limit processes of sequences of stochastic processes defined by partial sums of regression residuals (see MacNeill (1978*a*,*b*), and Bischoff (1998)). Under certain conditions, MacNeill (1978*b*) derived an explicit form of the residual partial sums limit processes for general linear regression residuals. This result has been generalized under weaker assumptions by Bischoff (1998).

We show that the uniformly most powerful test based on the residual partial sums limit process for the null hypothesis H_0 against the alternative K is a linear integral statistic. This result is derived in Section 4. Beforehand, in Section 3 we investigate the model under the null hypothesis and give the mathematical background necessary for our results.

It is worth mentioning that given a linear regression model, we have a bijection between the vector of least-squares residuals and the corresponding residual partial sums process; see formula (3.2). Hence, given a sequence of suitable regression models, the corresponding residual partial sums limit process contains "limit information" of the sequence of vectors of least-squares residuals. Thus, it can be said that we determine the uniformly most powerful test based on the "limit information" of the sequence of vectors of least-squares residuals; see Section 3.

Given a design $(t_1, \ldots, t_n) \in [a, b]^n$, $n \in \mathbb{N}$, for the test problem given above, we consider in Section 5 the well-known likelihood ratio test under the assumption that the error vector is normally distributed, i.e. we consider the F-test. If the sequence of designs suitably converges for $n \to \infty$, we can prove that the corresponding sequence of likelihood ratio tests "converges" to the asymptotically uniformly most powerful test based on the residual partial sums limit process. Hence, our result can also be stated as the asymptotic optimality of the likelihood ratio test (F-test) for the above test problem under a nonnormal error structure.

Moreover, we are interested in optimal experimental designs for the test problem considered above. Note that we use the functional central limit theorem of Bischoff (1998) for investigating the problem of experimental design. In Section 7, we determine optimal designs of the uniformly most powerful test mentioned above for some examples.

The results of the whole paper are illustrated by real examples of quality control. These can be described as change-point regression models. We introduce these examples and a general change-point regression model of some interest in Section 2.

For the situation of the change-point examples, in Section 6, we investigate by simulations the behavior of the asymptotic test for finite sample sizes.

Some technical proofs are given in the Appendix.

2. Change-point model

In practice, it is an important question whether the linear regression model holds true for the whole experimental region $[a, b]$ or whether a change-point $t^* \in (a, b)$ occurs. If a change-point t^* exists, then it is supposed that the regression model is given by

$$Y(t) = \begin{cases} f(t)^\top \boldsymbol{\beta} + \epsilon, & \text{if} \quad t \in [a, t^*), \\ f(t)^\top \boldsymbol{\beta} + f_{d+1,t^*}(t)\beta_{d+1} + \epsilon, & \text{if} \quad t \in [t^*, b], \end{cases}$$

where $(\boldsymbol{\beta}^\top, \beta_{d+1})^\top \in \mathbb{R}^{d+1}$ is the unknown parameter vector of the regression functions and $f_{d+1,t^*} : [a, b] \to \mathbb{R}$ with $f_{d+1,t^*}(t) = 0$ for $t \in [a, t^*)$. In the present paper, we assume that t^* is known. In the sequel, we write f_{d+1} for f_{d+1,t^*}. Note that, since t^* is known, b can be chosen in practice so that the period $[t^*, b]$ is suitable for the practician. Our verification problem can be formulated as a test:

$$H_0 : \text{There is no change at } t^*$$

against

$$K : \text{There is a change at } t^*.$$

It is worth mentioning, however, that some of our results do not depend on t^*. Then these statements hold true for unknown change-point t^* as well. In forthcoming papers, we will be concerned with the same problem as in the present paper but with an unknown change-point. The basis of these forthcoming papers will be the results of this paper.

In the literature on "detecting change-points" in linear regression models, it is familiar to consider the residual partial sums processes or variants of it; see, for instance, Gardner (1969), Brown *et al.* (1975), Sen and Srivastava (1975), MacNeill (1978*a*), Sen (1982), Jandhyala and MacNeill (1991), Tang and MacNeill (1993), Watson (1995) and the references cited therein. Change-point problems are mostly challenging if the change is small. Small changes, however, can only be discovered in an asymptotic way.

It is worth mentioning that Bischoff (1996) is concerned with properties of certain change-point test statistics and the sample path behavior of residual partial sums processes.

In Examples 2.1 and 2.2, we introduce two real problems of quality control.

Example 2.1. Let $d = 1$ and $f(t) = f_1(t) = 1$, $f_2(t) = \mathbf{1}_{[t^*,b]}(t)$ for $t \in [a, b]$, where

$$\mathbf{1}_{[t^*,b]}(t) := \begin{cases} 0, & \text{if} \quad t \in [a, t^*), \\ 1, & \text{if} \quad t \in [t^*, b]. \end{cases}$$

Thus the change of the parameter from β_1 to $\beta_1 + \beta_2$ can be described by the model

$$Y(t) = \beta_1 + \beta_2 \mathbf{1}_{[t^*,b]}(t) + \epsilon, \qquad t \in [a, b].$$

Such a model has been used in a quality control department of the German industrial company GKN Sinter Metals for controlling the density of toothed wheels produced during a certain period of time $[a, b]$. The engineers have taken an equidistant sample in time and assumed that the random variables of the sample are independent. Usually, the measurements can be assumed to be identically distributed, but if the machine is

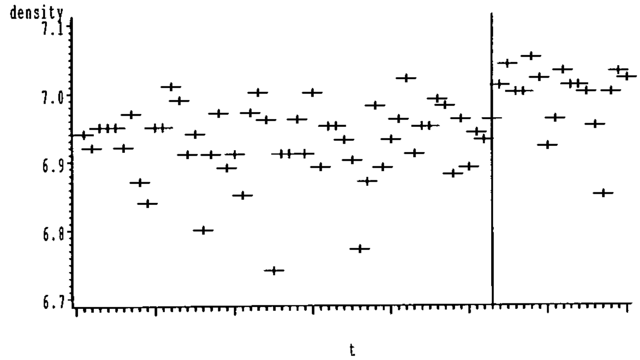

Fig. 1. Density of toothed wheels.

adjusted at time t^*, it is known that a change in the mean of the measurements can occur in t^*. It can be supposed that the measurements are identically distributed after the adjustment at the point of time t^*. The engineers are interested in knowing after the period of time $[a, b]$, on account of the whole sample, whether a change does or does not occur in $t^* \in (a, b)$. By measuring the density, the controlled toothed wheels are destroyed. So the engineers are interested in an optimal test. The data are shown in Fig. 1, where the point of time t^* is at the vertical reference line.

Example 2.2. Let $d = 1$ and $f(t) = f_1(t) = 1$, $f_2(t) = (t - t^*)^+$, $t \in [a, b]$, which means we have the following model, if a change occurs in t^*:

$$Y(t) = \beta_1 + \beta_2(t - t^*)^+ + \epsilon, \qquad t \in [a, b],$$

where

$$(t - t^*)^+ = \begin{cases} 0, & \text{if} \quad t \leq t^*, \\ t - t^*, & \text{if} \quad t > t^*. \end{cases}$$

Such a model has been used by the quality control department after another production step in which the stability of the toothed wheels is controlled. During the production, the toothed wheels were treated together with other oily work-pieces. After a while, it arose that the oil of the other work-pieces had a negative influence on the stability of the toothed wheels. From the point of time $t^* \in (a, b)$, say, the machine used in that production step has been reserved for the toothed wheels alone. The remaining oil in the machine decreases from the point of time t^*. So the above change-point model seems to be a possible choice for the data. Again, the controlled units are destroyed, the sample is taken equidistantly, and it is assumed that the measurements are stochastically independent. The engineers are interested in an optimal test, given the measurements in the period of time $[a, b]$. The data are shown in Fig. 2, where t^* is at the reference line.

Example 2.3. Let $d = 2$ and $f_1(t) = 1$, $f_2(t) = t$, $f_3(t) = (t - t^*)^+$, $t \in [a, b]$, which means we have the change-point model

$$Y(t) = \beta_1 + \beta_2 t + \beta_3(t - t^*)^+ + \epsilon, \qquad t \in [a, b].$$

Such a model and similar, more complicated models are investigated in detail as examples.

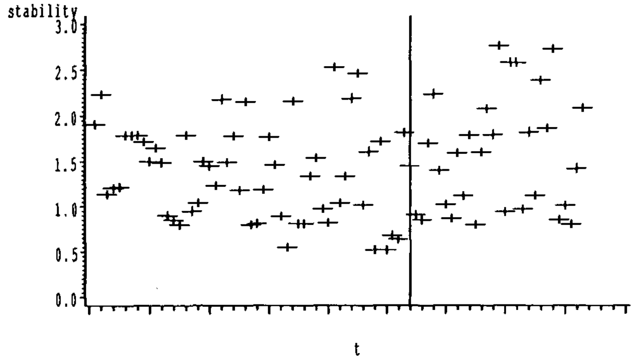

Fig. 2. Stability of toothed wheels.

3. Notation and preliminary results

Let $f_1, \ldots, f_d : \mathcal{E} \to \mathbb{R}$ be known measurable regression functions, where $\mathcal{E} = [a, b] \subseteq \mathbb{R}$ is the experimental region. As usual, we write $f(t)$ for $(f_1(t), \ldots, f_d(t))^\top$, $t \in \mathcal{E}$. Let us consider a triangular array t_{nj}, $1 \le j \le n$, $n \in \mathbb{N}$, of arbitrary experimental conditions, i.e. $t_{nj} \in \mathcal{E}$. For each $n \in \mathbb{N}$, we assume that $t_{n1} \le t_{n2} \le \cdots \le t_{nn}$. $(t_{n1}, \ldots, t_{nn})$ is called an (exact) design (for n observations). Note that we do not assume $t_{ni} \neq t_{ni+1}$. Corresponding to this array of experimental conditions, we have a triangular array of random variables Y_{nj}, $1 \le j \le n$, $n \in \mathbb{N}$, defined by

$$Y_{nj} = \sum_{i=1}^{d} \beta_i f_i(t_{nj}) + \epsilon_{nj},$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)^\top \in \mathbb{R}^d$ is the unknown parameter vector of interest and $\boldsymbol{\epsilon}_n = (\epsilon_{n1}, \ldots, \epsilon_{nn})^\top$ is a vector of stochastically independent and identically distributed real-valued random variables with $E(\epsilon_{nj}) = 0$ and $\mathrm{Var}(\epsilon_{nj}) = \sigma^2 \in (0, \infty)$.

Let $n \in \mathbb{N}$ be fixed. In the usual matrix formulation, we have

$$\boldsymbol{Y}_n = X_n \boldsymbol{\beta} + \boldsymbol{\epsilon}_n, \tag{3.1}$$

where X_n is the model matrix corresponding to the design $(t_{n1}, \ldots, t_{nn})$, i.e. the (s, r)-th component of X_n is $f_r(t_{ns})$. Then, for $\boldsymbol{\beta}$ to become estimable, $\mathrm{rank}(X_n)$ must be equal to d. We assume that $\mathrm{rank}(X_n) = d$ for all $n \ge n_0$ and, in the sequel, we consider $n \ge n_0$ only. Given model (3.1), the best linear unbiased estimation for $X_n\boldsymbol{\beta}$ is given by the least-squares estimation $\mathrm{pr}_{X_n} \boldsymbol{Y}_n = X_n(X_n^\top X_n)^{-1} X_n^\top \boldsymbol{Y}_n$ and the corresponding least-squares residuals vector is given by

$$\boldsymbol{r}_n = (r_{n1}, \ldots, r_{nn})^\top = \mathrm{pr}_{X_n^\perp} \boldsymbol{Y}_n = \mathrm{pr}_{X_n^\perp} \boldsymbol{\epsilon}_n,$$

where $\mathrm{pr}_{X_n} = X_n(X_n^\top X_n)^{-1} X_n^\top$ and $\mathrm{pr}_{X_n^\perp} = I_n - X_n(X_n^\top X_n)^{-1} X_n^\top$ are the orthogonal projectors onto $\mathrm{range}(X_n)$ and onto the orthogonal complement of $\mathrm{range}(X_n)$, respectively. The aim is to give a limit theorem for the sequence

$$\left(r_{n1}, r_{n1} + r_{n2}, \ldots, \sum_{j=1}^{n} r_{nj} \right)_{n \in \mathbb{N}}$$

of the partial sums of the least-squares residuals. For this, we need the functional $T_n : \mathbb{R}^n \to C[0,1]$ with

$$T_n(\boldsymbol{a})(z) = \sum_{i=1}^{[nz]} a_i + (nz - [nz])a_{[nz]+1}, \qquad z \in [0,1],$$

where $\boldsymbol{a} = (a_1, \ldots, a_n)^\top \in \mathbb{R}^n$, $[s] = \max\{n \in \mathbb{N}_0 \mid n \leq s\}$ and $\sum_{i=1}^{0} a_i = 0$. Thus we obtain

$$\left(T_n(\boldsymbol{r}_n)\left(\frac{1}{n}\right), T_n(\boldsymbol{r}_n)\left(\frac{2}{n}\right), \ldots, T_n(\boldsymbol{r}_n)(1)\right) = \left(r_{n1}, r_{n1} + r_{n2}, \ldots, \sum_{j=1}^{n} r_{nj}\right)_{n\in\mathbb{N}}. \tag{3.2}$$

Note that $(\sigma\sqrt{n})^{-1}T_n(\boldsymbol{r}_n)(z)$, $z \in [0,1]$, is a stochastic process in $C[0,1]$.

Next, each (exact) design $(t_{n1}, \ldots, t_{nn}) \in \mathcal{E}^n$ uniquely corresponds to a discrete probability measure P_n on $\mathcal{E}$ by

$$P_n = \frac{1}{n}\sum_{i=1}^{n} P_{\{t_{ni}\}}, \tag{3.3}$$

where $P_{\{t\}}$ denotes the one-point measure in t. In the sequel, we identify an exact design with its representation as a discrete probability measure and we call each probability measure on $\mathcal{E}$ a continuous design. Further, we do not distinguish between a design and its representation as a distribution function F_n, say, corresponding to P_n. For our results, we need that the sequence of designs F_n converges uniformly to a continuous design F_0, say

$$\sup_{t\in[a,b]} |F_n(t) - F_0(t)| \to 0 \quad \text{for} \quad n \to \infty. \tag{3.4}$$

To avoid misunderstand, we repeat that each probability measure P_0 on the experimental region as well as the corresponding distribution function F_0 are called continuous designs; but this does not imply that F_0 is continuous. In this paper, distribution functions are defined as right-continuous functions; hence, within our context, the class of all distribution functions is given by

$$\mathcal{F} := \{F : [a,b] \to [0,1] \mid F \text{ is increasing and right-continuous}, F(b) = 1\}. \tag{3.5}$$

We put

$$\left(\int_{\mathcal{E}} f_i(t)f_j(t)F_0(dt)\right)_{i,j=1}^{d} =: \int_{\mathcal{E}} f(t)f(t)^\top F_0(dt) =: J,$$

and assume that

$$\operatorname{rank}(J) = d. \tag{3.6}$$

Note that the above integral is defined pointwise. In the sequel, similar integrals are defined accordingly. It is obvious that (3.6) is fulfilled if and only if the regression functions $f_1, \ldots, f_d$ are linearly independent in $L_2(F_0)$, where $L_2(F_0)$ is the Hilbert space of square integrable functions with respect to F_0.

Remark 3.1. Given a continuous design F_0, we can construct, in a natural way, a sequence of designs (F_n) by $F_n = \frac{1}{n}\sum_{i=1}^n P_{\{t_{ni}\}}$ converging uniformly (according to (3.4)) to F_0: For example, we can choose $t_{ni} := Q_0(\frac{i}{n})$, $1 \le i \le n$, where Q_0 is the right-continuous quantile function of F_0 (this means $Q_0(z) := \sup\{t \in [a,b] : F_0(t) \le z\}$, where $\sup(\emptyset) := a$). In a similar way, Sacks and Ylvisacker (1966) introduced designs for regression models with correlated errors.

Now we can state the functional central limit theorem of Bischoff (1998). We use integration by parts for obtaining our formula instead of that given in Bischoff (1998).

THEOREM 3.1. *Let the regression functions $f_1, \ldots, f_d$ be continuous and of bounded variation. Let the conditions* (3.4) *and* (3.6) *be fulfilled for $f_1, \ldots, f_d$, the sequence of designs (F_n) and the continuous design F_0. Then $(\sigma\sqrt{n})^{-1}T_n(\boldsymbol{r}_n)$ converges weakly in $C[0,1]$ to the Gaussian process B_{f,F_0}, defined by*

$$B_{f,F_0}(z) = B(z) - \left(\int_{\mathcal{E}} f(t)F_{0;z}(dt)\right)^{\top} J^{-1}\left(\int_{\mathcal{E}} f(t)B(F_0(dt))\right),$$

where B is the Brownian motion on $[0,1]$ and $F_{0;z}(t) = \min\{F_0(t), z\}$.

It seems to be worth mentioning that all integrals are to be understood as μ-integrals, where μ is a signed measure. For example, let us consider for $i \in \{1, \ldots, d\}$ fixed

$$\int_{\mathcal{E}} f_i(t)B(F_0(dt)) = f_i(b)B(F_0(b)) - \int_{\mathcal{E}} B(F_0(t))f_i(dt),$$

where integration by parts is used. The left-hand integral can only be interpreted as a Riemann-Stieltjes integral, but the right-hand integral can be understood as a μ_i-integral, where μ_i is the signed measure corresponding to the function f_i of bounded variation. In this sense, we can consider all integrals as μ-integrals.

Example 3.1. (Examples 2.1 and 2.2 continued) In the case where $d = 1$, $f(t) = f_1(t) = 1$, $t \in \mathcal{E}$, it is easy to see by Theorem 3.1 that, independent of the design F_0, the Brownian bridge is the limit process

$$B_{f,F_0}(z) = B(z) - z \cdot B(1), \qquad z \in [0,1].$$

In the sequel, we need the covariance function of B_{f,F_0} given in the following lemma. The proof of this result can be found in the Appendix.

LEMMA 3.1. *The mean and the covariance function of the Gaussian process B_{f,F_0} stated in Theorem* 3.1 *are given by*

$$m(s) = 0, \qquad s \in [0,1],$$

and

$$K(s,z) = \min\{s,z\} - \left(\int_{\mathcal{E}} f(t)F_{0;s}(dt)\right)^{\top} J^{-1}\left(\int_{\mathcal{E}} f(t)F_{0;z}(dt)\right), \qquad s,z \in [0,1],$$

respectively.

Using the notation of Theorem 3.1, we call $(\sigma\sqrt{n})^{-1}T_n(\boldsymbol{r}_n)$ the residual partial sums process of f and the exact design F_n. B_{f,F_0} is called the residual partial sums limit process of f and the continuous design F_0.

4. Asymptotic considerations for the regression model

We want to test whether a function f_{d+1} does or does not belong to the linear regression model by observing $n \in \mathbb{N}$ independent responses at the design points $t_{ni} \in \mathcal{E} = [a, b]$, $i = 1, \dots, n$. Under the null hypothesis H_0, we assume that the model

$$Y_{ni} = f(t_{ni})^\top \boldsymbol{\beta} + \epsilon_{ni}, \qquad i = 1, \dots, n$$

is given. Under the alternative K, we assume that the model

$$Y_{ni} = f(t_{ni})^\top \boldsymbol{\beta} + f_{d+1}(t_{ni})\beta_{d+1} + \epsilon_{ni}, \qquad i = 1, \dots, n$$

occurs. So, for the exact design F_n corresponding to $(t_{n1}, \dots, t_{nn}) \in [a, b]^n$ we get the following linear models

$$\boldsymbol{Y}_n = X_n\boldsymbol{\beta} + \boldsymbol{\epsilon}_n \quad \text{under} \quad H_0 \tag{4.1}$$

and

$$\boldsymbol{Y}_n = \tilde{X}_n\tilde{\boldsymbol{\beta}} + \boldsymbol{\epsilon}_n \quad \text{under} \quad K, \tag{4.2}$$

where

$$\begin{aligned}
&\boldsymbol{\epsilon}_n = (\epsilon_{n1}, \dots, \epsilon_{nn})^\top \text{ with } E(\boldsymbol{\epsilon}_n) = 0 = (0, \dots, 0)^\top \in \mathbb{R}^n, \\
&\operatorname{Cov}(\boldsymbol{\epsilon}_n) = \sigma^2 I_n (I_n \text{ the } n \times n - \text{identity matrix}) \text{ and} \\
&\quad \epsilon_{n1}, \dots, \epsilon_{nn} \text{ independent and identically distributed,} \\
&X_n = (f_k(t_{ni}))_{i=1,k=1}^{n,d} \text{ is the model (design) matrix under } H_0, \\
&\tilde{X}_n = (f_k(t_{ni}))_{i=1,k=1}^{n,d+1} \text{ is the model (design) matrix under } K, \\
&\tilde{\boldsymbol{\beta}} = (\beta_1, \dots, \beta_d, \beta_{d+1})^\top \in \mathbb{R}^{d+1}.
\end{aligned}$$

We assume that the assumptions of Theorem 3.1 for $f = (f_1, \dots, f_d)^\top$, $(F_n)_{n\in\mathbb{N}}$ given above and for F_0, say, hold true. Let us consider the residual partial sums limit processes for both models H_0 and K under the null hypothesis H_0. If (4.1) is true, then we obtain

$$B_{f,F_0}(z) = \lim_{n\to\infty} (\sigma\sqrt{n})^{-1} T_n(\operatorname{pr}_{X_n^\perp}(X_n\boldsymbol{\beta} + \boldsymbol{\epsilon}_n)).$$

Next, we assume that (4.2) is true. Noting that T_n is a linear operator, we get

$$T_n(\operatorname{pr}_{X_n^\perp}(\tilde{X}_n\tilde{\boldsymbol{\beta}} + \boldsymbol{\epsilon}_n)) = T_n(\operatorname{pr}_{X_n^\perp}(X_n\boldsymbol{\beta} + \boldsymbol{\epsilon}_n)) + \beta_{d+1}T_n(\operatorname{pr}_{X_n^\perp}\boldsymbol{\xi}_n),$$

where

$$\boldsymbol{\xi}_n := (f_{d+1}(t_{n1}), \dots, f_{d+1}(t_{nn}))^\top.$$

The proof of the following result is contained in the Appendix.

LEMMA 4.1. *Let us assume that conditions (3.4) and (3.6) are satisfied. Further, let $f_1, \ldots, f_{d+1} \in L_2(F_0)$ be of bounded variation on $\mathcal{E}$, let $f_1, \ldots, f_d$ be continuous and let f_{d+1} be right-continuous on $\mathcal{E}$. Then it holds true that*

$$T_n\left(\frac{1}{n}\mathrm{pr}_{X_n^\perp}\boldsymbol{\xi}_n\right) \overset{n\to\infty}{\longrightarrow} h$$

in $(C[0,1], \|\cdot\|_\infty)$, where

$$\begin{aligned} h(z) &:= h_{F_0,f,f_{d+1}}(z) \\ &:= \int_{\mathcal{E}} f_{d+1}(t)F_{0;z}(dt) - \left(\int_{\mathcal{E}} f(t)F_{0;z}(dt)\right)^\top J^{-1}\left(\int_{\mathcal{E}} f_{d+1}(t)f(t)\,F_0(dt)\right). \end{aligned}$$

Thus, under the alternative K, we recognize that for a fixed and sufficiently large sample size n, $(\sigma\sqrt{n})^{-1}T_n(\mathrm{pr}_{X_n^\perp}\tilde{X}_n\boldsymbol{\beta}+\boldsymbol{\epsilon}_n)$ is approximately distributed as $\beta_{d+1}\sqrt{n}\sigma^{-1}h+B_{f,F_0}$. Hence, for obtaining an asymptotic test, we assume that $(\sigma\sqrt{n})^{-1}T_n(\mathrm{pr}_{X_n^\perp}\boldsymbol{Y}_n)$ is distributed as $\beta_{d+1}\sqrt{n}\sigma^{-1}h + B_{f,F_0}$ in $C[0,1]$, where n is fixed. Then the hypotheses corresponding to (4.1) and (4.2) can be stated as

$$H_0 : \beta_{d+1} = 0 \quad \text{against} \quad K : \beta_{d+1} \neq 0.$$

Remark 4.1. The above discussion shows that, for each $n \in \mathbb{N}$, we need a reparametrization of the parameter space for getting a limit distribution under the alternative. To this end we consider the random variable

$$\boldsymbol{Y}_n^* := \boldsymbol{\xi}_n\beta_{d+1}/\sqrt{n} + X_n\beta + \boldsymbol{\epsilon}_n$$

for testing

$$H_0 : \beta_{d+1}/\sqrt{n} = 0 \quad \text{against} \quad K : \beta_{d+1}/\sqrt{n} \neq 0.$$

Note that the above test problem is the same as the original one for each $n \in \mathbb{N}$. By the above considerations, we obtain

$$(\sigma\sqrt{n})^{-1}T_n(\mathrm{pr}_{X_n^\perp}\boldsymbol{Y}_n^*) \overset{\mathcal{D}}{\to} \beta_{d+1}\sigma^{-1}h + B_{f,F_0} \qquad (\text{for } n \to \infty),$$

where $\overset{\mathcal{D}}{\to}$ denotes weak convergence.

Remark 4.2. A uniformly most powerful test cannot exist for the test problem considered above. But for the one-sided test problem

$$H_0 : \beta_{d+1} = 0 \quad \text{against} \quad K : \beta_{d+1} > 0,$$

an asymptotically uniformly most powerful test exists as Theorem 4.1 shows.

For this test problem, we call a test statistic $\delta_n : C[0,1] \to \{0,1\}$, $n \in \mathbb{N}$, an asymptotically uniformly most powerful size α test (based on the residuals) if

$$\delta_n((\sigma\sqrt{n})^{-1}T_n(\mathrm{pr}_{X_n^\perp}\boldsymbol{Y}_n^*)) \overset{\mathcal{D}}{\to} \delta(\beta_{d+1}\sigma^{-1}h + B_{f,F_0}) \qquad (\text{for } n \to \infty), \tag{4.3}$$

where $\delta : C[0,1] \to \{0,1\}$ is a uniformly most powerful size α test for $H_0 : \beta_{d+1} = 0$ against $K : \beta_{d+1} > 0$ observing $\beta_{d+1}\sigma^{-1}h + B_{f,F_0}$. Note, that for the above definition, $T_n : \mathrm{pr}_{X_n^\perp}(\mathbb{R}^n) \to T_n(\mathrm{pr}_{X_n^\perp}(\mathbb{R}^n)) \subseteq C[0,1]$ is a bijection.

The above statement (4.3) implies that the power function of δ_n converges pointwise to the power function of δ for every continuity point of δ. If the power functions of δ_n and of δ are increasing with respect to β_{d+1} and the power function of δ is continuous and tends to 1 for $\beta_{d+1} \to \infty$, the convergence of the power functions is uniform. Note that the above conditions are fulfilled by the following test given in (4.4).

The proof of the following theorem is given in the Appendix.

THEOREM 4.1. *Let us consider the test problems* (4.1) *and* (4.2). *Let the conditions of Lemma* 4.1 *be fulfilled and let* $g(t) \equiv 1$, $t \in \mathcal{E}$, *be a linear combination of* $f_1, \ldots, f_d$. *Assume that* $f_1, \ldots, f_{d+1}$ *are linearly independent in* $L_2(F_0)$. *With the notation*

$$\Upsilon_{F_0} := \Upsilon_{F_0,f,f_{d+1}} := \left(-\int_{\mathcal{E}} h(F_0(t-))f_{d+1}(dt)\right)^{1/2},$$

the following is valid:

(a) *An asymptotically uniformly most powerful size* α *test for* $H_0 : \beta_{d+1} = 0$ *against* $K : \beta_{d+1} > 0$ *is given by*

$$\text{(4.4)} \qquad \text{“Reject } H_0 \Leftrightarrow S_{F_0,f,f_{d+1}}(y_n) := \Upsilon_{F_0}^{-1} \int_{\mathcal{E}} y_n(F_0(t-))f_{d+1}(dt) > \Phi^{-1}(1-\alpha)\text{”},$$

where $y_n(z) := (\sigma\sqrt{n})^{-1}T_n(\mathrm{pr}_{X_n^\perp}\boldsymbol{Y}_n)(z)$ *and* $\Phi^{-1}(1-\alpha)$ *is the* $(1-\alpha)$*-quantile of the* $N(0,1)$*-distribution.*

(b) *The test statistic is asymptotically normally distributed with mean* $\beta_{d+1}\sqrt{n}\sigma^{-1}\Upsilon_{F_0}$ *and variance* 1. *The test is consistent.*

Remark 4.3. In the notation of Remark 4.2, the test statistic in (4.4) is given by

$$\delta(u) = \delta_n(u) = \mathbf{1}\{S_{F_0,f,f_{d+1}}(u) > \Phi^{-1}(1-\alpha)\}, \qquad u \in C[0,1].$$

Remark 4.4. (a) The result of Theorem 4.1 holds also for the test problem

$$H_0 : \beta_{d+1} \le 0 \quad \text{against} \quad K : \beta_{d+1} > 0.$$

(b) For obtaining an asymptotically uniformly most powerful unbiased size α test for the test problem

$$H_0 : \beta_{d+1} = 0 \quad \text{against} \quad K : \beta_{d+1} \neq 0,$$

we have to change (4.4) into

$$\text{(4.5)} \qquad \text{“Reject } H_0 \Leftrightarrow \left|\Upsilon_{F_0}^{-1} \int_{\mathcal{E}} y_n(F_0(t-))f_{d+1}(dt)\right| > \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\text{”}.$$

(c) If σ^2 is unknown, then without altering the asymptotic distribution, σ^2 may be replaced by a consistent estimator.

Remark 4.5. In the case of a heteroscedastic error, similar results can be obtained. Such models will be investigated in forthcoming papers; see, for example, Bischoff *et al.* (1999).

For calculating $S_{F_0,f,f_{d+1}}$, the following formula for Υ_{F_0} can be useful.

LEMMA 4.2. *Under the assumptions of Lemma* 4.1 *and Theorem* 4.1, *we have*

$$\Upsilon_{F_0} = (J_0 - a_0^\top J^{-1} a_0)^{1/2},$$

where $J_0 := \int_{\mathcal{E}} f_{d+1}(t)^2 F_0(dt)$ *and* $a_0 := \int_{\mathcal{E}} f(t) f_{d+1}(t) F_0(dt)$.

The proof can be found in the Appendix.

Example 4.1. (Example 2.1 continued) Let us consider: $d = 1$, $f(t) = f_1(t) = 1$, $f_2(t) = \mathbf{1}_{[t^*,b]}(t)$, $t \in \mathcal{E}$. Let F_0 be an arbitrary continuous design with $s := F_0(t^*-)$, then

$$h_{F_0,f_1,f_2}(z) = \int_{[t^*,b]} 1 F_{0;z}(dt) - z(1-s) = \begin{cases} z(s-1), & \text{if} \quad z \in [0,s), \\ s(z-1), & \text{if} \quad z \in [s,1], \end{cases}$$
$$\Upsilon_{F_0} = (s(1-s))^{1/2}.$$

The optimal test statistic S_{F_0,f_1,f_2} stated in Theorem 4.1 is given by

$$S_{F_0,f_1,f_2}(u) = -\Upsilon_{F_0}^{-1} u(F_0(t^*-)) = -(s(1-s))^{-1/2} \cdot u(s), \qquad u \in C[0,1].$$

From the data of the quality control department, we have $S_{F_0,f_1,f_2}(y_n) = 3.93$, where σ in the computation of y_n is replaced with the square-root of the usual variance estimator under the H_0-model. Then the p-value of the two-sided test of Remark 4.4 (b) is 0.0001. Thus, the null hypothesis is rejected for all relevant α used in practice.

Example 4.2. (Example 2.2 continued) Let us consider: $d = 1$, $f(t) = f_1(t) = 1$, $f_2(t) = (t - t^*)^+$, $t \in \mathcal{E}$. Let F_0 be the uniform distribution on $\mathcal{E}$. With the definition $p := (t^* - a)/(b - a)$ we have

$$\begin{aligned} h_{F_0,f_1,f_2}(z) &= \int_{\mathcal{E}} (t-t^*)^+ F_{0;z}(dt) - \left(\int_{\mathcal{E}} (t-t^*)^+ F_0(dt) \right) z \\ &= \mathbf{1}_{[p,1]}(z) \cdot \left(\int_{[t^*,(b-a)z+a]} \frac{t-t^*}{b-a} dt \right) - \left(\int_{[t^*,b]} \frac{t-t^*}{b-a} dt \right) z \\ &= \frac{1}{2(b-a)} \left(\mathbf{1}_{[p,1]}(z)((b-a)z - (t^*-a))^2 - (b-t^*)^2 z \right) \\ &= (b-a)\frac{1}{2} \left(\mathbf{1}_{[p,1]}(z)(z-p)^2 - (1-p)^2 z \right). \end{aligned}$$

Further, analogously, we have

$$\begin{aligned} \Upsilon_{F_0} &= \left(\int_{\mathcal{E}} f_2(t)^2 F_0(dt) - \left(\int_{\mathcal{E}} f_2(t) F_0(dt) \right)^2 \right)^{1/2} \\ &= (b-a) \left(\frac{1}{12} (1-p)^3 (1+3p) \right)^{1/2}. \end{aligned}$$

Thus, the optimal statistic S_{F_0,f_1,f_2} stated in Theorem 4.1 is given by

$$S_{F_0,f_1,f_2}(u) = -\Upsilon_{F_0}^{-1} \int_{[t^*,b]} u\left(\frac{t-a}{b-a}\right) dt = -\Upsilon_{F_0}^{-1}(b-a) \int_p^1 u(z)dz, \qquad u \in C[0,1].$$

From the second data of the quality control department we have $S_{F_0,f_1,f_2}(y_n) = 1.85$. From the engineers' prior knowledge we can assume that the stability of the toothed wheels cannot deteriorate if the residuary oil decreases. So we take the one-sided test of Theorem 4.1 (a). The p-value is 0.032. Thus, the null hypothesis is rejected, for example, for $\alpha = 0.05$.

Example 4.3. We are concerned with the last model $d = 1$, $f(t) = f_1(t) = 1$, $f_2(t) = (t-t^*)^+$, $t \in \mathcal{E}$, once more. But now we consider the design F_0 that puts masses $\frac{1}{2}$ at the points a and b. Then, by simple calculations, we get

$$h_{F_0,f_1,f_2}(z) = \begin{cases} -\dfrac{b-a}{2}(1-p)z, & \text{if } z < \dfrac{1}{2}, \\ -\dfrac{b-a}{2}(1-p)(1-z), & \text{if } z \geq \dfrac{1}{2}, \end{cases}$$

$$\Upsilon_{F_0} = \frac{1}{2}(b-t^*) = \frac{b-a}{2}(1-p),$$

$$S_{F_0,f_1,f_2}(u) = -2u\left(\frac{1}{2}\right), \qquad u \in C[0,1].$$

Thus, the test of Theorem 4.1 rejects H_0 if and only if

$$S_{F_0,f_1,f_2}\left(y_n\left(\frac{1}{2}\right)\right) = -\frac{2}{\sigma\sqrt{n}}\left(\sum_{i=1}^{[n/2]} r_{ni} + \left(\frac{n}{2} - \left[\frac{n}{2}\right]\right) r_{n,[n/2]+1}\right) > \Phi^{-1}(1-\alpha)$$

holds true.

5. The likelihood ratio test (F-test)

In this section, we consider the original one-sided test problem

$$H_0 : \beta_{d+1} = 0 \quad \text{against} \quad K : \beta_{d+1} > 0$$

for a finite sample under the additional assumption that the error is normally distributed

$$\boldsymbol{\epsilon}_n \sim N_n(0, \sigma^2 I_n).$$

For simplifying the notation, we suppose that $\sigma^2 \in (0,\infty)$ is known. Further, we use the notation of Section 4 and assume the conditions of Theorem 4.1. Let $U := \mathrm{Im}X_n$, $\tilde{U} := \mathrm{Im}\tilde{X}_n$, $W := \tilde{U} \cap U^\perp$ and let $\boldsymbol{\xi}_n$ be the last column of $\tilde{X}_n$. Thus $\tilde{\boldsymbol{\xi}}_n := \boldsymbol{\xi}_n - \mathrm{pr}_U\boldsymbol{\xi}_n$ is a vector of W. Then the likelihood ratio test is given by

(5.1) $$\text{“Reject } H_0 \Leftrightarrow C_n(\boldsymbol{Y}_n) > \Phi(1-\alpha)\text{”},$$

where $C_n(\boldsymbol{Y}_n) := (\sigma\|\tilde{\boldsymbol{\xi}}_n\|)^{-1}\tilde{\boldsymbol{\xi}}_n^\top \boldsymbol{Y}_n$. The likelihood ratio test corresponds with the “F-test” for known σ^2. Thus, the test given above is uniformly most powerful in the class of all tests based on the residuals, see Arnold (1981), p. 108, in case $\epsilon_n \sim N_n(0, \sigma^2 I_n)$. Without supposing any distribution assumption, we call the test given in (5.1) the "likelihood ratio" test. Our aim is to compare the "likelihood ratio" test asymptotically with the test given in Theorem 4.1.

We can express the "likelihood ratio" statistic $C_n(\boldsymbol{Y}_n)$ in terms of the corresponding residual partial sums process. The following result is proved in the Appendix.

LEMMA 5.1. *Let us assume that condition* (3.4) *is satisfied and let* $f_1, \ldots, f_{d+1}$ *be of bounded variation on* $\mathcal{E}$ *and linearly independent in* $L_2(F_0)$. *Let* $f_1, \ldots, f_d$ *be continuous and let* f_{d+1} *be right-continuous.*

(a) *For the "likelihood ratio" statistics,* $C_n(\boldsymbol{Y}_n) = S_{F_n,f,f_{d+1}}(y_n)$, *hold true, where* $y_n(z) = (\sigma\sqrt{n})^{-1} T_n(\mathrm{pr}_{X_n^{\perp}} \boldsymbol{Y}_n)(z)$.

(b) *We have the pointwise convergence*

$$S_{F_n,f,f_{d+1}}(u) \to S_{F_0,f,f_{d+1}}(u), \qquad u \in C[0,1].$$

THEOREM 5.1. *Consider the test problems* (4.1) *and* (4.2). *Let the assumptions of Theorem* 4.1 *be fulfilled. Then the "likelihood ratio" test*

$$\text{"Reject } H_0 \Leftrightarrow C_n(\boldsymbol{Y}_n) = S_{F_n,f,f_{d+1}}(y_n) > \Phi^{-1}(1-\alpha)\text{"} \tag{5.2}$$

for $H_0 : \beta_{d+1} = 0$ *against* $K : \beta_{d+1} > 0$ *is an asymptotically uniformly most powerful size* α *test.*

PROOF. If $u_n \to u_0$, $u_0, u_n \in C[0,1]$, we have $S_{F_n,f,f_{d+1}}(u_n) \to S_{F_0,f,f_{d+1}}(u_0)$ because of Lemma 5.1 (b) and because $\|S_{F_n,f,f_{d+1}}\|$ is bounded. So Theorem 5.5 of Billingsley (1968) implies that

$$S_{F_n,f,f_{d+1}}((\sigma\sqrt{n})^{-1} T_n(\mathrm{pr}_{X_n^{\perp}} \boldsymbol{Y}_n^*)) \xrightarrow{\mathcal{D}} S_{F_0,f,f_{d+1}}(\beta_{d+1}\sigma^{-1}h + B_{f,F_0}).$$

Thus, the assertion of the theorem follows from Remark 4.2 and Theorem 4.1 □.

Remark 5.1. Theorem 5.1 holds also for the test problem

$$H_0 : \beta_{d+1} \le 0 \quad \text{against} \quad K : \beta_{d+1} > 0.$$

Remark 5.2. The difference between the two asymptotically uniformly most powerful tests (test (4.4) of Theorem 4.1 and the "likelihood ratio" test (5.2)) is only that the test of Theorem 4.1 uses $S_{F_0,f,f_{d+1}}$ and the "likelihood ratio" test uses $S_{F_n,f,f_{d+1}}$ as statistics for finite samples.

Example 5.1. (Examples 2.1 and 2.2 continued) From the data of the quality control department, the values of $S_{F_n,f_1,f_2}(y_n)$ are only slightly different from the values of $S_{F_0,f_1,f_2}(y_n)$ computed in Examples 4.1 and 4.2. The rounded values (3.93 and 1.85) are the same.

Table 1. Probabilities of rejection of the two-sided size α test (4.5) for the situation of Example 4.1: $\mathcal{E} = [0,1]$, $d = 1$, $f_1 = 1$, $f_2 = \mathbf{1}_{[t^*,1]}(t)$, $t^* = 53/70$, $n = 70$ under H_0 : "$\beta_2 = 0$" and certain alternatives $\beta_2 > 0$. 10^6 simulations.

	distribution	β_2				
α	of the error	0	.25	.5	.75	1
.01	normal	.009	.043	.203	.524	.829
.01	negative Gumbel	.009	.040	.214	.545	.829
.05	normal	.049	.145	.434	.769	.949
.05	negative Gumbel	.049	.149	.453	.776	.942

Table 2. Probabilities of rejection for the one-sided size α test (4.4) for the situation of Example 4.2: $\mathcal{E} = [0,1]$, $d = 1$, $f_1 = 1$, $f_2 = (t - t^*)^+$, $t^* = 54/83$, $n = 83$ under H_0 : "$\beta_2 = 0$" and certain alternatives $\beta_2 > 0$. 10^6 simulations.

	distribution	β_2				
α	of the error	0	1	2	3	4
.01	normal	.010	.080	.319	.679	.920
.01	Gumbel	.013	.086	.321	.677	.918
.05	normal	.050	.242	.598	.883	.984
.05	Gumbel	.055	.242	.592	.884	.985

Table 3. Probabilities of rejection of the size α test (5.2) for the situation $\mathcal{E} = [0,1]$, $d = 1$, $f_1 = 1$, $f_2 = (t - t^*)^+$, $t^* = 54/83$, $n = 83$ under H_0 : "$\beta_2 = 0$" for certain observation sizes n. 10^6 simulations.

distribution	n				
of the error	25	83	250	1000	∞
negative Gumbel	.039	.045	.047	.048	.05
Gumbel	.060	.055	.053	.052	.05

6. Simulations

In this section, we investigate by simulations the behavior of the asymptotic test used in Examples 4.1 and 4.2 for finite sample sizes. For these examples, the error distribution of the data does not seem to be normally distributed. An extreme value distribution seems to be more appropriate. Therefore, for Example 4.2 we consider the standardized Gumbel distribution with mean zero and variance one: $G(s) = \exp(-\exp(-s\pi/\sqrt{6}-\gamma))$, $s \in \mathbb{R}$, γ = Euler's constant. For Example 4.1 we consider the "negative Gumbel distribution" $1 - G(-s)$, $s \in \mathbb{R}$.

For the model of Example 4.1 ($f_1 = 1, f_2(t) = \mathbf{1}_{[t^*,b]}(t)$) and the situation of the quality-control data, i.e. $n = 70$, $t_{ni} = i/n, i = 1, \ldots, 70$, $t^* = 53/70$, $\mathcal{E} = [0,1]$ and F_0 the uniform distribution on the points t_{ni}, $i = 1, \ldots, 70$, we simulate the probabilities of rejection of the two-sided test given in Remark 4.4 (b) for the cases $\alpha \in \{.01, .05\}$, $\beta_2 \in \{0, .25, .5, .75, 1\}$. The simulations are done with the error $N(0,1)$ and with a negative Gumbel distribution error. Our results in Table 1 are based on 10^6 simulations. We do not assume that σ is known. Hence, for our simulations, we have to estimate σ for computing the test statistic.

For the model of Example 4.2 ($f_1 = 1, f_2(t) = (t - t^*)^+$) and the situation of the quality-control data, that is $n = 83$, $t_{ni} = i/n$, $i = 1, \ldots, 83$, $t^* = 54/83$, $\mathcal{E} = [0,1]$ and F_0 the uniform distribution on the points t_{ni}, $i = 1, \ldots, 83$, we simulate the probabilities of rejection of the test given in Theorem 4.1 for the cases $\alpha \in \{.01, .05\}$, $\beta_2 \in \{0, 1, 2, 3, 4\}$. The simulations are carried out as above. For this example, we have used the error as $N(0,1)$ and a Gumbel distribution, respectively. The results shown in Table 2 are based on 10^6 simulations.

We conclude this section with some simulations for increasing sample size. We investigate the model of Example 4.2 once more. For F_0, the uniform distribution on the interval $[0,1]$, we simulate the probability of rejection under the hypotheses ($\beta_2 = 0$) for $\alpha = 0.05$. We use the sample sizes $n \in \{25, 83, 250, 1000\}$ and the Gumbel and the negative Gumbel distribution. The rounded results based on 10^6 simulations for the statistic (4.4) and the statistic (5.2) are the same. They are shown in Table 3.

7. Design of experiments for change-point problems

In the previous sections, we have considered the test problem for an arbitrary but fixed asymptotic design F. Under weak assumptions, we have derived the asymptotically uniformly most powerful size α test in Theorem 4.1 for a given design F. If we want to compare the (asymptotic) power of two tests with respect to two different designs F_1, F_2, we have to look at $(\Upsilon_{F_1}/\Upsilon_{F_2})^2$ (see Theorem 4.1 (b)) because of

$$-\beta_{d+1}\sqrt{n_1}\sigma^{-1}\Upsilon_{F_1} = -\beta_{d+1}\sqrt{n_2}\sigma^{-1}\Upsilon_{F_2} \Leftrightarrow n_2 = (\Upsilon_{F_1}/\Upsilon_{F_2})^2 n_1.$$

So, if we choose the design F_2 instead of the design F_1, we need (asymptotically) $(\Upsilon_{F_1}/\Upsilon_{F_2})^2$ times the number of design points in order to get (approximately) the same power. The ratio $(\Upsilon_{F_1}/\Upsilon_{F_2})^2$ is the asymptotic relative efficiency of the test with design F_1 relative to the test with design F_2.

Example 7.1. (Example 2.2 continued) Let us again consider: $d = 1$, $f(t) = f_1(t) = 1$, $f_2(t) = (t - t^*)^+$, $t \in \mathcal{E}$. Let F_1 be the uniform distribution on $\mathcal{E}$, and let F_2 be the measure that has masses $\frac{1}{2}$ at the points a and b. Let $p := (t^* - a)/(b - a)$. By Examples 4.2 and 4.3 it follows that

$$(\Upsilon_{F_1}/\Upsilon_{F_2})^2 = \frac{1}{3}(1-p)(1+3p).$$

For the data of Example 2.2 ($p = .651$) we have $(\Upsilon_{F_1}/\Upsilon_{F_2})^2 = .344$, so we need only 34.4 percent of the number of design points if we use F_2 instead of F_1 for getting the same power. This is very important in practice especially if the measurements are expensive. Moreover, $(\Upsilon_{F_1}/\Upsilon_{F_2})^2 \leq \frac{4}{9}$ holds for each $p \in (0,1)$.

In the previous example, we have seen that the design has a great influence on the power of the test. In this section, we look for the best choice of $F \in \mathcal{F}$, where $\mathcal{F}$ is defined in (3.5). Theorem 4.1 shows that, given $F \in \mathcal{F}$, the power of the optimal test is characterized by Υ_F. Thus, the "optimal" design maximizes the functional Υ_F with respect to $F \in \mathcal{F}$. So we get the following definition:

DEFINITION 7.1. Let the test problems (4.1) and (4.2) be given. Then a design $F_0 \in \mathcal{F}$ fulfilling

(7.1) $$\max_{F \in \mathcal{F}} \Upsilon_F = \Upsilon_{F_0}$$

is called an asymptotically optimal design (for testing H_0: "the function f_{d+1} does not belong to the model" against the alternative K: "f_{d+1} belongs to the model").

By Lemma 4.2, we know that

$$\Upsilon_F = (J_0 - a_0^\top J^{-1} a_0)^{1/2}, \qquad F \in \mathcal{F},$$

where $J_0 := \int_{\mathcal{E}} f_{d+1}(t)^2 F(dt)$ and $a_0 := \int_{\mathcal{E}} f(t) f_{d+1}(t) F(dt)$. With the definitions $\tilde{f} := (f_1, \ldots, f_{d+1})^\top$ and $e_{d+1} := (0, \ldots, 0, 1)^\top \in \mathbb{R}^{d+1}$, we can write

$$\Upsilon_F = \left(e_{d+1}^\top \left(\int_{\mathcal{E}} \tilde{f} \tilde{f}^\top dF \right)^{-1} e_{d+1} \right)^{-1/2}.$$

Solutions of the maximization problem (7.1) are known for the following special regression functions $f_1, \ldots, f_{d+1}$:

(a) $f_j(t) = t^{j-1}$, $j = 1, \ldots, d+1$, $t \in [a, b]$, see Dette (1994) and Dette and Studden (1997),

(b) $f_{2j+1}(t) = \cos(jt)$, $j = 0, \ldots, [(d-1)/2]$, $f_{2j}(t) = \sin(jt)$, $j = 1, \ldots, [d/2]$, $t \in [a, b]$, see Dette and Haller (1998).

For the special change-point models considered in our examples, we are able to determine optimal designs.

Example 7.1. Let $d = 1$ and $f_1(t) = f(t) \equiv 1$. For an arbitrary $F \in \mathcal{F}$ and an arbitrary function $f_2 : \mathcal{E} \to \mathbb{R}$ of bounded variation, by Lemma 4.2, we obtain:

$$\Upsilon_F^2 = \int_{\mathcal{E}} f_2(t)^2 F(dt) - \left(\int_{\mathcal{E}} f_2(t) F(dt) \right)^2.$$

Further, we assume that

$$\arg\max_{x \in \mathcal{E}} f_2(x) \neq \emptyset, \qquad \arg\min_{x \in \mathcal{E}} f_2(x) \neq \emptyset.$$

One can show that each $F_0 \in \mathcal{F}$ with mass $\frac{1}{2}$ at the points $\arg\min_{x \in \mathcal{E}} f_2(x)$ and with mass $\frac{1}{2}$ at the points $\arg\max_{x \in \mathcal{E}} f_2(x)$ solves Equation (7.1); the assertion can be proved by the same technique used in the following example.

If we especially choose $f_2 = 1_{[t^*, b]}(t)$ (Example 2.1), the optimal designs are given by the set of all $F_0 \in \mathcal{F}$ with $F_0(t^*-) = \frac{1}{2}$. Note that in the set of optimal designs, there exists a design that is independent of the change-point t^*, namely the design F_0^* that puts equal masses $\frac{1}{2}$ at a and b. So this design together with the optimal test, is even optimal in the case that the change-point is unknown!

If we choose $f_2(t) = (t - t^*)^+$ (Example 2.2), then for a known change-point $t^* \in (a, b)$ a design $F_0 \in \mathcal{F}$ is optimal if and only if $F_0(t^*-) = \frac{1}{2}$ and $F_0(b-) = \frac{1}{2}$. The design F_0^* is again an optimal design in the case where the change-point t^* is unknown.

Example 7.2. Let us consider the case $[a, b] = [0, 1]$, $t^* \in (0, 1)$ and $d = 2$ with the regression functions $f_1 \equiv 1$, $f_2(t) = t$, $f_3(t) = ((t - t^*)^+)^k$, where $k \geq 1$. Υ_F^2 is maximized by the design F_0 that has masses $\frac{1-s}{2}$, $\frac{1}{2}$, $\frac{s}{2}$ at the points 0, s, 1, respectively, where s is defined by

$$s := \begin{cases} t^*, & \text{if} \quad k = 1, \\ t^* + (1 - t^*)^{k/(k-1)} k^{-1/(k-1)}, & \text{if} \quad k > 1. \end{cases}$$

This result is proved by using an equivalence criterion; see, for example, Silvey (1980), p. 45, or Pukelsheim (1993), p. 176. For F_0 being optimal, we have to show that

$$(f_3(z) - a_0^\top J^{-1}(f_1(z), f_2(z))^\top)^2 \leq J_0 - a_0^\top J^{-1} a_0, \tag{7.2}$$

holds for every $z \in [0,1]$, where $J_0 := \int_\mathcal{E} f_{d+1}(t)^2 F_0(dt)$ and $a_0 := \int_\mathcal{E} f(t) f_{d+1}(t) F_0(dt)$. For the design F_0 given above, we have

$$J_0 = \frac{1}{2}((s-t^*)^+)^{2k} + \frac{s}{2}(1-t^*)^{2k},$$

$$a_0 = \begin{pmatrix} \frac{1}{2}((s-t^*)^+)^k + \frac{s}{2}(1-t^*)^k \\ \frac{1}{2}s((s-t^*)^+)^k + \frac{s}{2}(1-t^*)^k \end{pmatrix}, \qquad J = \begin{pmatrix} 1 & s \\ s & s(1+s)/2 \end{pmatrix}.$$

Simple calculations yield

$$J_0 - a_0^\top J^{-1} a_0 = \frac{1}{4}(s(1-t^*)^k - (s-t^*)^k)^2.$$

So we can evaluate the difference of the right and the left-hand sides in (7.2) as follows:

$$\begin{aligned} &J_0 - a_0^\top J^{-1} a_0 - (((z-t^*)^+)^k - a_0^\top J^{-1}(1,z)^\top)^2 \\ &= \begin{cases} (1-t^*)^k z(s(1-t^*)^k - (s-t^*)^k - (1-t^*)^k z), & z \in [0,t^*), \\ ((1-t^*)^k z - (z-t^*)^k) \\ \quad \cdot((z-t^*)^k - (1-t^*)^k z + s(1-t^*)^k - (s-t^*)^k), & z \in [t^*,1]. \end{cases} \end{aligned} \tag{7.3}$$

In the case where $k = 1$ (that means $s = t^*$), it is easy to see that this expression is nonnegative for every $z \in [0,1]$. In the case where $k > 1$, some elementary calculations lead to the same result. So (7.2) follows.

Acknowledgements

We thank the referee for helpful comments on improving the representation of our results. Further, we wish to thank the quality control department of the GKN Sinter Metals company for supplying the data.

Appendix

Proof of Lemma 3.1. $m(s) = 0$, $s \in [0,1]$, is obvious. Next, for simplifying the notation, we define

$$v_z := \int_\mathcal{E} f(t) F_{0;z}(dt). \tag{A.1}$$

We have

$$\begin{aligned} K(s,z) = {} & \mathrm{Cov}(B(s), B(z)) - \mathrm{Cov}\left(B(s), v_z^\top J^{-1} \int_\mathcal{E} f(t) B(F_0(dt))\right) \\ & - \mathrm{Cov}\left(B(z), v_s^\top J^{-1} \int_\mathcal{E} f(t) B(F_0(dt))\right) \\ & + \mathrm{Cov}\left(v_s^\top J^{-1} \int_\mathcal{E} f(t) B(F_0(dt)), v_z^\top J^{-1} \int_\mathcal{E} f(t) B(F_0(dt))\right). \end{aligned}$$

For every $m \in \mathbb{N}$, let $a \le x_{m1} \le \cdots \le x_{mm} \le b$ be a partition of $\mathcal{E}$ with $\sup_{j=2}^{m}\{x_{mj} - x_{mj-1}\} \to 0$ for $m \to \infty$. Then we have

$$\begin{aligned}
&\operatorname{Cov}\left(B(s), \int_{\mathcal{E}} f(x)B(F_0(dx))\right) \\
&\quad = \operatorname{Cov}\left(B(s), \lim_{m\to\infty} \sum_{i=1}^{m} f(x_{mi})(B(F_0(x_{mi})) - B(F_0(x_{mi-1})))\right) \\
&\quad = \lim_{m\to\infty} \sum_{i=1}^{m} f(x_{mi})(\min\{s, F_0(x_{mi})\} - \min\{s, F_0(x_{mi-1})\}) \\
&\quad = \lim_{m\to\infty} \sum_{i=1}^{m} f(x_{mi})(F_{0;s}(x_{mi}) - F_{0;s}(x_{mi-1})) \\
&\quad = \int_{\mathcal{E}} f(x)F_{0;s}(dx) = v_s
\end{aligned}$$

and

$$\operatorname{Cov}\left(\int_{\mathcal{E}} f(x)B(F_0(dx)), \int_{\mathcal{E}} f(y)B(F_0(dy))\right) = \int_{\mathcal{E}} f(x)f(x)^\top F_0(dx) = J.$$

Thus, the assertion of the lemma follows. □

PROOF OF LEMMA 4.1. We define $F_{n;z}(t) = \min\{F_n(t), z\}$ and $\mathbf{1}_{n;z} = (\gamma_i)_{i=1}^{n} \in \mathbb{R}^n$, where $\gamma_i = 1$, if $i \le [nz]$, $\gamma_{[nz]+1} = nz - [nz]$ and $\gamma_i = 0$, if $i > [nz] + 1$. For each $z \in [0, 1]$, we have

$$\begin{aligned}
&T_n\left(\frac{1}{n}\mathrm{pr}_{X_n^\perp}\boldsymbol{\xi}_n\right)(z) \\
&\quad = T_n\left(\frac{1}{n}\boldsymbol{\xi}_n\right)(z) - T_n\left(\frac{1}{n}X_n\left(\frac{1}{n}X_n^\top X_n\right)^{-1}\left(\frac{1}{n}X_n^\top\boldsymbol{\xi}_n\right)\right)(z) \\
&\quad = \left(\frac{1}{n}\mathbf{1}_{n;z}^\top\boldsymbol{\xi}_n\right) - \left(\frac{1}{n}\mathbf{1}_{n;z}^\top X_n\right)\left(\frac{1}{n}X_n^\top X_n\right)^{-1}\left(\frac{1}{n}X_n^\top\boldsymbol{\xi}_n\right) \\
&\quad = \int_{\mathcal{E}} f_{d+1}(t)F_{n;z}(dt) \\
&\qquad - \left(\int_{\mathcal{E}} f(t)F_{n;z}(dt)\right)^\top\left(\int_{\mathcal{E}} f(t)^\top f(t)F_n(dt)\right)^{-1}\left(\int_{\mathcal{E}} f_{d+1}(t)f(t)F_n(dt)\right) \\
&\quad \overset{n\to\infty}{\longrightarrow} \int_{\mathcal{E}} f_{d+1}(t)F_{0;z}(dt) \\
&\qquad - \left(\int_{\mathcal{E}} f(t)F_{0;z}(dt)\right)^\top\left(\int_{\mathcal{E}} f(t)^\top f(t)F_0(dt)\right)^{-1}\left(\int_{\mathcal{E}} f_{d+1}(t)f(t)F_0(dt)\right).
\end{aligned}$$

The last pointwise convergence (for each $z \in [0, 1]$) holds true because (3.4) implies the weak convergence of F_n to F_0 and $F_{n;z}$ to $F_{0;z}$, respectively, if f_{d+1} is continuous. But if we assume that f_{d+1} is of bounded variation, then the convergence can be seen as follows: It is easy to show that the product of two functions of bounded variation has this property as well. So we get by integration by parts for signed measures (cf., for example, Hewitt and Stromberg (1969) p. 419) and by Lebesgue's dominated convergence theorem:

$$\int_{\mathcal{E}} f_{d+1}(t)f(t)F_n(dt)$$

$$= f_{d+1}(b)f(b) - \int_{\mathcal{E}} F_n(t-)(f_{d+1}f)(dt)$$
$$\overset{n\to\infty}{\longrightarrow} f_{d+1}(b)f(b) - \int_{\mathcal{E}} F_0(t-)(f_{d+1}f)(dt)$$
$$= \int_{\mathcal{E}} f_{d+1}(t)f(t)F_0(dt),$$

where the convergence $F_n(t-) \to F_0(t-)$ follows by assumption (3.4). Next, we show uniform convergence. For the first integral

$$\sup_{z\in[0,1]}\left|\int_{\mathcal{E}} f_{d+1}(t)(F_{n;z} - F_{0;z})(dt)\right| = \sup_{z\in[0,1]}\left|\int_{\mathcal{E}} F_{n;z}(t-) - F_{0;z}(t-)f_{d+1}(dt)\right|$$
$$\leq \sup_{z\in[0,1]}\int_{\mathcal{E}} |F_{n;z}(t-) - F_{0;z}(t-)||f_{d+1}|(dt)$$
$$= \int_{\mathcal{E}} |F_n(t-) - F_0(t-)||f_{d+1}|(dt) \overset{n\to\infty}{\longrightarrow} 0$$

holds true. The uniform convergence of the second integral can be shown in the same way. □

LEMMA A.1. *Let $K(\cdot,\cdot)$ be the covariance function of B_{f,F_0} given in Lemma 3.1, let the assumptions of Lemma 4.1 be fulfilled, let $f_1,\ldots,f_{d+1}$ be linearly independent in $L_2(F_0)$, and let $g(t)\equiv 1$, $t\in\mathcal{E}$, be a linear combination of $f_1,\ldots,f_d$. Let h be the function defined in Lemma 4.1. Then it holds true that*
(a) $\int_{[0,1]} K(s,z)\,(-f_{d+1}(Q_0(ds))) = h(z)$,
(b) $h \neq 0$ *in* $C[0,1]$.

PROOF. (a) We use integration by parts for signed measures (cf. Hewitt and Stromberg (1969) p. 419) for the following result:

$$-\int_{[0,1]} \min\{s,z\} f_{d+1}(Q_0(ds))$$
$$= -\int_{[0,z]} s f_{d+1}(Q_0(ds)) - \int_{(z,1]} z f_{d+1}(Q_0(ds))$$
$$= -z f_{d+1}(Q_0(z)) + \int_{[0,z]} f_{d+1}(Q_0(s-))ds - z f_{d+1}(Q_0(1)) + z f_{d+1}(Q_0(z))$$
$$= \int_{[0,z]} f_{d+1}(Q_0(s-))ds - z f_{d+1}(Q_0(1))$$
$$= \int_{\mathcal{E}} f_{d+1}(t)F_{0;z}(dt) - z f_{d+1}(Q_0(1)).$$

Further, by Fubini, we get

$$-\int_{[0,1]}\int_{\mathcal{E}} f(t)F_{0;s}(dt) f_{d+1}(Q_0(ds))$$
$$= -\int_{[0,1]}\int_{[0,s]} f(Q_0(z-))dz\, f_{d+1}(Q_0(ds))$$

$$= -\int_{[0,1]}\int_{[z,1]} f_{d+1}(Q_0(ds))f(Q_0(z-))dz$$
$$= -\int_{[0,1]} (f_{d+1}(Q_0(1)) - f_{d+1}(Q_0(z-)))f(Q_0(z-))dz$$
$$= -\int_{\mathcal{E}} (f_{d+1}(Q_0(1)) - f_{d+1}(t))f(t)F_0(dt).$$

So we obtain

$$\text{(A.2)} \quad \int_{[0,1]} K(s,z)(-f_{d+1}(Q_0(ds))) = h(z) + f_{d+1}(Q_0(1))\left(v_z^\top J^{-1}\int_{\mathcal{E}} f(t)F_0(dt) - z\right),$$

where v_z is defined in (A.1). Without loss of generality, we assume that $f_1 \equiv 1$ and $\{f_1, \ldots, f_d\}$ is a set of orthonormal vectors in $L_2(F_0)$ (In the general case, we can choose a $d \times d$-matrix A of full rank and put $g(t) := A \cdot f(t)$ so that $g_1 \equiv 1$ and $\{g_1, \ldots, g_d\}$ is a set of orthonormal vectors). So we obtain $J = I_d$ and $\int_{\mathcal{E}} f(t)F_0(dt) = (1, 0, \ldots, 0)^\top$. Because of $v_z = (z, \ldots)^\top$ we see that the second summand in (A.2) is zero.

(b) Obviously,

$$p(t) := f_{d+1}(t) - f(t)^\top J^{-1}\int f_{d+1}(s)f(s)F_0(ds) \neq 0 \quad \text{in} \quad L_2(F_0)$$

by the assumption that $f_1, \ldots, f_{d+1}$ are linearly independent in $L_2(F_0)$. Hence, the assertion follows since $h(z) = \int_{\mathcal{E}} p(t)F_{0;z}(dt)$. □

Proof of Theorem 4.1. (a) Given $y \sim \gamma h + B_{f,F_0}$ and given the test problem $H_0 : \gamma = 0$ against $K : \gamma > 0$, it follows by Luschgy ((1991), Proposition 1 and the following considerations) that the uniformly most powerful test statistic is given by

$$u \mapsto \left(\int_{[0,1]} h(z)\mu(dz)\right)^{-1/2} \cdot \int_{[0,1]} u(z)\mu(dz), \qquad u \in C[0,1],$$

where μ is a signed measure being the solution of

$$\text{(A.3)} \qquad \int_{[0,1]} K(s,z)\mu(ds) = h(z).$$

Note that $K(s,z)$ is the covariance function of B_{f,F_0} given in Lemma 3.1 and h is given in Lemma 4.1. Lemma A.1 (a) yields that the signed measure μ induced by the function $-f_{d+1} \circ Q_0$ which is of bounded variation solves (A.3). Note that the assumptions (A2) and (A3) of Luschgy (1991) are satisfied because we have a solution of (A.3) and $h \neq 0$ in $C[0,1]$, see Lemma A.1 (b). Note further that Υ_{F_0} is well defined and positive because Υ_{F_0} is the norm of h in the reproducing kernel Hilbert space considered in Luschgy (1991).

The signed measure induced by $f_{d+1} \circ Q_0$ is the image measure of the signed measure induced by f_{d+1} under the mapping $t \mapsto F_0(t-)$. So it follows, from Hewitt and Stromberg ((1969), p. 180), that

$$\text{(A.4)} \qquad \int_{[0,1]} u(z)f_{d+1}(Q_0(dz)) = \int_{\mathcal{E}} u(F_0(t-))f_{d+1}(dt)$$

for $u(z) \in C[0,1]$ (especially also for $u(z) = h(z)$).
(b) The assertions of (b) follow directly by (1.3) in Luschgy (1991). □

PROOF OF LEMMA 4.2. We have (cp. the proof of Lemma A.1 (a)):

$$\int_{[0,1]} \int_{\mathcal{E}} f(t) F_{0;z}(dt) f_{d+1}(Q_0(dz)) = \int_{\mathcal{E}} f(t)(f_{d+1}(Q_0(1)) - f_{d+1}(t)) F_0(dt).$$

Analogously, we can show that

$$\int_{[0,1]} \int_{\mathcal{E}} f_{d+1}(t) F_{0;z}(dt) f_{d+1}(Q_0(dz)) = \int_{\mathcal{E}} f_{d+1}(t)(f_{d+1}(Q_0(1)) - f_{d+1}(t)) F_0(dt).$$

Hence, using (A.4), we get

$$\begin{aligned}
\Upsilon_{F_0}^2 &= -\int_{[0,1]} h(z) f_{d+1}(Q_0(dz)) \\
&= -\int_{[0,1]} \int_{\mathcal{E}} f_{d+1}(t) F_{0;z}(dt) f_{d+1}(Q_0(dz)) \\
&\quad + \left(\int_{[0,1]} \int_{\mathcal{E}} f(t)\, F_{0;z}(dt) f_{d+1}(Q_0(dz)) \right)^{\top} J^{-1} \left(\int_{\mathcal{E}} f_{d+1}(t) f(t)\, F_0(dt) \right) \\
&= -\int_{\mathcal{E}} f_{d+1}(t)(f_{d+1}(Q_0(1)) - f_{d+1}(t)) F_0(dt) \\
&\quad + \left(\int_{\mathcal{E}} f(t)(f_{d+1}(Q_0(1)) - f_{d+1}(t)) F_0(dt) \right)^{\top} J^{-1} \left(\int_{\mathcal{E}} f_{d+1}(t) f(t)\, F_0(dt) \right).
\end{aligned}$$

Similarly, as at the end of the proof of Lemma A.1 (a), we recognize

$$f_{d+1}(Q_0(1)) \left(\int_{\mathcal{E}} f_{d+1}(t) F_0(dt) - \left(\int_{\mathcal{E}} f(t) F_0(dt) \right)^{\top} J^{-1} \left(\int_{\mathcal{E}} f_{d+1}(t) f(t) F_0(dt) \right) \right) = 0.$$

Thus, Lemma 4.2 follows. □

PROOF OF LEMMA 5.1. (a) The "likelihood ratio" statistic can be written in the following way:

$$C_n(\boldsymbol{Y}_n) = \sigma^{-1}(\tilde{\boldsymbol{\xi}}_n^{\top} \tilde{\boldsymbol{\xi}}_n)^{-1/2} \tilde{\boldsymbol{\xi}}_n^{\top} \boldsymbol{Y}_n = (n^{-1} \boldsymbol{\xi}_n^{\top} \mathrm{pr}_{X_n^{\perp}} \boldsymbol{\xi}_n)^{-1/2} (\sigma\sqrt{n})^{-1} \boldsymbol{\xi}_n^{\top} \mathrm{pr}_{X_n^{\perp}} \boldsymbol{Y}_n$$

One can show that the following two equations hold true:

$$\begin{aligned}
n^{-1} \boldsymbol{\xi}_n^{\top} \mathrm{pr}_{X_n^{\perp}} \boldsymbol{\xi}_n &= -\int_{\mathcal{E}} h_{F_n, f, f_{d+1}}(F_n(t-)) f_{d+1}(dt) = \Upsilon_{F_n}^2, \\
(\sigma\sqrt{n})^{-1} \boldsymbol{\xi}_n^{\top} \mathrm{pr}_{X_n^{\perp}} \boldsymbol{Y}_n &= -\int_{\mathcal{E}} y_n(F_n(t-)) f_{d+1}(dt).
\end{aligned}$$

Hence, the assertion of part (a) follows.

(b) We have

$$\Upsilon_{F_n}^2 = \int_{\mathcal{E}} f_{d+1}(t)^2 F_n(dt) - a_n^\top \left(\int_{\mathcal{E}} f(t) f(t)^\top F_n(dt) \right)^{-1} a_n$$
$$\stackrel{n\to\infty}{\longrightarrow} \int_{\mathcal{E}} f_{d+1}(t)^2 F_0(dt) - a_0^\top J^{-1} a_0 = \Upsilon_{F_0}^2,$$

where $a_n := \int_{\mathcal{E}} f(t) f_{d+1}(t) F_n(dt)$, $n \in \mathbb{N} \cup \{0\}$. The convergence can be shown in a similar way as in the proof of Lemma 4.1 in this appendix. Further, it holds true for $u \in C[0,1]$:

$$\lim_{n\to\infty} \int_{\mathcal{E}} u(F_n(t-)) f_{d+1}(dt) = \int_{\mathcal{E}} u(F_0(t-)) f_{d+1}(dt).$$

So the assertion of part (b) follows. □

References

Arnold, S. F. (1981). *The Theory of Linear Models*, Wiley, New York.

Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.

Bischoff, W. (1996). Properties of certain change-point test statistics and sample path behavior of residual partial sums processes (preprint).

Bischoff, W. (1998). A functional central limit theorem for regression models, *Ann. Statist.*, **26**, 1398–1410.

Bischoff, W., Lo Huang, M. and Yang, L. (1999). Growth curve models with random parameter for stochastic modelling and analyzing of natural disinfection of wastewater (preprint).

Brown, R. L., Durbin, J. and Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time, *J. Roy. Statist. Soc. Ser. B*, **37**, 149–192.

Dette, H. (1994). Discrimination designs for polynomial regression on a compact interval, *Ann. Statist.*, **23**, 1248–1267.

Dette, H. and Haller, G. (1998). Optimal designs for the identification of the order of a Fourier regression, *Ann. Statist.*, **26**, 1496–1521.

Dette, H. and Studden, W. J. (1997). *The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis*, Wiley, New York.

Gardner, L. A. (1969). On detecting changes in the mean of normal variates, *Ann. Math. Statist.*, **40**, 116–126.

Hewitt, E. and Stromberg, K. (1969). *Real and Abstract Analysis*, Springer, Berlin.

Jandhyala, V. K. and MacNeill, I. B. (1991). Tests for parameter changes at unknown times in linear regression models, *J. Statist. Plann. Inference*, **27**, 291–316.

Luschgy, H. (1991). Testing one-sided hypotheses for the mean of a Gaussian process, *Metrika*, **38**, 179–194.

MacNeill, I. B. (1978*a*). Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times, *Ann. Statist.*, **6**, 422–433.

MacNeill, I. B. (1978*b*). Limit processes for sequences of partial sums of regression residuals, *Ann. Probab.*, **6**, 695–698.

Pukelsheim, F. (1993). *Optimal Design of Experiments*, Wiley, New York.

Sacks, J. and Ylvisaker, D. (1966). Design for regression problems with correlated errors, *Ann. Math. Statist.*, **37**, 66–89.

Sen, A. and Srivastava, M. S. (1975). On tests for detecting change in mean, *Ann. Statist.*, **3**, 98–108.

Sen, P. K. (1982). Invariance principles for recursive residuals, *Ann. Statist.*, **10**, 307–312.

Silvey, S. D. (1980). *Optimal design*, Chapman and Hall, London.

Tang, S. M. and MacNeill, I. B. (1993). The effect of serial correlation on tests for parameter change at unknown time, *Ann. Statist.*, **21**, 552–575.

Watson, G. S. (1995). Detecting a change in the intercept in multiple regression, *Statist. Probab. Lett.*, **23**, 69–72.