Journal of Multivariate Analysis 102 (2011) 281–291

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

# Partial sum process to check regression models with multiple correlated response: With an application for testing a change-point in profile data

W. Bischoff*, A. Gegg

*Faculty of Mathematics and Geography, Catholic University Eichstätt-Ingolstadt, D-85071 Eichstätt, Germany*

ARTICLE INFO

*Article history:*
Received 1 December 2009
Available online 19 September 2010

*AMS subject classifications:*
60F05
60G15
62G10
62G20
62J05

*Keywords:*
Multiple linear regression model
Multiple residual partial sum limit process
Multivariate Brownian motion
Change-point problem
Panel data
Profile data
Repeated measurements

ABSTRACT

We consider regression models with multiple correlated responses for each design point. Under the null hypothesis, a linear regression is assumed. For the least-squares residuals of this linear regression, we establish the limit of the partial sums. This limit is a projection on a certain subspace of the reproducing Kernel Hilbert space of a multivariate Brownian motion. Based on this limit, we propose a significance test of Kolmogorov–Smirnov type to test the null hypothesis and show that this result can be used to study a change-point problem in the case of linear profile data (panel data). We compare our proposed method, which does not rely on any distributional assumptions, with the likelihood ratio test in a simulation study.



## 1. Introduction

In order to find change points in regression models, it is popular to investigate the partial sums of the least-squares residuals. A non-parametric test (as for instance a test of Kolmogorov–Smirnov type) can then be applied to the limit process of the partial sums in order to test whether a change point does or does not occur. This technique can also be used to check asymptotically for linear regression. Note that MacNeill [12,11] and Bischoff [2,3] studied this residual partial sum process (RPSP) for univariate response and univariate experimental region whereas Xie and MacNeill [16] as well as Bischoff and Somayasa [4] investigated the case of univariate response with a multivariate experimental region. In this paper, we consider linear regression models with multivariate correlated responses for each design point of a univariate experimental region.

In Section 2, we describe in detail the regression model under consideration in this paper. It is a regression model with multiple correlated responses. Moreover, we formulate the linear regression model under the null hypothesis. In Section 3, we present our main result: the $p$-dimensional RPSP for a regression model with $p$-variate response. We use these results to attack the change-point problem in the case of linear profile data (in econometrics also known as "panel data", see Section 4) and establish a size $\alpha$-test in Section 5. Mahmoud et al. [13] proposed a modification of a likelihood ratio test (LRT) for the case of simple linear regression with normally distributed error terms and showed in a simulation study that this test is superior to other competing control charting approaches ($F$-test by Mahmoud and Woodall [14] and a Shewart-type control

* Corresponding author.
*E-mail address:* wolfgang.bischoff@ku-eichstaett.de (W. Bischoff).

0047-259X/$ – see front matter 
doi:10.1016/j.jmva.2010.08.014

chart by Kim et al. [10]). For these approaches, an exact specification of the alternative hypothesis is necessary as opposed to our procedure. Furthermore, by asymptotic considerations our method does not need assumptions about the distribution of the error terms. In contrast, the approach of Mahmoud et al. [13] heavily depends on the normal distribution. This is also shown in our simulation study (Section 6). The paper ends with a conclusion in Section 7. Proofs can be found in the last section.

## 2. Regression model with multiple correlated responses

In this paper, we consider a compact interval $[a, b] \subseteq \mathbb{R}$ as the experimental region. Without loss of generality, it is assumed $[a, b] = [0, 1]$ in the following. In practice, one has to decide a change-point problem by the information of data coming from an experimental design with a finite number of design points. To be able to carry out asymptotic considerations, we consider a triangular array of design points:

$$0 \le t_{m,1} < t_{m,2} < \cdots < t_{m,m} \le 1, \quad m \in \mathbb{N}.$$

Moreover, we suppose an equidistant design (i.e. $t_{m,k} = \frac{k}{m}$ for $k = 1, \ldots, m$) to simplify notations. At a first glance let $m$ be fixed, that is we have $m$ distinct design points. Later we are interested in large sample results, i.e. $m \to \infty$. In each design point, we observe $p$ responses. Let the vector of responses in $t_{m,k}$ be denoted by $Y^{(k)} = (Y_1^{(k)}, \ldots, Y_p^{(k)})^\top \in \mathbb{R}^p$. As usual, we consider an element $x \in \mathbb{R}^p$ as a column vector and let $x^\top$ be the corresponding row vector. For $Y^{(k)},\ k = 1, \ldots, m$, the true regression model is given by

$$Y^{(k)^\top} = g^\top(t_{m,k}) + Z^{(k)^\top}, \quad \text{with } \mathbf{E}\, Z^{(k)} = \mathbf{0},\ \mathbf{Cov}\, Z^{(k)} = \Sigma, \tag{1}$$

where $g^\top = (g_1, \ldots, g_p) : [0, 1] \to \mathbb{R}^p$ is the true, but unknown vector of regression functions. For technical reasons, we assume that $g$ is of bounded variation (i.e. $g \in BV([0, 1], \mathbb{R}^p)$), which is no restriction in practice. Furthermore, $\mathbf{0} \in \mathbb{R}^p$ is the vector with components equal to 0 and $\Sigma$ is a positive definite $p \times p$ matrix. In the application of our results to panel data below, $\Sigma$ will be known. Let

$$Y := \begin{pmatrix} Y^{(1)^\top} \\ \vdots \\ Y^{(m)^\top} \end{pmatrix}, \qquad \tau_m := \begin{pmatrix} t_{m,1} \\ \vdots \\ t_{m,m} \end{pmatrix}, \qquad g^\top(\tau_m) := \begin{pmatrix} g^\top(t_{m,1}) \\ \vdots \\ g^\top(t_{m,m}) \end{pmatrix}, \qquad Z := \begin{pmatrix} Z^{(1)^\top} \\ \vdots \\ Z^{(m)^\top} \end{pmatrix}.$$

We assume that $Z^{(1)}, \ldots, Z^{(m)}$ are independent. Hence, we get the following regression model for the response when the design $\tau_m$ is used:

$$Y = g^\top(\tau_m) + Z, \quad \text{with } \mathbf{E}\, Z = \mathbf{0},\ \mathbf{Cov}\,(\mathrm{vec}(Z^\top)) = \mathbf{I}_m \otimes \Sigma. \tag{2}$$

Thereby $\mathbf{I}_m$ is the $m \times m$ identity matrix, "$\otimes$" denotes the Kronecker-Product and "vec" is the well-known vec-operator (cf. [7, Ch. 16]).

Let $f_1, \ldots, f_d : [0, 1] \to \mathbb{R}$ be $d \in \mathbb{N}$ linearly independent functions in $C([0, 1]) \cap BV([0, 1])$, which are known, where $C([0, 1]) := C([0, 1], \mathbb{R})$ is the space of continuous real-valued functions on $[0, 1]$. We are interested in testing whether model (2) is a linear regression model, that is whether the null hypothesis

$$g^\top(\tau_m) = (f_1(\tau_m), \ldots, f_d(\tau_m))\, \Gamma \quad \text{with } \Gamma \in \mathbb{R}^{d\times p} \text{ unknown parameter matrix} \tag{3}$$

holds true. Hence, if the null hypothesis (3) is true, then we have the linear regression model

$$\begin{aligned} Y &= (f_1(\tau_m), \ldots, f_d(\tau_m))\, \Gamma + Z, \quad \text{with } \mathbf{E}\, Z = \mathbf{0},\ \mathbf{Cov}\,(\mathrm{vec}(Z^\top)) = \mathbf{I}_m \otimes \Sigma \\ &\quad \text{and } \Gamma \in \mathbb{R}^{d\times p} \text{ unknown parameter matrix.} \end{aligned} \tag{4}$$

Let $W := [f_1, \ldots, f_d]$ be the linear subspace spanned by $f_1, \ldots, f_d$. Then, we can formulate the above test problem by

$$H_0 : g_j \in W, \quad j = 1, \ldots, p \tag{5}$$

against the alternative

$$H_1 \colon \exists j \in \{1, \ldots, p\} \quad \text{with } g_j \notin W. \tag{6}$$

Putting $W^p := W \times \cdots \times W$, Eqs. (5)–(6) are equivalent to

$$H_0 \colon g = (g_1, \ldots, g_p)^\top \in W^p \quad \text{vs.} \quad H_1 : g \notin W^p. \tag{7}$$

Under the null hypothesis (5), we have for all $j \in \{1, \ldots, p\}$

$$g_j(\tau_m) \in [f_1(\tau_m), \ldots, f_d(\tau_m)] =: W_m \le \mathbb{R}^m, \tag{8}$$

where "$W_m \leq \mathbb{R}^m$" means that $W_m$ is a linear subspace of $\mathbb{R}^m$. We can reformulate Eq. (8) according to

$$g(\tau_m) \in W_m^p := W_m \times \cdots \times W_m \leq \mathbb{R}^{m\times p} = \mathbb{R}^m \times \cdots \times \mathbb{R}^m.$$

We assume that $m$ is large enough that the vectors $f_1(\tau_m), \ldots, f_d(\tau_m)$ are linearly independent and so form a basis of $W_m$.

Furthermore, for $m$ large enough, we have $g(\tau_m) \notin W_m^p$ in the case $g \notin W^p$. Consequently, we can decide the original test by

$$H_0 : g(\tau_m) \in W_m^p \quad \text{vs.} \quad H_1 : g(\tau_m) \notin W_m^p.$$

Note that the space of real $m \times p$-matrices is a Hilbert space endowed with the inner product $\langle A, B\rangle_{\mathbb{R}^{m\times p}} := \sum_{j=1}^p \langle a^{(j)}, b^{(j)}\rangle_{\mathbb{R}^m} = \operatorname{trace}(A^T B)$ for $A = (a^{(1)}, \ldots, a^{(p)}), B = (b^{(1)}, \ldots, b^{(p)}) \in \mathbb{R}^{m\times p}$. Thereby $\langle \cdot, \cdot\rangle_{\mathbb{R}^m}$ is the Euclidean scalar product in $\mathbb{R}^m$.

## 3. Residual partial sum process

Let $L_2 := L_2^1([0, 1], \lambda)$ be the Hilbert space of real functions on $[0, 1]$, which are square integrable with respect to the Lebesgue measure $\lambda$. As usual $L_2$ is furnished with the inner product $\langle h, \tilde{h}\rangle_{L_2} := \int_{[0,1]} h\tilde{h}\mathrm{d}\lambda$, $h, \tilde{h} \in L_2$.

Then $L_2^p = \{h : [0, 1] \to \mathbb{R}^p \mid \int_{[0,1]} h^\top h \mathrm{d}\lambda < \infty\} = L_2 \times \cdots \times L_2$, endowed with the inner product $\langle h, \tilde{h}\rangle_{L_2^p} := \sum_{j=1}^p \langle h_j, \tilde{h}_j\rangle_{L_2} = \int_{[0,1]} h^\top \tilde{h}\mathrm{d}\lambda$ for $h = (h_1, \ldots, h_p)^\top, \tilde{h} = (\tilde{h}_1, \ldots, \tilde{h}_p)^\top \in L_2^p$, is a Hilbert space. Furthermore, $W^p = [f_1, \ldots, f_d]^p$ is a linear subspace of $L_2^p$, where $f_1, \ldots, f_d$ are known functions, see Section 2.

In order to test the hypotheses (7), we investigate the partial sums of the $p$-dimensional residuals in model (3). For that we use the partial sum operator $T_m$, which embeds a vector $a = (a_1, \ldots, a_m)^\top \in \mathbb{R}^m$ in the space $C([0, 1])$ by

$$T_m\left(\begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix}\right)(z) = \sum_{i=1}^{\lfloor mz\rfloor} a_i + (mz - \lfloor mz\rfloor)a_{\lfloor mz\rfloor+1}, \quad z \in [0, 1]$$

where $\lfloor z\rfloor := \max\{\tilde{z} \in \mathbb{Z} \mid \tilde{z} \leq z\}$ and $\sum_{i=1}^0 a_i = 0$. The partial sum operator $T_{m\times p}$ embeds $\mathbb{R}^{m\times p} = \mathbb{R}^m \times \cdots \times \mathbb{R}^m$ in the space $C([0, 1], \mathbb{R}^p) = C([0, 1]) \times \cdots \times C([0, 1])$. We define $T_{m\times p}$ with the help of $T_m$. For this, let $A \in \mathbb{R}^{m\times p}$ be a $m \times p$ matrix with columns $a^{(1)}, \ldots, a^{(p)}$, then

$$T_{m\times p} : \begin{cases} \mathbb{R}^{m\times p} & \to \quad C([0, 1]) \times \cdots \times C([0, 1]) \\ \quad A & \mapsto \quad T_{m\times p}(A)(z) = \left(T_m(a^{(1)})(z), \ldots, T_m(a^{(p)})(z)\right), \quad z \in [0, 1]. \end{cases}$$

Note that $T_m$, and so $T_{m\times p}$, is a linear mapping.

In the following, we make use of the vector-valued version of the Donsker theorem (see [8]).

**Theorem 1.** *Let $(\xi_i)_{i\geq 1}$ be an i.i.d. sequence of random variables with values in $\mathbb{R}^p$ and $\mathbf{E}\,\xi_1 = \mathbf{0}$, $\mathbf{Cov}\,\xi_1 = \Sigma$, $\Sigma$ positive definite. Then*

$$\frac{1}{\sqrt{m}}\Sigma^{-1/2}T_{m\times p}\left(\begin{pmatrix} \xi_1^\top \\ \vdots \\ \xi_m^\top \end{pmatrix}\right)^\top \xrightarrow{\mathcal{D}} B^p \quad \text{with } m \to \infty,$$

*whereas $B^p$ is the p-dimensional Brownian motion with independent components and "$\xrightarrow{\mathcal{D}}$" means weak convergence.*

The residuals of the linear model (2) are correlated and by that they do not fulfill the i.i.d. assumption of the preceding theorem. Hence, we cannot directly use this vector-valued version of the Donsker theorem to establish the $p$-dimensional RPSP.

For a mapping $h : [0, 1] \to \mathbb{R}$, we define in the following $h(\tau_m)$ as above (i.e. $h(\tau_m) = \left(h(t_{m,1}), \ldots, h(t_{m,m})\right)^\top \in \mathbb{R}^m$) and

$$s_h : [0, 1] \to \mathbb{R} \quad \text{with } s_h(t) = \int_0^t h(z)\mathrm{d}z, \quad t \in [0, 1].$$

The function $s_h$ is also known as the signal coming from $h$.

The functions $g_j, f_i$ are of bounded variation and consequently Riemann integrable. Furthermore, the design $\tau_m$ is uniform (i.e. $t_{m,k} = \frac{k}{m}$ for $k = 1, \ldots, m$) and so we obtain for $m \to \infty$,

$$\frac{1}{m}T_m\left(g_j(\tau_m)\right) \longrightarrow s_{g_j} \quad \text{in } C([0, 1]) \text{ for } j = 1, \ldots, p \tag{9}$$

$$\frac{1}{m}T_m\left(f_i(\tau_m)\right) \longrightarrow s_{f_i} \quad \text{in } C([0, 1]) \text{ for } i = 1, \ldots, d. \tag{10}$$

Putting $W_{\mathcal{H}} := \left[s_{f_1}, \ldots, s_{f_d}\right]$ the test problem (7) is equivalent to

$$H_0 : s_g \in W_{\mathcal{H}}^p := W_{\mathcal{H}} \times \cdots \times W_{\mathcal{H}} \quad \text{vs.} \quad H_1 : s_g \notin W_{\mathcal{H}}^p \tag{11}$$

where $s_g := \left(s_{g_1}, \ldots, s_{g_p}\right)^\top$. $W_{\mathcal{H}}$ is a linear subspace of

$$\mathcal{H} := \mathcal{H}^1 := \{s_h : [0, 1] \to \mathbb{R} \mid h \in L_2\},$$

the reproducing kernel Hilbert space (RKHS) of the one-dimensional Brownian motion (see [3]), which is furnished with the inner product $\langle s_h, s_{\tilde{h}}\rangle_{\mathcal{H}} = \langle h, \tilde{h}\rangle_{L_2}$ $(s_h, s_{\tilde{h}} \in \mathcal{H})$. Analogously

$$\mathcal{H}^p := \left\{s_h : [0, 1] \to \mathbb{R}^p \mid h \in L_2^p\right\} = \mathcal{H} \times \cdots \times \mathcal{H}$$

is the RKHS of the $p$-dimensional Brownian motion. $W_{\mathcal{H}}^p$ is a linear subspace of $\mathcal{H}^p$, where $\mathcal{H}^p$ is furnished with the inner product

$$\langle s_h, s_{\tilde{h}}\rangle_{\mathcal{H}^p} := \sum_{j=1}^{p} \langle s_{h_j}, s_{\tilde{h}_j}\rangle_{\mathcal{H}} = \langle h, \tilde{h}\rangle_{L_2^p} = \int_{[0,1]} h^\top \tilde{h} \mathrm{d}\lambda$$

for $s_h = (s_{h_1}, \ldots, s_{h_p})^\top, s_{\tilde{h}} = (s_{\tilde{h}_1}, \ldots, s_{\tilde{h}_p})^\top \in \mathcal{H}^p$.

Note that $\mathcal{H}^p$ is also a linear subspace of $C([0, 1], \mathbb{R}^p)$. Bischoff [3], see also [4], showed that there exists a projector $\mathrm{pr}_{W_{\mathcal{H}}} : C([0, 1]) \to W_{\mathcal{H}}$ which coincide with the orthogonal projector onto $W_{\mathcal{H}}$ in $\mathcal{H}$ if $\mathrm{pr}_{W_{\mathcal{H}}}$ is restricted to $\mathcal{H}$. Consequently,

$$\mathrm{pr}_{W_{\mathcal{H}}^p} : \begin{cases} C([0, 1]) \times \cdots \times C([0, 1]) & \to \quad W_{\mathcal{H}} \times \cdots \times W_{\mathcal{H}} \\ \quad\left(u_1, \ldots, u_p\right)^\top & \mapsto \quad \left(\mathrm{pr}_{W_{\mathcal{H}}}(u_1), \ldots, \mathrm{pr}_{W_{\mathcal{H}}}(u_p)\right)^\top \end{cases} \tag{12}$$

is the corresponding projector for the multivariate case. For a subspace $U$, we denote in the following, by $U^\perp$ the orthogonal space and by $\mathrm{pr}_U$ (resp. $\mathrm{pr}_{U^\perp} = \mathrm{id} - \mathrm{pr}_U$) the orthogonal projector onto $U$ (resp. $U^\perp$). With this notation, the least-squares residuals of the linear model under $H_0$ given in Eq. (4) can be expressed by $\mathrm{pr}_{W_m^{p\perp}}(Y) = Y - \mathrm{pr}_{W_m^p}(Y)$. Now we are in the position to state our main result.

**Theorem 2.** *Let $f_1, \ldots, f_d \in C([0, 1])$ be known, linearly independent functions with bounded variation, $W^p := [f_1, \ldots, f_d]^p$, $W_m^p := [f_1(\tau_m), \ldots, f_d(\tau_m)]^p$ and $W_{\mathcal{H}}^p := [s_{f_1}, \ldots, s_{f_d}]^p$.*

*Consider model (2) under the null hypothesis, i.e.*

$$Y = (f_1(\tau_m), \ldots, f_d(\tau_m))\, \Gamma + Z \quad \textit{with } \mathbf{E}\, Z = \mathbf{0},\ \mathbf{Cov}\,(Z^{(1)}) = \Sigma, Z^{(1)}, \ldots, Z^{(m)} \text{ i.i.d.}$$
$$\textit{and } \Gamma \in \mathbb{R}^{d\times p} \textit{ unknown parameter matrix.}$$

*Then we have for $m \to \infty$,*

$$\frac{1}{\sqrt{m}} \Sigma^{-1/2} T_{m\times p} \left(\mathrm{pr}_{W_m^{p\perp}}(Y)\right)^\top \xrightarrow{\mathcal{D}} \mathrm{pr}_{W_{\mathcal{H}}^{p\perp}}\left(B^p\right), \tag{13}$$

*where $B^p$ is the p-dimensional Brownian motion.*

The proof of this theorem can be found in the last section of this paper.

Note that we do not make any assumptions about the distribution of the error terms $Z^{(1)}, \ldots, Z^{(m)}$- except for the mean value and the covariance matrix.

## 4. Linear profile data

In this section, we show that a change-point problem for linear profile data can be described by the linear regression model introduced in Section 2. We get this result by a two-step procedure: first, we estimate the parameter vectors for each profile. In the second step, we analyze these estimates by a linear model with multiple correlated responses under the null hypothesis that the profile data have no change point.

In practice, it is of common interest to test whether all of a fixed number $m$, say, of independent samples follow the same known linear model. To be more precise let

$$W^{(j)} = \mathbf{X}\beta^{(j)} + \epsilon^{(j)}, \qquad \beta^{(j)} \in \mathbb{R}^p \text{ unknown} \tag{14}$$

be the true linear model for each $j \in \{1, \ldots, m\}$, where

(A1) $\mathbf{X} \in \mathbb{R}^{n\times p}$ is the corresponding design matrix of explanatory variables with $\mathrm{rank}(\mathbf{X}) = p \le n$, and
(A2) $\epsilon^{(j)}$ is the vector with i.i.d. components $\epsilon_1^{(j)}, \ldots, \epsilon_n^{(j)}$ having mean 0 and variance $\sigma^2$.

Model (14) is called the "$j$th linear profile" and the aim is to test

$$H_0{:}\beta = \beta^{(1)} = \cdots = \beta^{(m)}. \tag{15}$$

For that we estimate $\beta^{(j)}$ by the least-squares estimator $\hat{\beta}^{(j)}$. With our assumptions (A1)–(A2), we have $\hat{\beta}^{(j)} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T W^{(j)}$ with

$$\mathbf{E}\left(\hat{\beta}^{(j)}\right) = \beta^{(j)} \quad \text{and} \quad \mathbf{Cov}\left(\hat{\beta}^{(j)}\right) = \sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1} =: \Sigma. \tag{16}$$

In case $\sigma^2$ is unknown, our proposed procedure can also be used by replacing $\sigma^2$ with a consistent estimator for $\sigma^2$ under $H_0$ (see Section 5). So we can assume that without loss of generality, $\Sigma$ is a known positive definite matrix. Furthermore $\hat{\beta}^{(1)}, \ldots, \hat{\beta}^{(m)}$ are independent since the different samples

$W^{(1)}, \ldots, W^{(m)}$ are assumed to be independent.

Let $Y := \begin{pmatrix} \hat{\beta}^{(1)\top} \\ \vdots \\ \hat{\beta}^{(m)\top} \end{pmatrix}$ be the $m \times p$ matrix containing the least-squares estimations and let $Z := \begin{pmatrix} \hat{\beta}^{(1)\top} - \beta^\top \\ \vdots \\ \hat{\beta}^{(m)\top} - \beta^\top \end{pmatrix}$. Specifying the notation of Section 2, we have $d = 1$ and put $f_1 = \mathbf{1}$ where $\mathbf{1}$ is the constant mapping identical 1, i.e. $\mathbf{1}(t) \equiv 1 \in \mathbb{R}$ for all $t \in [0, 1]$. Furthermore $\Gamma = \beta^\top \in \mathbb{R}^{1\times p}$. If (15) holds true, then using the above notation we have the linear regression model

$$Y = f_1(\tau_m)\Gamma + Z \quad \text{with } \mathbf{E}\,Z = \mathbf{0}, \quad \mathbf{Cov}\left(\mathrm{vec}(Z^T)\right) = \mathbf{I}_m \otimes \Sigma \tag{17}$$

and $\Gamma^\top = \beta \in \mathbb{R}^p$ unknown parameter vector.

Thereby $f_1(\tau_m) =: \mathbf{1}_m \in \mathbb{R}^m$ is the vector whose components are all equal to 1.

Conversely, if (15) is false, then a change point occurred and (17) does not hold. Therefore, we can test hypothesis (15) by checking the linear model (17), see Section 5.

## 5. Test for linear profile data

As already mentioned in Section 4, model (17) is a specific form of model Eq. (4) with $d = 1$ and $f_1 = \mathbf{1}$. Consequently, we have $W = [\mathbf{1}]$ and $s_{\mathbf{1}}(t) = \int_{[0,t]} \mathbf{1}\mathrm{d}\lambda = t$ for $t \in [0, 1]$. Furthermore, as already mentioned in an example by Bischoff [3], $\mathrm{pr}_{[\mathbf{1}]}(u)(t) = (u(1) - u(0)) \cdot t$ for $u \in C([0, 1])$ and $t \in [0, 1]$. Consequently, projection (12) can be written as

$$\mathrm{pr}_{W^p_{\mathcal{H}}} : \begin{cases} C([0,1], \mathbb{R}^p) & \rightarrow & W^p_{\mathcal{H}} \\ u(\cdot) & \mapsto & (u(1) - u(0)) \cdot \mathrm{id}(\cdot) \end{cases}$$

whereas $\mathrm{id}(t) = t$ for all $t \in [0, 1]$. Under the null hypothesis "$g(\cdot) \in W^p$", the residual partial sum limit process (cf. Theorem 2) is given by $\mathrm{pr}_{W^{p\,\perp}_{\mathcal{H}}}(B^p) = B^p_0$, the so-called standard $p$-dimensional Brownian bridge on [0, 1]. An intuitive one-dimensional test statistic is the maximum of the Euclidean norm of the $p$-dimensional process. To be more precise, let

$$R_m(t) := \frac{1}{\sqrt{m}}\Sigma^{-1/2} T_{m\times p}\left(\mathrm{pr}_{W^{p\,\perp}_m}(Y)\right)^\top(t), \quad t \in [0, 1].$$

Then, our proposed test statistic is $\max_{t\in[0,1]} \|R_m(t)\|_{\mathbb{R}^p}$.

On http://www.ku-eichstaett.de/mgf/statistik/Forschung/R.en you can find a way how to compute the proposed test statistic in a convenient way using R [15].

Because of the "Continuous Mapping theorem", see [1], we have the following convergence under $H_0$:

$$\|R_m\|_{\mathbb{R}^p} \xrightarrow{\mathcal{D}} \|B^p_0\|_{\mathbb{R}^p} \quad \text{for } m \to \infty. \tag{18}$$

Note that the limit process is the well-known Bessel bridge. In order to check (17), we apply a test of Kolmogorov–Smirnov type to the Bessel bridge and we get an asymptotic size $\alpha$-test, $\alpha \in (0, 1)$, by

$$\text{Reject } H_0 \iff \sup_{t\in[0,1]} \|R_m(t)\|_{\mathbb{R}^p} > k_\alpha.$$

Thereby $k_\alpha > 0$ is a constant such that $\mathcal{P}(\sup_{t\in[0,1]} \|B^{0,p}(t)\|_{\mathbb{R}^p} > k_\alpha) = \alpha$. Note that for given $\alpha$, the corresponding value $k_\alpha$ can be explicitly calculated by the formula (see [9, Eq. (3.21)]):

$$P(\max_{t\in[0,1]} \|B^{0,p}_t\|_{\mathbb{R}^p} \le k_\alpha) = \frac{4}{\Gamma\left(\frac{p}{2}\right)\sqrt{2}^p k^p_\alpha} \sum_{n=1}^{\infty} \frac{\left(\gamma_{\frac{p-2}{2},n}\right)^{p-2} \cdot \exp\left(-\frac{\left[\gamma_{\frac{p-2}{2},n}\right]^2}{2k^2_\alpha}\right)}{\left[J_{p/2}\left(\gamma_{\frac{p-2}{2},n}\right)\right]^2} \tag{19}$$

where $\Gamma(\cdot)$ is the gamma function, $J_{\frac{p-2}{2}}$ is the modified Bessel function of the first kind and $\gamma_{\frac{p-2}{2},n}$ is the positive zero of $J_{\frac{p-2}{2}}, n = 1, 2, \ldots$.

Note that in case the model under study is not correct, the RPSP for large enough $m$ is approximatively given by

$$\mathrm{pr}_{W_{\mathcal{H}}^{p}\,\perp}(B^p) + \text{ drift-term.}$$

For such a result, a consistent estimation for the variance is needed under the alternative. There are several possibilities for such estimations, see, for instance, Hall and Marron [6] and the papers cited by Dette et al. [5]. Dette et al. [5] give a comparison of estimations for the variance. They point out that such estimations can be bad for a finite sample and a good choice of the estimator heavily depends on the true regression function. For our simulations under alternatives we apply, however, the usual variance estimator of the linear model. So we underestimate the power by our simulations on the one hand, but on the other hand we have not to assume any further knowledge on the regression function under the alternative.

## 6. Simulation study

We present simulations concerning the change-point situation in the case of linear profile data (see Section 4) for our proposed method (residual partial sum method: RPSM, see Section 5) and for the LRT developed by Mahmoud et al. [13]. For our analysis, we used R [15].

In the following, we use in principle the same simulation conditions as Mahmoud et al. [13]. So we have, by using the notation of Section 4, $m$ independent profiles, each of size $n = 10$ with a polynomial regression having $p$ parameters (i.e. $p = 2$ means straight line regression). The design matrix for each profile is assumed to be

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 0^2 & \cdots & 0^{p-1} \\ 1 & 0.2 & 0.2^2 & \cdots & 0.2^{p-1} \\ \vdots & & & & \\ 1 & 1.8 & 1.8^2 & \cdots & 1.8^{p-1} \end{pmatrix} \in \mathbb{R}^{n \times p}.$$

We set the variance $\sigma^2 := 1$ and the significance level for the test to $\alpha := 0.04$, as in [13]. Consequently, we get with (19) for the critical value of our proposed test $k_\alpha = 1.622751$.

For the error terms, we simulate three different distributions: normal, lognormal and gamma (each standardized).

### 6.1. *Number of profiles m fixed*

In this section, we only consider the case of a straight-line regression (i.e. $p = 2$) and two cases for $m$: 20 and 10. All results in this section are based on 1000 runs.

#### 6.1.1. *Situation under $H_0$*

Our simulations for $m = 20$ profiles show that the LRT do not meet the significance level in case we have non-normal distributed error terms. The significance level $\alpha = 0.04$ can only be achieved by the normal distribution:

- normal distribution: relative number of rejection = 0.037
- lognormal distribution: 0.757
- gamma distribution: 0.465.

So, the LRT heavily depends on the assumption of normally distributed errors, as even stated by Mahmoud et al. [13]. Consequently, for the lognormal and the gamma distribution the LRT cannot be applied. In contrast to that, our proposed RPSM is a distribution-free method. It achieves the significance level for all distributions under study (see Fig. 2).

#### 6.1.2. *Comparison under alternatives: normal error terms*

For the normal distribution, we compare the power of LRT with the power of RPSM. The types of shifts alternatives investigated in our simulation for the $m = 20$ profiles are sustained step shifts, as in [13], taking place after sample $m_0$ with $m_0 \in \{10, 15, 18, 19\}$. In particular, the following step shifts in the parameters were considered:

1. shifts in the intercept from 1 to $1 + \lambda \frac{\sigma}{\sqrt{n}} = 1 + \frac{\lambda}{\sqrt{10}}$ and
2. shifts in the slope from 1 to $1 + \delta \frac{\sigma}{\sqrt{S_{XX}}} = 1 + \frac{\delta}{\sqrt{8.124038}}$ (for a definition of $S_{XX}$ see [13])

with $\lambda, \delta \in \{0, 0.5, \ldots, 5\}$. Each value is based on 1000 runs.

Our comparison of RPSM and LRT in the normal distribution case leads to the results presented in Fig. 1. Consequently, in the case of normally distributed error terms, RPSM is similar to LRT if the change point occurs in or near the middle of profiles. Otherwise, LRT is superior. However, the proposed test by RPSM can be improved by a weighted test procedure.

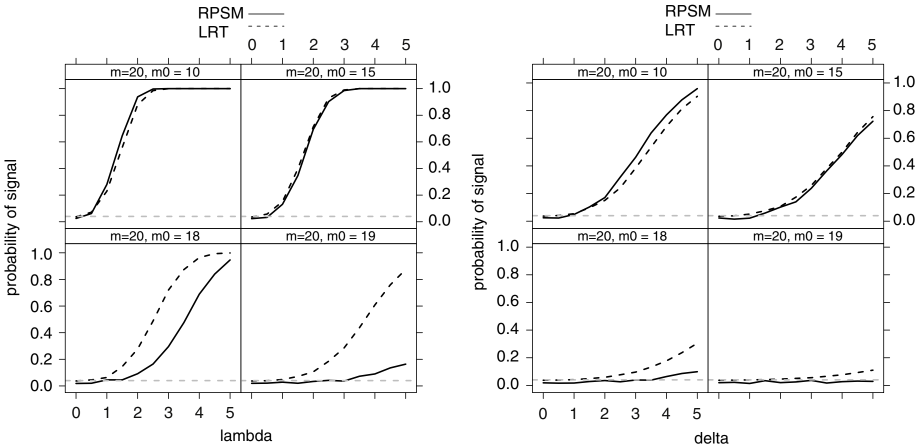


**Fig. 1.** Power of RPSM and LRT with normal distributed error terms by a change in the intercept (left) and slope (right) in $m_0$.

### *6.1.3. Comparison under alternatives: non-normal error terms*

In the non-normal case, Fig. 2 shows the power for $m = 10$ (with change point after profile $m_0 = 5, 8, 9$) and $m = 20$ (with $m_0 = 10, 15, 18, 19$) profiles. Again we investigate the case of a change point in the intercept term as well as in the slope term, see Section 6.1.2. Our proposed method fulfills the conditions for a size $\alpha$-test and the power increases for larger shifts. To be more precise, RPSM has the largest power, when the change point $m_0$ occurs in the middle of the $m$ profiles and power decreases as $m_0$ tends to the edge of its domain $\{0, 1, \ldots, m\}$. Furthermore, RPSM leads to moderate results already for $m = 10$ and $m_0 = 5, 8, 9$ (pictures in the lower row). As our asymptotic is in $m \to \infty$, power increases from $m = 10$ to $m = 20$ (see also Section 6.2).

In order to have a feeling about the variation of the value of our test statistic, we also present in Fig. 3 boxplots of the test statistics for the case $m = 10, m_0 = 5$ and $m = 20, m_0 = 10$ together with the reference value $k_\alpha = 1.622751$. Each value in both Fig. 2 and Fig. 3 is based on 1000 runs. The boxplots for all the other parameters look similar and are therefore omitted here.

### *6.2. Asymptotic considerations*

In this section, we present simulation results for $m \to \infty$, our asymptotic in Theorem 2. By this we want to give a feeling about the accuracy of the approximation by the asymptotic distribution. For that, we simulated values for $m = 10, 20, 30, \ldots, 200$ profiles and $p = 2, 4, 8$ as well as different values for $n$, whereby the number of simulation runs was 4000.

*Quality of the limit process.* The plots shown in the left column of Fig. 4 indicate that for $n = 10$ fixed and $p$ large (i.e. $p = 8$), the significance level $\alpha = 0.04$ (dotted line) can hardly be achieved. This is due to the fact that a polynomial regression with $p = 8$ parameters for $n = 10$ observations is an over-fitted model. If the ratio of parameters per profile $p$ to the number of observations per profile $n$ is more natural, then the $\alpha$-level can be achieved for moderate values of $m$. This can be seen by the right column of Fig. 4, where we have chosen $n = 5p$.

*Power.* Since the significance level $\alpha$ cannot be achieved for $n = 10$ and $p$ near $n$ (see preceding section), we only investigate the case $n = 5p$ in this section. All data points are based on 4000 runs.

In Fig. 5 we considered the case of a small change in the intercept (i.e. $\lambda = 1$) after profile $m_0 := \lfloor \frac{m}{3} \rfloor$ for $m = 10, 20, \ldots, 200$. Thereby $\lfloor z \rfloor := \max\{\tilde{z} \in \mathbb{Z} \mid \tilde{z} \leq z\}$ for $z \in \mathbb{R}$ as above.

## 7. Conclusion

We established the asymptotic distribution of the multivariate RPSP in a regression model with multivariate independent response variables $Y^{(1)}, \ldots, Y^{(m)} \in \mathbb{R}^p$ in Theorem 2.

We used this result in order to decide the change-point problem in the case of linear profile data, where each profile has the same design matrix $\mathbf{X}$. To be more precise, our proposed method can be used to test

$$H_0 : \beta^{(1)} = \beta^{(2)} = \cdots = \beta^{(m)}$$

where $\beta^{(j)} \in \mathbb{R}^p$ is the unknown parameter vector of the $j$th linear profile ($j = 1, \ldots, m$). The method under consideration is an asymptotic size $\alpha$-test, whereby the asymptotic is for $m \to \infty$.

We studied our RPSM in a simulation study and compared it with the modified LRT by Mahmoud et al. [13]:

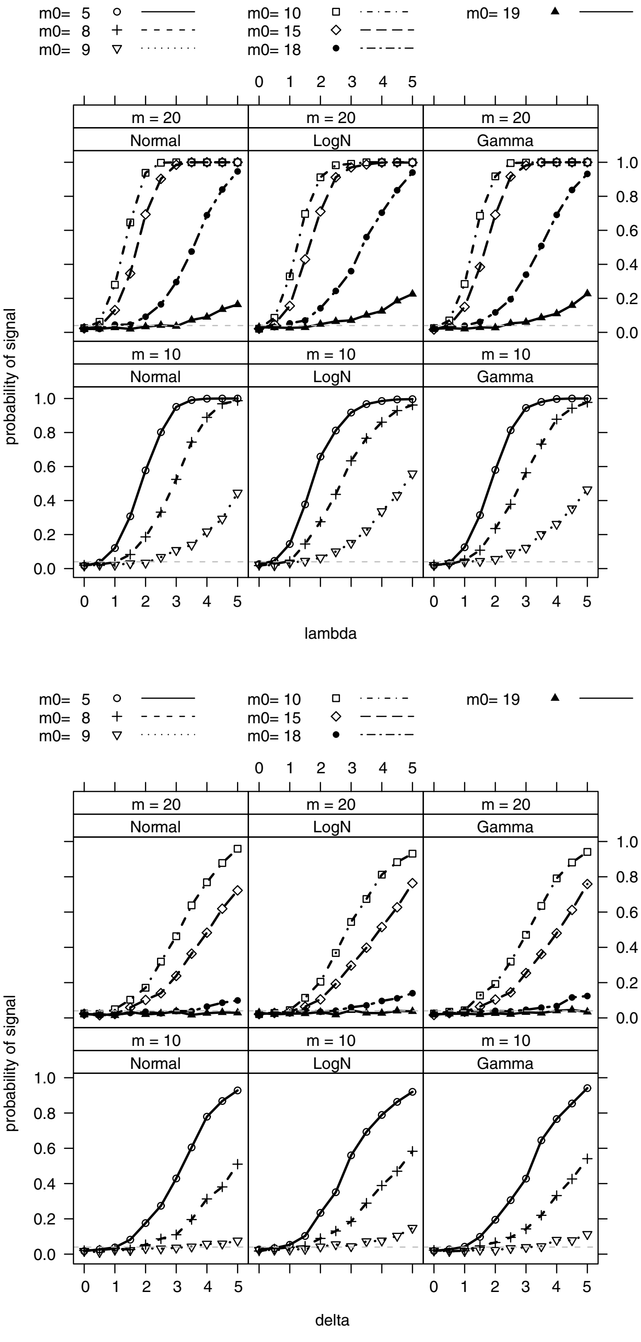


**Fig. 2.** Power of RPSM by a change in the intercept (top) and slope.

- In contrast to the RPSM, the LRT does not achieve the significance level for $m = 20$ and $p = 2$ in the case of Gamma or log-normally distributed error terms.
- The LRT was in Mahmoud et al. [13] only presented for the case $p = 2$, whereas our procedure can also be applied to larger values of $p$. RPSM needs, however, the same design matrix $\mathbf{X}$ for each profile.
- In the case of normally distributed data, we can compare LRT and RPSM: if the change point occurs in the middle of the profiles, the LRT performs slightly better than RPSM. Otherwise, if the change point is near the boundary of its domain, the LRT is superior to RPSM.
- RPSM can be implemented in statistical packages using standard procedures. Consequently, the implementation is quite fast.
- The RPSM leads to good (power) results for all the three distributions under study.

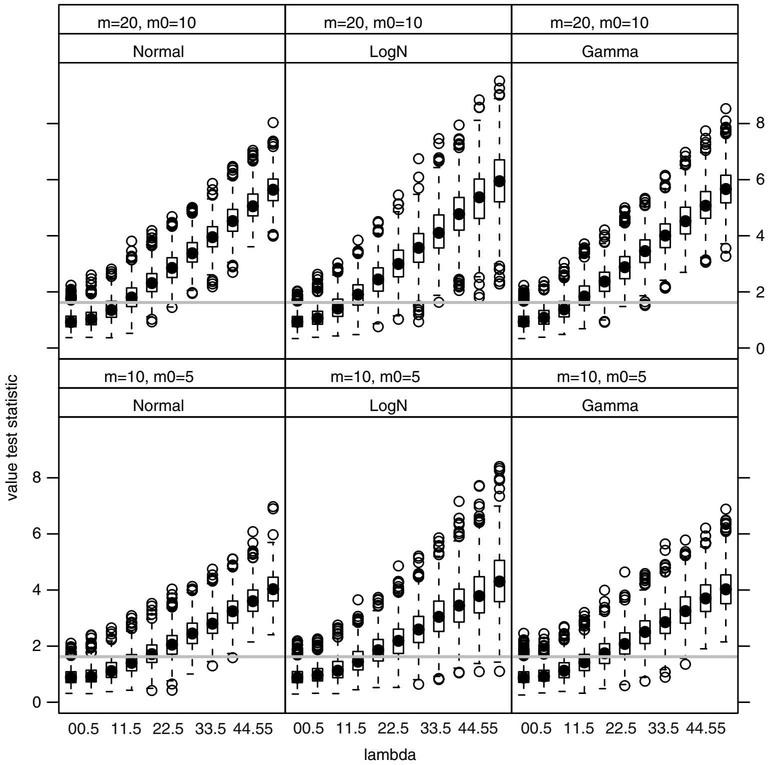


**Fig. 3.** Values of the test statistic (RPSM) by a change in the intercept.

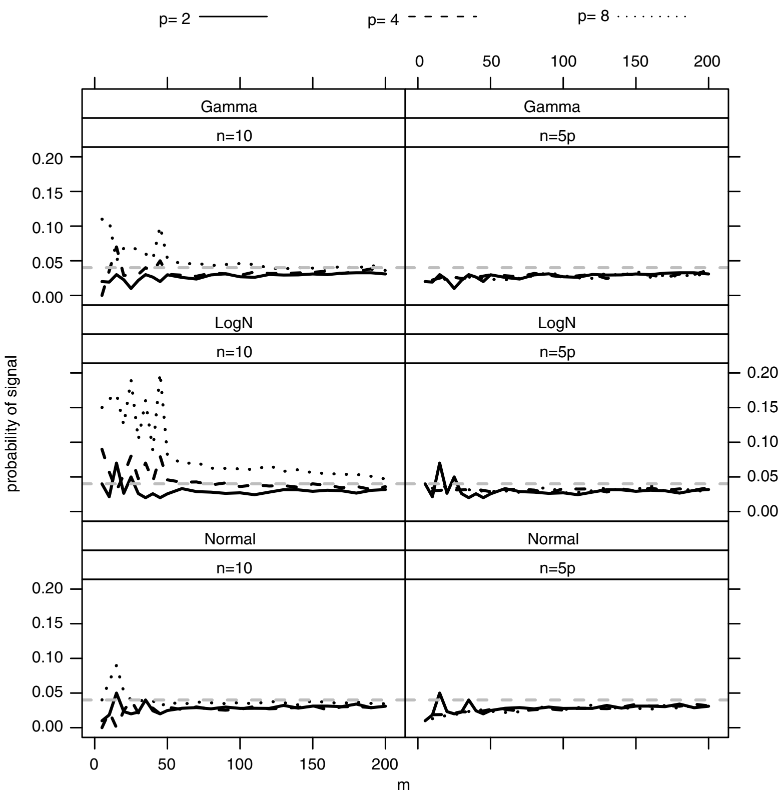


**Fig. 4.** Quality of the limit process (situation under $H_0$).

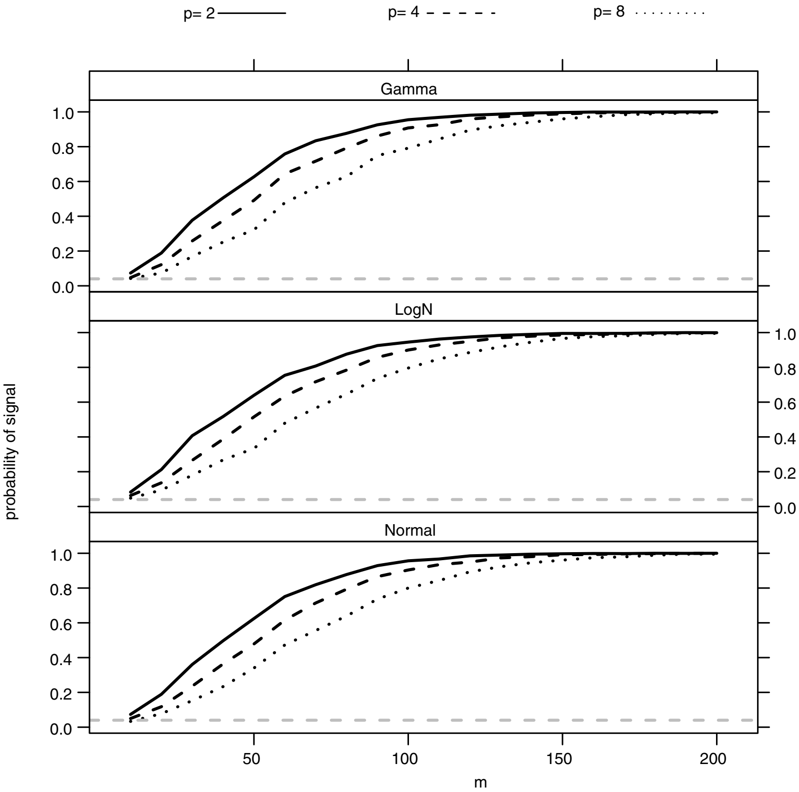


**Fig. 5.** Asymptotic power for $n = 5p$ and $m_0 = \lfloor \frac{m}{3} \rfloor$ by a small change in intercept ($\lambda = 1$).

- Even for small $m$ (i.e. $m = 10, 20$), the RPSM performs already well if the ratio of the number of parameters per profile $p$ to the number of observations per profile $n$ is not too large.

For the change point problem in the case of linear profile data with the same design matrix $\mathbf{X}$, we recommend the RPSM when the error distribution is unknown.

## 8. Proofs

In order to prove the results given in Section 3, we have to introduce some further notation.

Consider the space $W_m = [f_1(\tau_m), \ldots, f_d(\tau_m)]$ given in (8), which is a linear subspace of $\mathbb{R}^m$. With $W_{m,\mathcal{H}} := T_m(W_m)$ and

$$W^p_{m,\mathcal{H}} := W_{m,\mathcal{H}} \times \cdots \times W_{m,\mathcal{H}}, \quad \text{we have } W^p_{m,\mathcal{H}} = T_{m\times p}(W^p_m).$$

Furthermore $W^p_{m,\mathcal{H}}$ is a linear subspace of $T_{m\times p}(\mathbb{R}^{m\times p})$ and consequently also a subspace of the RKHS $\mathcal{H}^p$. Consider the following inner product on $T_{m\times p}(\mathbb{R}^{m\times p})$:

$$\langle T_{m\times p}(A), T_{m\times p}(B)\rangle_{m\times p} := m \cdot \langle A, B\rangle_{\mathbb{R}^{m\times p}} \quad \text{for } A, B \in \mathbb{R}^{m\times p},$$

which is the multivariate analogous to the univariate case considered by Bischoff [3]. Lemma 3.1 in [3] states that for $a \in \mathbb{R}^m$, we have

$$T_m(\mathrm{pr}_{W_m}(a)) = \mathrm{pr}_{T_m(W_m)}(T_m(a)).$$

Consequently, for a matrix $A \in \mathbb{R}^{m\times p}$,

$$T_{m\times p}(\mathrm{pr}_{W^p_m}(A))^\top = \mathrm{pr}_{T_{m\times p}(W^p_m)}(T_{m\times p}(A)^\top) = \mathrm{pr}_{W^p_{m,\mathcal{H}}}(T_{m\times p}(A)^\top) \tag{20}$$

holds true. With this, we get for the response variable $Y = g(\tau_m) + Z$ and for the associated residuals in Theorem 2

$$\begin{aligned} T_{m\times p}(\mathrm{pr}_{W^{p\perp}_m}(Y))^\top &= T_{m\times p}(Y)^\top - \mathrm{pr}_{T_{m\times p}(W^p_m)}\left(T_{m\times p}(Y)^\top\right) \\ &= \mathrm{pr}_{W^{p}_{m,\mathcal{H}}{}^{\perp}}\left(T_{m\times p}(Y)^\top\right). \end{aligned} \tag{21}$$

In case $u$ and $(u_m)_{m\in\mathbb{N}}$ are real continuous functions on [0, 1] with $u_m \longrightarrow u$ in $C([0,1])$, [3, p. 5] proved that

$$\mathrm{pr}_{W_{m,\mathcal{H}}}(u_m) \longrightarrow \mathrm{pr}_{W_{\mathcal{H}}}(u). \tag{22}$$

Since the corresponding projections $\mathrm{pr}_{W^p_{m,\mathcal{H}}}$ and $\mathrm{pr}_{W^p_{\mathcal{H}}}$ are defined component wise, the following lemma is an immediate consequence of (22).

**Lemma 3.** *Let $u, (u_m)_{m\in\mathbb{N}} \in C([0,1],\mathbb{R}^p)$ with $u_m$ converges to $u$ in $C([0,1],\mathbb{R}^p)$ for $m \to \infty$; then,*

$$\mathrm{pr}_{W^p_{m,\mathcal{H}}}(u_m) \longrightarrow \mathrm{pr}_{W^p_{\mathcal{H}}}(u) \quad \text{in } C([0,1],\mathbb{R}^p) \text{ for } m \to \infty.$$

Now we are in the position to prove our main result (Theorem 2):

**Proof.** We have $Y = g(\tau_m) + Z$ with $\mathbf{Cov}(\mathrm{vec}(Z^\top)) = \mathbf{I}_m \otimes \Sigma$. Let $c = (c_1,\ldots,c_p)^\top \in \mathbb{R}^p$ and $h = (h_1,\ldots,h_p)^\top \in W^p_{m,\mathcal{H}}$. Then, one gets

$$c^\top \mathrm{pr}_{W^p_{m,\mathcal{H}}}(h) = \sum_{j=1}^{p} c_j \mathrm{pr}_{W_{m,\mathcal{H}}}(h_j) = \mathrm{pr}_{W_{m,\mathcal{H}}}(c^\top h).$$

This leads to

$$A\mathrm{pr}_{W^p_{m,\mathcal{H}}}(h) = \mathrm{pr}_{W^p_{m,\mathcal{H}}}(Ah) \quad \text{for any } A \in \mathbb{R}^{p\times p}.$$

Consequently, we have under $H_0$ in connection with (21):

$$\begin{aligned}\frac{1}{\sqrt{m}}\Sigma^{-1/2}T_{m\times p}\left(\mathrm{pr}_{W_m^{p\perp}}(Y)\right)^\top &= \frac{1}{\sqrt{m}}\Sigma^{-1/2}T_{m\times p}\left(\mathrm{pr}_{W_m^{p\perp}}(Z)\right)^\top \\ &= \frac{1}{\sqrt{m}}\Sigma^{-1/2}\mathrm{pr}_{W^{p\ \perp}_{m,\mathcal{H}}}\left(T_{m\times p}(Z)^\top\right) \\ &= \mathrm{pr}_{W^{p\ \perp}_{m,\mathcal{H}}}\left(\frac{1}{\sqrt{m}}\Sigma^{-1/2}\,T_{m\times p}(Z)^\top\right).\end{aligned} \tag{23}$$

Since $Z$ fulfills the conditions of the vector-valued Donsker theorem (Theorem 1), we have

$$\frac{1}{\sqrt{m}}\Sigma^{-1/2}\,T_{m\times p}(Z)^\top \xrightarrow{\mathcal{D}} B^p \quad \text{for } m \to \infty.$$

Applying the "Continuous Mapping theorem" [1], one gets for $m \to \infty$ with Lemma 3, that the RPSP (23) converges weakly to $\mathrm{pr}_{W^{p\ \perp}_{\mathcal{H}}}(B^p)$. This proofs the theorem. □

## Acknowledgments

We would like to thank two unknown referees for their valuable suggestions which lead to an improvement in readability and to an extension of the simulation experiments.

## References

[1] P. Billingsley, Convergence of Probability Measures, 2nd ed., Wiley, New York, 1999.
[2] W. Bischoff, A functional central limit theorem for regression models, The Annals of Statistics 26 (4) (1998) 1398–1410.
[3] W. Bischoff, The structure of residual partial sums limit processes of linear regression models, The Theory of Stochastic Processes 8 (2002) 23–28.
[4] W. Bischoff, W. Somayasa, The limit of the partial sums process of spatial least squares residuals, Journal of Multivariate Analysis 100 (2009) 2167–2177.
[5] H. Dette, A. Munk, T. Wagner, Estimating the variance in nonparametric regression-what is a reasonable choice? Journal of the Royal Statistical Society. Series B (Statistical Methodology) 60 (4) (1998) 751–764.
[6] P. Hall, J.S. Marron, On variance estimation in nonparametric regression, Biometrika 77 (2) (1990) 415–419.
[7] D.A. Harville, Matrix Algebra from a Statistician's Perspective, Springer, New York, 1997.
[8] D. Iglehart, Weak convergence of probability measures on product spaces with applications to sums of random vectors, Tech. Rep., Dept. of Operations Res., Stanford Univ. 109, 1968.
[9] J. Kiefer, $K$-sample analogues of the Kolmogorov–Smirnov and Cramer-V. Mises tests, The Annals of Mathematical Statistics 30 (2) (1959) 420–447.
[10] K. Kim, M. Mahmoud, W. Woodall, On the monitoring of linear profiles, Journal of Quality Technology 35 (3) (2003) 317–328.
[11] I. MacNeill, Limit processes for sequences of partial sums of regression residuals, The Annals of Probability 6 (4) (1978) 695–698.
[12] I. MacNeill, Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times, The Annals of Statistics 6 (2) (1978) 422–433.
[13] M. Mahmoud, P. Parker, W. Woodall, D. Hawkins, A change point method for linear profile data, Quality and Reliability Engineering International 23 (2) (2007) 247–268.
[14] M. Mahmoud, W. Woodall, Phase I analysis of linear profiles with calibration applications, Technometrics 46 (4) (2004) 380–391.
[15] R Development Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2010, ISBN: 3-900051-07-0. URL: http://www.R-project.org.
[16] L. Xie, I. MacNeill, Spatial residual processes and boundary detection, South African Statistical Journal 40 (1) (2006) 33–53.