JOURNAL
OF
APPLIED
QUANTITATIVE
METHODS

A NON-PARAMETRIC TEST FOR A CHANGE-POINT IN LINEAR PROFILE DATA

Wolfgang BISCHOFF
Prof., Faculty of Mathematics and Geography,
Catholic University Eichstätt-Ingolstadt,
Eichstätt, Germany

E-mail: wolfgang.bischoff@ku-eichstaett.de

Andreas GEGG
PhD Candidate, Faculty of Mathematics and Geography,
Catholic University Eichstätt-Ingolstadt,
Eichstätt, Germany

E-mail: andreas.gegg@ku-eichstaett.de

Abstract: *We propose a change-point approach for testing the constancy of regression parameters in a linear profile data set (panel data in econometrics).*
Each sample collected over time in the historical data set consists of several multivariate observations for which a linear regression model is appropriate. The question now is whether all of the profiles follow a linear regression model with the same parameter vector or whether a change occurred in one or more model parameters after a special sample.
We use the partial sum operator in several dimensions to test the null hypothesis "H_0: no change-point occurred" and propose a non-parametric size α-test.
In Bischoff and Gegg (2010) we compared our proposed method with the likelihood-ratio-test by Mahmoud et al. (2007) in a simulation study. By these simulations we could show that our procedure can, in contrast to the likelihood-ratio-test, even be applied to the non-normal case. In this paper, however, we show how to compute our proposed test statistic step-by-step by considering an artificial data set.

Key words: *change-point problem; panel data; statistical process control; linear regression*

1. Introduction

We investigate a linear profile data set for change-points. In economics, linear profile data are also known as, "panel data". Note that linear profile data are assumed to be ordered in a natural way. In most of the applications the profiles will be sampled sequentially and so time is the ordering variable. We attack this problem by using the results given in Bischoff and Gegg (2010), where a linear regression model with p-variate response was considered. In order to find change-points in regression models we investigated there the partial sums of the least squares residuals. Without specifying the error a nonparametric 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. MacNeill (1978a,b) and Bischoff (1998, 2002) give basic theoretical results concerning the residual partial sums process for univariate response, whereas Bischoff (2010) demonstrates this approach in case of univariate regression by using an example from quality control. Note that the residual partial sums technique can also be used to check asymptotically for regression with multivariate correlated response (Bischoff and Gegg, 2010).

Our proposed method is a two-step-procedure: In a first step, we estimate the parameter vectors for every profile. In a second step we analyze these estimations which build a linear model with multiple correlated response under the null hypothesis that the profile data have no change-point.

Mahmoud et al. (2007) also attacked the described change-point problem and proposed a modification of a likelihood ratio test (LRT) for the case of simple linear regression with normally distributed error terms. By asymptotic considerations our method does not need assumptions about the distribution of the error terms and so it is more robust against departure from normal distribution, see Bischoff and Gegg (2010). A further advantage of our procedure is that the alternative hypothesis does not have to be specified.

2. Linear Profile Data

In practice, one often wants to test whether all of a fixed number m, say, of independent samples follow the same known linear model. To be more precise let

$$W^{(j)} = \mathbf{X}\,\beta^{(j)} + \varepsilon^{(j)}, \qquad \beta^{(j)} \in \mathbb{R}^p \text{ unknown}, \tag{1}$$

be a linear model for every profile $j \in \{1,\dots,m\}$, where

(A1) $\mathbf{X} \in \mathbb{R}^{n\times p}$ is the corresponding design matrix of explanatory variables with rank$(\mathbf{X}) = p \le n$,

(A2) and $\varepsilon^{(j)}$ is the vector with iid components $\varepsilon_1^{(j)},\dots,\varepsilon_n^{(j)}$ having mean 0 and variance σ^2.

Note that assumption **(A1)** in particular claims that the same design is used for each profile j. Furthermore, since different profiles are assumed to be independent, so are $\varepsilon^{(1)},\dots,\varepsilon^{(m)}$. In the sequel we assume model (1), together with the assumptions **(A1) - (A2)**, to be true for each j. Model (1) is called the "j-th Linear Profile" and the aim is to test for a change-point in the parameter vector. Since the profiles have a natural ordering we can formulate the corresponding hypothesis by

$$H_0 : \beta = \beta^{(1)} = ... = \beta^{(m)} \quad \text{vs.} \quad H_1 : \exists m_0 \in \{1,...,m-1\} : \beta^{(1)} = \beta^{(2)} = ... = \beta^{(m_0)} \neq \beta^{(m_0+1)} \qquad (2)$$

and so the testing problem is indeed a change-point problem. In order to check (2) 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

$$\mathbf{E}(\hat{\beta}^{(j)}) = \beta^{(j)} \quad \text{and} \quad \mathbf{Cov}(\hat{\beta}^{(j)}) = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} =: \Sigma. \qquad (3)$$

In case σ^2 is unknown, our proposed procedure can also be used by replacing σ^2 with a consistent estimator for σ^2 under H_0. So we can assume without loss of generality, Σ is a known positive definite matrix. Furthermore $\hat{\beta}^{(1)},...,\hat{\beta}^{(m)}$ are independent since the different samples $W^{(1)},...,W^{(m)}$ are assumed to be independent. Let $Y := \begin{pmatrix} \hat{\beta}^{(1)^T} \\ \vdots \\ \hat{\beta}^{(m)^T} \end{pmatrix}$ be the $m \times p$ matrix containing the least-squares estimations and let $Z := \begin{pmatrix} \hat{\beta}^{(1)^T} - \beta^T \\ \vdots \\ \hat{\beta}^{(m)^T} - \beta^T \end{pmatrix}$. If (1) and (2) hold true, then (3) leads to the following model:

$$Y = \mathbf{1}_m \beta^T + Z \quad \text{with} \quad \mathbf{EZ} = \mathbf{0},\ \mathbf{Cov}(\text{vec}(Z^T)) = \mathbf{I}_m \otimes \Sigma \quad \text{and } \beta \in \mathbb{R}^p \text{ unknown parameter vector.} \qquad (4)$$

Thereby $1_m \in \mathbb{R}^m$ is the vector whose components are all equal to 1, I_m is the $m \times m$ identity matrix, "$\otimes$" denotes the Kronecker-Product and "vec" is the well-known vec-operator (Harville, 1997).

Conversely, if (2) is false and (1) together with **(A1)–(A2)** still holds true, then a change-point occured and (4) does not hold. Therefore we can test hypothesis (2) by checking the linear model (4). Bischoff and Gegg (2010) formulated a procedure which can be used to check a more general model by using multiple partial sum processes. Below we apply this method to our problem.

3. Residual Partial Sums Process

In order to test the hypotheses (2), we investigate the partial sums of the p-dimensional residuals in model (4). For that we use the partial sum operator T_m, which embeds a vector $a = (a_1,...,a_m)^T \in \mathbb{R}^m$ in the space C([0,1]) by

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

where $\lfloor z \rfloor := \max\{l \in Z \mid l \leq z\}$ and $\sum_{i=1}^{0} a_i = 0$. Figure 1 shows the resulting graph of the partial sum operator T_m applied to a vector $a \in R^m$. The partial sum operator $T_{m \times p}$ embeds

$R^{m\times p} = R^m \times \ldots \times R^m$ in the space $C([0,1], R^p) = C([0,1]) \times \ldots \times C([0,1])$.

We define $T_{m\times p}$ with the help of T_m. For this, let A $\in$ $R^{m\times p}$ be an $m\times p$ matrix with columns $a^{(1)},\ldots,a^{(p)}$, then:

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

The partial sum operator has gained a lot of interest especially because of the well-known Donsker-Theorem for an iid sequence of centered random variables. Iglehart (1968) formulated a vector-valued version of this theorem:

Theorem 1

Let $(\xi_i)_{i\geq 1}$ *be an iid sequence of random variables with values in* R^p *and*

E $\xi_1 = \mathbf{0}$, ***Cov****(vec(Z*^{T}*))* = ***I***$_m \otimes \Sigma$

with Σ *positive definite. Then:*

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

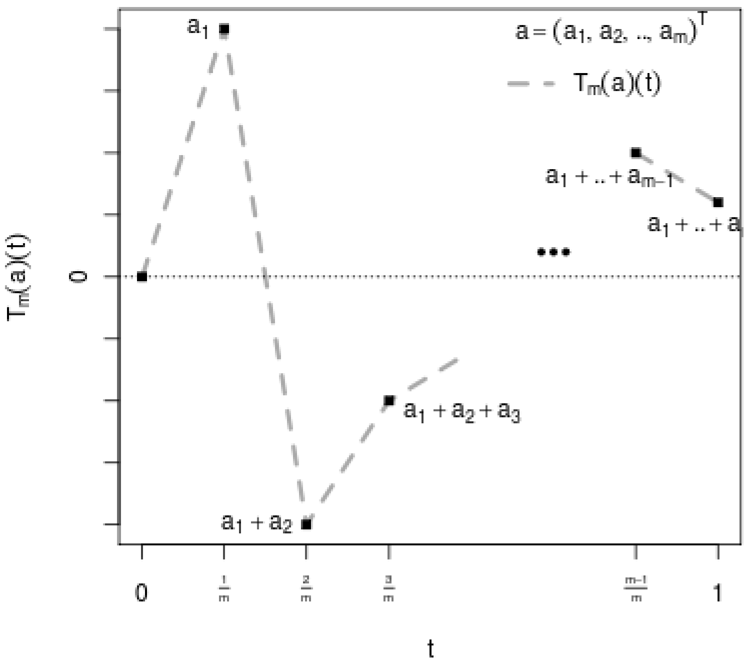

Figure1. Partial sum process $T_m(a)$

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

The residuals of the linear model (4) are correlated and through this they do not fulfil the iid assumption of the preceding theorem. However, Bischoff and Gegg (2010) used the vector-valued version of the Donsker-Theorem to establish the *p*-dimensional residual partial sums process in case of a multivariate linear model with multiple response. It is a projection of the Brownian motion B^{p} on a certain subspace. As a special case, we state the following result:

Theorem 2

Consider model (4), i.e. $Y = \mathbf{1}_m \beta^T + Z$ *with* $\mathbf{E}Z = \mathbf{0}$, $\mathbf{Cov}(vec(Z^T)) = \mathbf{I}_m \otimes \Sigma$ *and* $\beta \in R^p$ *unknown parameter vector.*

Then, under "$\mathrm{H}_0 : \beta = \beta^{(1)} = \ldots = \beta^{(m)}$*", we have for the residuals* $\mathrm{Y} - \hat{\mathrm{Y}}$ *and* $\mathrm{m} \to \infty$,

$$\frac{1}{\sqrt{\mathrm{m}}} \Sigma^{-1/2} \mathrm{T}_{\mathrm{m} \times \mathrm{p}} (\mathrm{Y} - \hat{\mathrm{Y}})^{\mathrm{T}} \xrightarrow{\mathrm{D}} \mathrm{B}_0^{\mathrm{p}}, \qquad (5)$$

where $\mathrm{B}_0^{\mathrm{p}}$ *is the p-dimensional Brownian bridge.*

4. Test for Linear Profile Data

Under the null hypothesis the residual partial sums limit process (cf. Theorem 2) is given by $\mathrm{B}_0^{\mathrm{p}}$, 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

$$\mathrm{R}_{\mathrm{m}}(\mathrm{t}) := \frac{1}{\sqrt{\mathrm{m}}} \Sigma^{-1/2} \mathrm{T}_{\mathrm{m} \times \mathrm{p}} (\mathrm{Y} - \hat{\mathrm{Y}})^{\mathrm{T}}(\mathrm{t}), \qquad \mathrm{t} \in [0,1].$$

Then our proposed test statistic is $\max_{t \in [0,1]} \| R_m(t) \|$, where $\| \cdot \|$ is the Euclidean norm in R^p, i.e. $\| (x_1, \ldots, x_p)^T \|^2 = \sum_{i=1}^{p} x_i^2$. Because of the "Continuous Mapping Theorem" (Billingsley 1999), we have the following convergence under H_0:

$$\| R_m \| \xrightarrow{D} \| B^p \| \qquad \text{for } \mathrm{m} \to \infty. \qquad (6)$$

Note that the limit process is the well-known Bessel bridge. In order to check (4) we apply a test of Kolmogorov-Smirnov type to the Bessel bridge and we get an asymptotic size α-test, $\alpha \in (0,1)$, by

Reject $H_0 \Leftrightarrow \sup_{t \in [0,1]} \| R_m(t) \| > k_\alpha$.

Thereby $\kappa_\alpha > 0$ is a constant such that $P\left(\sup_{t \in [0,1]} \| B_0^p(t) \| > k_\alpha\right) = \alpha$. Note that for given α, the corresponding value κ_α can be explicitly calculated. Kiefer (1959) gives a closed form for the cumulative distribution function of the Bessel bridge. We cite from his article concrete values for κ_α in case *p* = 2,...,5 and α = 0.1, 0.05, 0.01 in Table 1:

Table 1. Critical values for the *p*-dimensional Bessel bridge

κ_α	***p* = 2**	***p* = 3**	***p* = 4**	***p* = 5**
α=0.1	1.45399	1.61960	1.75593	1.87462
α=0.05	1.58379	1.74726	1.88226	2.00005
α=0.01	1.84273	2.00092	2.13257	2.24798

5. Numerical Example

5.1. Profile Data and Parameter Estimation

As a concrete example of application, we study the situation, that a change takes place after profile 4 of *m* = 6 profiles – both in the intercept and in the quadratic term in a quadratic model. For each profile, we simulated *n* = 10 observations according to

$$W_i^{(j)} = \alpha_1^{(j)} + x_i + \alpha_2^{(j)} x_i^2 + 2 \cdot \varepsilon_{ij}, i = 1,\ldots,10, j = 1,\ldots,6.$$

Thereby

- $\alpha_1^{(j)} = 0$ for j = 1,…,4 and $\alpha_1^{(j)} = 1$ for j = 5, 6 (shift in intercept)
- $\alpha_2^{(j)} = 0.1$ for j =1,…,4 and $\alpha_2^{(j)} = 0.12$ for j = 5, 6 (shift in quadratic term)
- $x_1 = 0, x_2 = \frac{10}{9}, x_3 = \frac{20}{9}, \ldots, x_{10} = 10,$
- ε_{ij} is a sequence of standardized iid random variables having lognormal distribution.

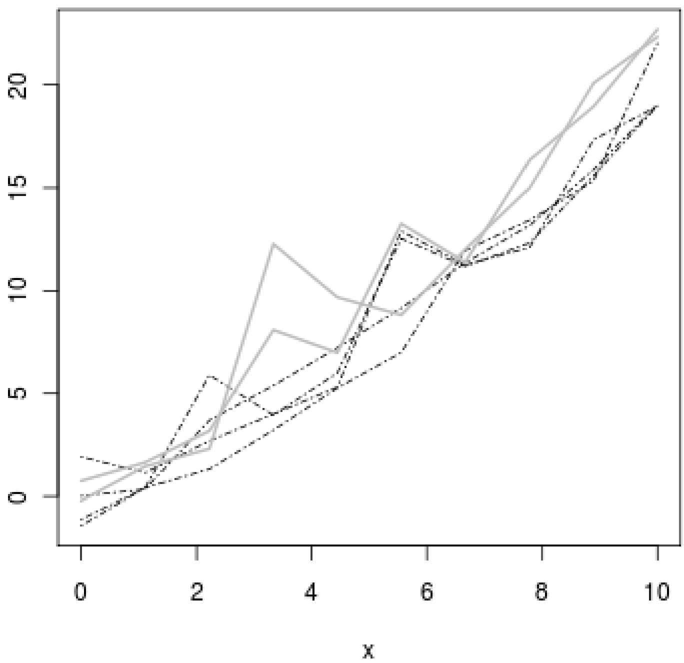

Figure 2. Simulated profiles $W^{(1)},\ldots,W^{(4)}$ (color in plot: black) and $W^{(5)},W^{(6)}$ (grey).

Figure 2 shows the simulated profile data under study. Let

$$\mathbf{X} := \begin{pmatrix} 1 & x_1 & x_1^2 \\ \vdots & & \\ 1 & x_{10} & x_{10}^2 \end{pmatrix} \in R^{m \times p}$$

be the design matrix for each profile. Consequently, we have $p = rank(\mathbf{X}) = 3$. We fit for each profile the model

$$W^{(j)} = \mathbf{X} \begin{pmatrix} \beta_0^{(j)} \\ \vdots \\ \beta_2^{(j)} \end{pmatrix} + \varepsilon^{(j)}, \tag{7}$$

where $\varepsilon^{(j)}$ is a random vector with $\mathbf{E}\,\varepsilon^{(j)} = 0$ and $\mathbf{Cov}\ \varepsilon^{(j)} = \sigma^2 \mathbf{I}_p$. We estimate the coefficients by least squares method and get the values $\beta^{(1)}, \ldots, \beta^{(6)}$ shown in Figure 3. With these estimations, we can fit model (4) with $\Sigma = \sigma^2 (XX^T)^{-1}$.

Furthermore, by model (7), we get for each profile j an estimation for the variance, namely the usual variance estimation

$$\hat{\sigma}_j^2 := \frac{1}{10-3} (W^{(j)^T} (\mathbf{I}_3 - \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T) W^{(j)}), j = 1, \ldots, 6.$$

Consequently, with our assumptions **(A1)-(A2)**, we can estimate σ^2 by

$$\hat{\sigma}^2 := \frac{1}{m} \sum_{j=1}^{m} \hat{\sigma}_j^2 = \frac{1}{6} \sum_{j=1}^{6} \hat{\sigma}_j^2.$$

In case of the simulated data mentioned above, we have $\hat{\sigma}^2 = 3.004228$.

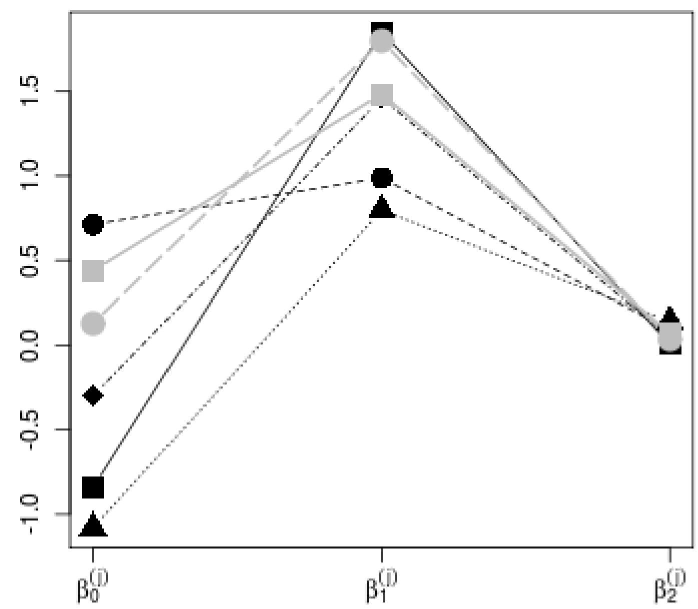

Figure 3. Estimations for parameter vector $\beta^{(1)}, \ldots, \beta^{(4)}$ (black) and $\beta^{(5)}, \beta^{(6)}$ (grey)

5.2. Test for a Change-Point

Now we are in the position, to calculate our proposed test statistic. Therefore, with $\mathrm{Y} := \begin{pmatrix} \hat{\beta}^{(1)^{\mathrm{T}}} \\ \vdots \\ \hat{\beta}^{(6)^{\mathrm{T}}} \end{pmatrix}$ and $\hat{Y} \in R^{6\times 3}$ being the matrix of estimated values in (4), we get:

$$\mathrm{R}_6(\mathrm{t}) := \frac{1}{\sqrt{6}\hat{\sigma}} (\mathbf{X}^{\mathrm{T}} \mathbf{X})^{1/2} \mathrm{T}_{6\times 3}(\mathrm{Y} - \hat{\mathrm{Y}})^{\mathrm{T}}(\mathrm{t}), \mathrm{t} \in [0,1].$$

Then we can determine the value of our test-statistic $\max_{t\in[0,1]} \| R_6(t) \|$ and compare with the critical values given in Table 1.

Figure 4 shows the process $\| R_6(t) \|$ for the data set under study. We get a value for the test statistic of 2.01976 and so we can reject the null hypothesis "H_0 no change-point" even for $\alpha = 0.01$ since the corresponding critical value is $\mathrm{k}_{0.01} = 2.00092.$

Note that, by using the same random numbers in case "no change-point" (i.e. $\alpha_1^{(j)} = 0, \alpha_2^{(j)} = 0.1$ for all *j* = 1,...,6), the test statistic is 0.5036005 and so we consistently cannot reject the null hypothesis to usual sizes.

Consequently, our proposed method leads to good results even in the case of small shifts (see Figure 2) and also in case of non-normal error terms (in the example, we have used log-normally distributed error terms).

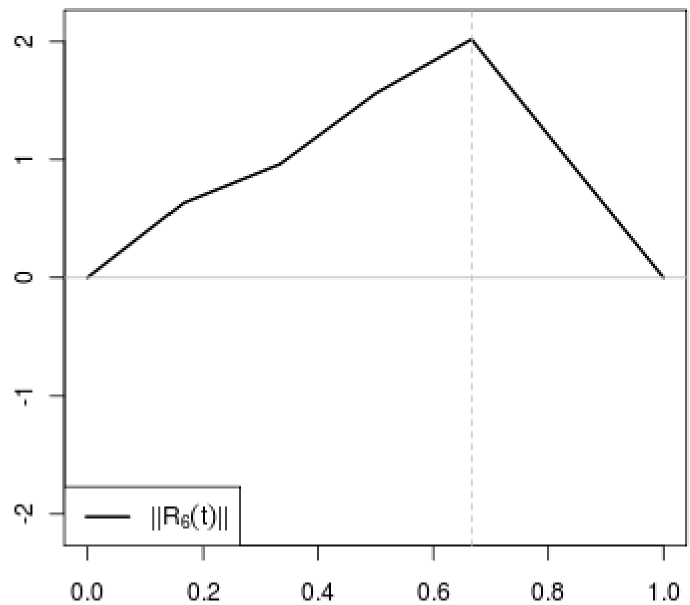

Figure 4. $\| R_6(t) \|$ with the true position of the change-point (dotted grey line)

References

1. Billingsley, P. **Convergence of Probability Measures** (2nd ed.), New York: Wiley, 1999
2. Bischoff, W. **A functional central limit theorem for regression models**, Ann. Stat., 26, 1998, pp. 1398-1410

3. Bischoff, W. **The structure of residual partial sums limit processes of linear regression models**, Theory Stoch. Process., 8, 24, N1-2, 2002, pp. 23-28
4. Bischoff, W. **Residual partial sums techniques to find change-points in linear regression**, Computer Modelling and New Technologies, 2010 (to appear)
5. Bischoff, W. and Gegg, A. **Partial sums process to check regression models with multiple correlated response with an application: Test for a change-point in profile data**, J. Multivariate Anal., 2010 (to appear)
6. Harville, D.A. **Matrix Algebra from a Statistician's Perspective**, New York: Springer, 1997
7. Iglehart, D. **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
8. Kiefer, J. **K-sample analogues of the Kolmogorov-Smirnov and Cramer-v. Mises tests**, Ann. Math. Stat., 30, 2, 1959, pp. 420-447
9. MacNeill, I.B. **Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times**, Ann. Statist., 6, 1978, pp. 422-433
10. MacNeill, I.B. **Limit processes for sequences of partial sums of regression residuals**, Ann. Prob., 6, 1978, pp. 695-698
11. Mahmoud, M., Parker,P., Woodall, W. and Hawkins, D. **A change point method for linear profile data**, Quality and Reliability Engineering International, 23, 2, 2007, pp. 247-268