Detecting relevant changes in time series models

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
e-mail: holger.dette@rub.de

Dominik Wied
Technische Universität Dortmund
Fakultät Statistik
44221 Dortmund, Germany
email: wied@statistik.tu-dortmund.de

March 2, 2015

Abstract

Most of the literature on change-point analysis by means of hypothesis testing considers hypotheses of the form $H_0 : \theta_1 = \theta_2$ vs. $H_1 : \theta_1 \neq \theta_2$, where θ_1 and θ_2 denote parameters of the process before and after a change point. This paper takes a different perspective and investigates the null hypotheses of *no relevant changes*, i.e. $H_0 : ||\theta_1 - \theta_2|| \leq \Delta$, where $\| \cdot \|$ is an appropriate norm. This formulation of the testing problem is motivated by the fact that in many applications a modification of the statistical analysis might not be necessary, if the difference between the parameters before and after the change-point is small. A general approach to problems of this type is developed which is based on the CUSUM principle. For the asymptotic analysis weak convergence of the sequential empirical process has to be established under the alternative of non-stationarity, and it is shown that the resulting test statistic is asymptotically normal distributed. The results can also be used to establish similarity of the parameters, i.e. $H_1 : ||\theta_1 - \theta_2|| \leq \Delta$, at a controlled type one error and to estimate the magnitude $||\theta_1 - \theta_2||$ of the change with a corresponding confidence interval. Several applications of the methodology are given including tests for relevant changes in the mean, variance, parameter in a linear regression model and distribution function among others. The finite sample properties of the new tests are investigated by means of a simulation study and illustrated by analyzing a data example from portfolio management.

Keywords: change-point analysis, CUSUM, relevant changes, precise hypotheses, strong mixing, weak convergence under the alternative

AMS Subject Classification: 62M10, 62F05, 62G10

1 Introduction

The analysis of structural breaks in a sequence $(Z_t)_{t=1}^n$ of random variables has a long history. Early work on this problem can be found in Page (1954, 1955) who investigated quality control problems. Since these seminal papers numerous authors have worked on the problem of detecting structural breaks or change-points in various statistical models [see Chow (1960), Brown et al. (1975), Krämer et al. (1988), among others]. Usually methodology is firstly developed for independent observations and – in a second step – extended to more complex dependent processes. Prominent examples of change-point analysis are the detection of instabilities in mean and variance [see Horváth et al. (1999) and Aue et al. (2009) among others]. These results have been extended to more complex regression models [see Andrews (1993) and Bai and Perron (1998)] and to change-point inference on the second order characteristics of a time series [see Berkes et al. (2009), Wied et al. (2012) and Preuss et al. (2014)]. A rather extensive list of references can be found in the recent work of Aue and Horváth (2013) who described how popular procedures investigated under the assumption of independent observations can be modified to analyse structural breaks in data exhibiting serial dependence.

A large portion of the literature attacks the problem of structural breaks by means of hypothesis testing instead of directly focusing on e.g. estimating the potential break points [compare the introduction in Jandhyala et al. (2013)]. Usually the hypothesis of no structural break is formulated as

$$H_0 : \theta_{(1)} = \theta_{(2)} = \cdots = \theta_{(n)} \tag{1.1}$$

where $\theta_{(t)}$ denotes a (not necessarily finite dimensional) parameter of the distribution of the random variable Z_t ($t = 1, \ldots, n$), such as the mean, variance, etc. The alternative is then formulated (in the simplest case of one structural break) as

$$H_1 : \theta_1 = \theta_{(1)} = \theta_{(2)} = \cdots = \theta_{(k)} \neq \theta_{(k+1)} = \theta_{(k+2)} = \cdots = \theta_{(n)} = \theta_2, \tag{1.2}$$

where $k \in \{1, \ldots, n\}$ denotes the (unknown) location of the change-point. If the null hypothesis of structural breaks has been rejected, the location of the change has to be estimated [see Csörgő and Horváth (1997) or Bai and Perron (1998) among others] and the statistical analysis has to be modified to address the different stochastic properties before and after the change-point.

The present work is motivated by the observation that such a modification of the statistical analysis might not be necessary if the difference between the parameters before and after the change-point is rather small. For example, in risk management situations, one is interested in fitting a suitable model for forecasting Value at Risk from "uncontaminated data", that means from data after the last change-point [see e.g. Wied (2013)]. But in practice, small changes in the parameter are perhaps not very interesting because they do not yield large changes in the Value at Risk. The forecasting quality might only improve slightly, but this benefit could be negatively overcompensated by transaction costs. On the other hand, as an illustration with real interest rates at the end of this paper indicates, a relevant difference can potentially be linked to significant real-world events. One could also think of an application to inflation rates in the sense that only “large” changes call for interventions of, for example, the European Central Bank. With this point of view it might be more reasonable to replace the hypothesis (1.2) by the null hypothesis of *no relevant structural break*, that is

$$H_0 : \|\theta_1 - \theta_2\| \leq \Delta \quad \text{versus} \quad H_1 : \|\theta_1 - \theta_2\| > \Delta, \tag{1.3}$$

where θ_1 and θ_2 are the parameters before and after the change-point, $\|\cdot\|$ denotes a (semi-)norm on the parameter space and Δ is a pre-specified constant representing the “maximal” change accepted by statisticians without modifying the statistical analysis. Note that this formulation of the change-point problem avoids the consistency problem as mentioned in Berkson (1938), that is: any consistent test will detect any arbitrary small change in the parameters if the sample size is sufficiently large. Moreover, the “classical” formulation of the change-point problem in formula (1.1) does not allow to control the type II error if the null hypothesis of no structural break cannot be rejected, and as a consequence the statistical uncertainty in the subsequent data analysis (under the assumption of stationarity) cannot be quantified. On the other hand, a decision of “no small structural” break at a controlled type I error can be easily achieved by interchanging the null hypothesis and alternative in (1.3), that is

$$H_0 : \|\theta_1 - \theta_2\| > \Delta \quad \text{versus} \quad H_1 : \|\theta_1 - \theta_2\| \leq \Delta. \tag{1.4}$$

The new approach requires the specification of the quantity $\Delta > 0$, which depends on the specific application. ”Classical” hypotheses tests simply use $\Delta = 0$, but we argue that from a practical point of view it might be very reasonable to think about this choice more carefully and to define the size of change in which one is really interested. The relevance of testing hypotheses of the form (1.3), which are also called *precise hypotheses* in the literature [see Berger and Delampady (1987)], has nowadays been widely recognized in various fields of statistical inference including medical, pharmaceutical, chemistry or environmental statistics [see Chow and Liu (1992), McBride (1999)]. On the other hand – to our best knowledge – the problem of testing for relevant structural breaks has not been discussed in the literature so far.

In this paper we present a general approach to address this problem, which is based on the CUSUM principle. The basic ideas are illustrated in Section 2 for the problem of detecting a relevant change in the mean of a multivariate sequence of independent observations. The general methodology is introduced in Section 3 and is applicable to several other situations including changes in the variance, the parameter in regression models and changes in the distribution function (the nonparametric change-point problem). Additionally, if it is difficult to specify the threshold Δ, it can be used to estimate the magnitude $||\theta_1 - \theta_2||$ with a corresponding confidence interval.

It turns out that - in contrast to the classical change-point problem - testing relevant hypotheses of the type (1.3) requires results on the weak convergence of the sequential empirical process under non-stationarity (more precisely under the alternative H_1), which - to our best knowledge - have not been developed so far. The reference which is most similar in spirit to investigations of this type is Zhou (2013), who considered the asymptotic properties of tests for the classical hypothesis of a change in the mean, i.e. $H_0 : \mu_1 = \mu_2$, under piecewise local stationarity. The present paper takes a different and more general perspective using weak convergence of the sequential empirical process in the case $\theta_1 \neq \theta_2$. These asymptotic properties depend sensitively on the dependence structure of the basic time series $(Z_t)_{t\in\mathbb{Z}}$ and are developed in Section A.1 of the online supplement for the concept of strong mixing triangular arrays. Although the analysis of the sequential process under non-stationarities of the type (1.2) is very complicated, the resulting test statistics for the hypothesis of *no relevant structural break* have a very simple asymptotic distribution, namely a normal distribution. Consequently, statistical analysis can be performed estimating a variance and using quantiles of the standard normal distribution. In Section 4 we illustrate the methodology and develop tests for the hypothesis (1.3) of a relevant change in the mean, parameters in a linear regression model and distribution function. In particular, we consider the situation of testing for a change in the mean with possibly simultaneously changing variance, which occurs frequently in applications. Note that none of the classical change-point tests are able to address this problem. In fact it was pointed out by Zhou (2013) that the classical CUSUM approach and similar methods are not pivotal in this case leading to severe biased testing results. Section 5 presents some finite sample evidence of the new test revealing appealing size and power properties. We also give an illustration in a data example from portfolio management. In an Online Appendix we provide some theoretical results, which demonstrate that the assumptions made in Section 3 are satisfied for strong mixing processes, give some of the more technical proofs and discuss the problem of detecting relevant changes in the variance and correlation.

2 Relevant Changes in the Mean - Motivation

This section serves as a motivation for the general approach to detect relevant changes in time series which will be discussed in Section 3. For illustration purposes we consider independent d-dimensional random variables $Z_1, ..., Z_n$ with common positive definite variance $\mathrm{Var}(Z_i) = \Sigma$, such that for some unknown $t \in (0, 1)$

$$\mu_1 = \mathbb{E}[Z_1] = \ldots = \mathbb{E}[Z_{\lfloor nt\rfloor}] \ ; \ \ \mathbb{E}[Z_{\lfloor nt\rfloor+1}] = \ldots = \mathbb{E}[Z_n] = \mu_2.$$

The case of a variance simultaneously changing with the mean will be discussed in Section 4. We are interested in the problem of testing for a relevant change in the mean, that is

$$H_0 : \|\mu_1 - \mu_2\| \leq \Delta \qquad \text{versus} \qquad H_1 : \|\mu_1 - \mu_2\| > \Delta, \tag{2.1}$$

where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^d$. For this purpose we consider the CUSUM statistic $\{\mathbb{U}_n(s)\}_{s\in[0,1]}$ defined by

$$\hat{\mathbb{U}}_n(s) = \frac{1}{n}\sum_{j=1}^{\lfloor ns\rfloor} Z_j - \frac{s}{n}\sum_{j=1}^{n} Z_j = \frac{1-s}{n}\sum_{j=1}^{\lfloor ns\rfloor} Z_j - \frac{s}{n}\sum_{j=\lfloor ns\rfloor+1}^{n} Z_j$$

Note that a straightforward computation gives $\mathbb{E}[\hat{\mathbb{U}}_n(s)] = (s \wedge t - st)(\mu_1 - \mu_2)(1 + o(1))$. A similar calculation yields

$$\mathbb{E}[\|\hat{\mathbb{U}}_n(s)\|^2] = \Big\{\frac{\sigma^2}{n}s(1-s) + \|\mu_1 - \mu_2\|^2(s \wedge t - st)^2\Big\}\ (1 + o(1)).$$

Consequently, we obtain

$$\begin{aligned}(2.2)\qquad \mathbb{E}\Big[\int_0^1 \|\hat{\mathbb{U}}_n(s)\|^2 ds\Big] &= \Big\{\int_0^1 \frac{\sigma^2}{n}s(1-s) + \|\mu_1 - \mu_2\|^2(s \wedge t - st)^2 ds\Big\}(1 + o(1)) \\ &= \Big\{\frac{\sigma^2}{6n} + \|\mu_1 - \mu_2\|^2\frac{(t(1-t))^2}{3}\Big\}(1 + o(1)),\end{aligned}$$

and therefore it is reasonable to consider the statistic

$$\frac{3}{(t(1-t))^2}\int_0^1 \|\hat{\mathbb{U}}_n(s)\|^2 ds$$

as an estimator of the distance $\|\mu_1 - \mu_2\|^2$ (a bias correction addressing for the term $\sigma^2/(6n)$ will be discussed later). The following result specifies the asymptotic properties of this statistic. Throughout this paper the symbol $\stackrel{\mathcal{D}}{\Longrightarrow}$ means weak convergence in the appropriate space under consideration.

Theorem 2.1 *For any $t \in (0, 1)$ we have as $n \to \infty$*

$$(2.3)\qquad L_n = \sqrt{n}\Big(\frac{3}{(t(1-t))^2}\int_0^1 \|\hat{\mathbb{U}}_n(s)\|^2 ds \ - \ \|\mu_1 - \mu_2\|^2\Big) \stackrel{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0, \tau^2)\ ,$$

where the asymptotic variance is given by

$$(2.4)\qquad \tau^2 = (\mu_1 - \mu_2)^T\Sigma(\mu_1 - \mu_2)\frac{4(1 + 2t(1-t))}{5t^2(1-t)^2}.$$

Proof. We start calculating the covariance $\mathrm{Cov}(\hat{\mathbb{U}}_n(s_1), \hat{\mathbb{U}}_n(s_2))$ using the decomposition

$$\hat{\mathbb{U}}_n(s) = (1-s)\mathbb{U}_n^{(1)}(s) - s\mathbb{U}_n^{(2)}(s),$$

where

$$\mathbb{U}_n^{(1)}(s) = \frac{1}{n}\sum_{j=1}^{\lfloor ns\rfloor} Z_j\ ;\quad \mathbb{U}_n^{(2)}(s) = \frac{1}{n}\sum_{j=\lfloor ns\rfloor+1}^{n} Z_j.$$

For this purpose we first assume that $s_1 \le s_2$ and note that $\mathrm{Cov}(\mathbb{U}_n^{(1)}(s_1), \mathbb{U}_n^{(2)}(s_2)) = 0$ in this case. Moreover, the remaining covariances are obtained as follows

$$\begin{aligned}\mathrm{Cov}(\mathbb{U}_n^{(1)}(s_1), \mathbb{U}_n^{(1)}(s_2)) &= \frac{1}{n^2}\sum_{j,k=1}^{\lfloor ns_1\rfloor}\mathrm{Cov}(Z_j, Z_k) = \frac{s_1}{n}\Sigma\ (1 + o(1)), \\ \mathrm{Cov}(\mathbb{U}_n^{(1)}(s_2), \mathbb{U}_n^{(2)}(s_1)) &= \frac{1}{n^2}\sum_{j=1}^{\lfloor ns_2\rfloor}\sum_{k=\lfloor ns_1\rfloor+1}^{n}\mathrm{Cov}(Z_j, Z_k) = \frac{s_2 - s_1}{n}\Sigma\ (1 + o(1)), \\ \mathrm{Cov}(\mathbb{U}_n^{(2)}(s_1), \mathbb{U}_n^{(2)}(s_2)) &= \frac{1}{n^2}\sum_{j=\lfloor ns_1\rfloor+1}^{n}\sum_{s=\lfloor ns_2\rfloor+1}^{n}\mathrm{Cov}(Z_j, Z_k) = \frac{1 - s_2}{n}\Sigma\ (1 + o(1)),\end{aligned}$$

which gives $\mathrm{Cov}(\hat{\mathbb{U}}_n(s_1), \hat{\mathbb{U}}_n(s_2)) = \frac{s_1(1-s_2)}{n}\Sigma\ (1+o(1))$ if $s_1 \leq s_2$. A similar calculation for the case $s_1 \geq s_2$ finally yields

$$\lim_{n\to\infty} n\ \mathrm{Cov}(\hat{\mathbb{U}}_n(s_1), \hat{\mathbb{U}}_n(s_2)) = (s_1 \wedge s_2 - s_1 s_2)\Sigma.$$

It can be shown (note that for illustration purposes the random variables $Z_1, \ldots, Z_n$ are assumed to be independent and a corresponding statement under the assumption of a strong mixing process is given in Section A.1) that an appropriately standardized version of the process $\hat{\mathbb{U}}_n$ converges weakly, that is

$$\{\sqrt{n}(\hat{\mathbb{U}}_n(s) - \mu(s,t))\}_{s\in[0,1]} \stackrel{\mathcal{D}}{\Longrightarrow} \Sigma^{1/2}\{B(s)\}_{s\in[0,1]},$$

where $\mu(s,t) = (s \wedge t - st)(\mu_1 - \mu_2)$ and B denotes a vector of independent Brownian bridges on the interval $[0,1]$. This gives for the random variable L_n in (2.3)

$$\begin{aligned} L_n &= \frac{3\sqrt{n}}{(t(1-t))^2}\Big\{\int_0^1 \|\hat{\mathbb{U}}_n(s)\|^2 ds - \|\mu_1 - \mu_2\|^2 \frac{(t(1-t))^2}{3}\Big\} \\ &= \frac{3\sqrt{n}}{(t(1-t))^2}\Big\{\int_0^1 (\|\hat{\mathbb{U}}_n(s)\|^2 - \|\mu(s,t)\|^2) ds\Big\} \\ &= \frac{3\sqrt{n}}{(t(1-t))^2}\Big\{\int_0^1 \|\hat{\mathbb{U}}_n(s) - \mu(s,t))\|^2 ds\ +\ 2\int_0^1 \mu^T(s,t)\{\hat{\mathbb{U}}_n(s) - \mu(s,t)\} ds\Big\} \\ &\stackrel{\mathcal{D}}{\Longrightarrow} \frac{6}{(t(1-t))^2}\int_0^1 \mu^T(s,t)\Sigma^{1/2}B(s)ds. \end{aligned}$$

It is well known that the distribution on the right hand side is a centered normal distribution with variance.

$$\frac{36}{(t(1-t))^4}\int_0^1\int_0^1 \mu^T(s_1,t)\Sigma\mu(s_2,t)(s_1 \wedge s_2 - s_1 s_2) ds_1 ds_2,$$

and it follows by a straightforward calculation that this expression is given in (2.4). □

The test statistic for the hypothesis (2.1) is finally defined as

$$\hat{\mathbb{M}}_n^2 = \frac{3}{(\hat{t}(1-\hat{t}))^2}\int_0^1 \|\hat{\mathbb{U}}_n(s)\|^2 ds - \frac{\hat{\sigma}^2}{6n}$$

where $\hat{t}$ and $\hat{\sigma}^2$ are consistent estimators of t and σ^2, respectively. Note that this definition corrects for the additional bias in (2.2) which is asymptotically negligible. The null hypothesis of no relevant change-point is finally rejected, whenever

$$\hat{\mathbb{M}}_n^2 \geq \Delta^2 + u_{1-\alpha}\frac{\hat{\tau}}{\sqrt{n}}, \tag{2.5}$$

where $u_{1-\alpha}$ is the $(1-\alpha)$-quantile of the standard normal distribution and $\hat{\tau}$ is an appropriate estimator of τ. An estimator of the change-point can be obtained by the argmax-principle, that is $\hat{t} = \text{argmax}_{s\in[0,1]} \|\hat{\mathbb{U}}_n(s)\|$ [see Carlstein (1988)]. For the estimation of the residual variance we denote by

$$(2.6) \qquad \hat{\mu}_1 = \frac{1}{\lfloor n\hat{t} \rfloor} \sum_{i=1}^{\lfloor n\hat{t} \rfloor} Z_i \ ; \qquad \hat{\mu}_2 = \frac{1}{\lfloor (1-\hat{t})n \rfloor} \sum_{i=\lfloor n\hat{t} \rfloor+1}^{n} Z_i$$

the estimates of the mean "before" and "after" the change-point and define a variance estimator by

$$(2.7) \qquad \hat{\Sigma}_1 = \frac{1}{n} \Big\{ \sum_{i=1}^{\lfloor n\hat{t} \rfloor} (Z_i - \hat{\mu}_1)(Z_i - \hat{\mu}_1)^T + \sum_{i=\lfloor n\hat{t} \rfloor+1}^{n} (Z_i - \hat{\mu}_2)(Z_i - \hat{\mu}_2)^T \Big\}.$$

This yields

$$(2.8) \qquad \hat{\tau} = \frac{2\hat{\nu}}{\sqrt{5}\hat{t}(1-\hat{t})} \sqrt{1 + 2\hat{t}(1-\hat{t})}$$

as an estimation of τ, where $\hat{\nu}^2 = (\hat{\mu}_1 - \hat{\mu}_2)^T \hat{\Sigma}_1 (\hat{\mu}_1 - \hat{\mu}_2)$. It will be shown in Section 3 that the test defined by (2.5) is consistent and has asymptotic level α.

Remark 2.1 Our motivation for considering a statistic based on the integral $\int_0^1 \|\hat{\mathbb{U}}_n^2(s)\|^2 ds$ is twofold. On the one hand, we want to consider a CUSUM-type statistic, which has some optimality properties for testing the "classical" change-point hypothesis $H_0 : \mu_1 = \mu_2$ [see e.g. Lorden (1971) and the subsequent literature for early references and Moustakides (2004) for a more recent reference]. On the other hand, we are interested in a test statistic with a simple limit distribution, such as a normal distribution. This allows us to use classical results on UMPU-tests for interval hypotheses in exponential families [see Lehmann (1986), Chapter 7] to derive powerful tests for the hypothesis of a relevant change point.
As pointed out by a referee, an alternative test could be based on the statistic

$$\hat{\Theta}_n = \max_{0<s<1} \Big\| \frac{1}{\lfloor ns \rfloor} \sum_{j=1}^{\lfloor ns \rfloor} Z_j - \frac{1}{n - \lfloor ns \rfloor} \sum_{j=\lfloor ns \rfloor+1}^{n} Z_j \Big\|,$$

which estimates the jump size directly. The null hypothesis is rejected for large values of $\hat{\Theta}_n$. However, the asymptotic distribution of the statistic $\hat{\Theta}_n$ (appropriately standardized) is not known in the case $\mu_1 \neq \mu_2$ and, as a consequence, the classical UMPU-test theory for interval hypotheses is not applicable here.

3 Testing for Relevant Changes - A General Approach

3.1 ***General Formulation of the Problem***: Let $Z_1, \ldots, Z_n$ denote d-dimensional random variables such that

$$(3.1) \qquad Z_1, \ldots, Z_{\lfloor nt \rfloor} \sim F_1, \qquad Z_{\lfloor nt \rfloor+1}, \ldots, Z_n \sim F_2,$$

where F_1 and F_2 denote continuous distribution functions before and after the change-point. Let $\mathcal{S}$ be a Hilbert space with (semi-)norm $||\cdot||$, define $\ell^\infty(\mathbb{R}^d|\mathcal{S})$ as the set of all bounded functions $g : \mathbb{R}^d \to \mathcal{S}$ and consider $\mathcal{F} \subset \ell^\infty(\mathbb{R}^d|\mathcal{S})$. We denote by

$$\theta : \mathcal{F} \to \mathcal{S} \; ; \; F \to \theta(F) \tag{3.2}$$

a given function defining the parameter of interest. Typical examples include the mean ($\theta(F) = \int z dF$) or the distribution function (here θ is the identity map). We are interested in testing the hypothesis of *no relevant change* in the functional $\theta(F)$, that is

$$H_0 : \|\theta(F_1) - \theta(F_2)\| \leq \Delta \qquad\qquad H_1 : \|\theta(F_1) - \theta(F_2)\| > \Delta, \tag{3.3}$$

where $\Delta > 0$ is a pre-specified constant. If $\mathcal{S} \subset \mathbb{R}^k$ with $k \leq d$, then $\|\cdot\|$ denotes always the Euclidean norm, if not specified otherwise.
Our general approach will be based on an estimator of the distance $\|\theta(F_1)-\theta(F_2)\|^2$ by a CUSUM type statistic. For this purpose we assume for a moment linearity of the functional θ in (3.2), that is

$$\theta(\alpha F_1 + \beta F_2) = \alpha\theta(F_1) + \beta\theta(F_2) \tag{3.4}$$

for all $\alpha, \beta \in \mathbb{R},\ F_1, F_2 \in \mathcal{F}$. We introduce the statistic

$$\hat{\mathbb{F}}_n(s, z) = \frac{1}{n} \sum_{j=1}^{\lfloor ns \rfloor} I\{Z_j \leq z\}, \tag{3.5}$$

where $s \in [0, 1]$, $z \in \mathbb{R}^d$ and the inequality is understood component-wise. Note that for fixed $s \in (0, 1]$ the function $\frac{n}{\lfloor ns \rfloor}\hat{\mathbb{F}}(s, \cdot)$ is a distribution function and that a straightforward calculation yields

$$\lim_{n\to\infty} \mathbb{E}[\hat{\mathbb{F}}_n(s, z)] = E_{F_1,F_2,t}(s, z) := (s \wedge t)F_1(z) + (s - t)_+ \; F_2(z). \tag{3.6}$$

We also introduce the function

$$Z_{F_1,F_2,t}(s, z) := E_{F_1,F_2,t}(s, z) - sE_{F_1,F_2,t}(1, z) = (s \wedge t - st)(F_1(z) - F_2(z)) \tag{3.7}$$

and note that $Z_{F_1,F_2,t}$ vanishes on $[0, 1] \times \mathbb{R}^d$ if and only if $F_1 = F_2$. If $\Phi_{\text{lin}} : \ell^\infty([0, 1] \times \mathbb{R}^d|\mathbb{R}) \to \ell^\infty([0, 1]|\mathcal{S})$ denotes the (linear) operator defined by

$$\Phi_{\text{lin}}(E_{F_1,F_2,t})(s) := \theta(E_{F_1,F_2,t}(s, \cdot) - sE_{F_1,F_2,t}(1, \cdot)) = \theta(Z_{F_1,F_2,t})(s),$$

we obtain from (3.4) and (3.7) for the function $\mathbb{U} := \Phi_{\text{lin}}(E_{F_1,F_2,t})$ the representation

$$\mathbb{U}(s) := \Phi_{\text{lin}}(E_{F_1,F_2,t})(s) = (s \wedge t - st)(\theta(F_1) - \theta(F_2)). \tag{3.8}$$

Consequently, the norm of this function is given by

$$\mathbb{T}^2(s) = \|\mathbb{U}(s)\|^2 = \|\theta(Z_{F_1,F_2}(s,\cdot)\|^2 = (s \wedge t - st)^2 \|\theta(F_1) - \theta(F_2)\|^2, \tag{3.9}$$

which can be used as the basis for estimating the distance between the parameters $\theta(F_1)$ and $\theta(F_2)$. Before we explain the construction of this estimate in more detail, we "remove" assumption (3.4) and consider more general nonlinear functionals.

In this case the situation is slightly more complicated and we assume throughout this paper that there exists a mapping

$$\Phi : \ell^\infty([0,1] \times \mathbb{R}^d|\mathbb{R}) \to \ell^\infty([0,1]|\mathcal{S}), \tag{3.10}$$

such that the difference between $\theta(F_1)$ and $\theta(F_2)$ can be expressed as a functional of the function $E_{F_1,F_2,t}$ in (3.6), that is

$$\mathbb{U}(s) := \Phi(E_{F_1,F_2,t})(s) = (s \wedge t - st)(\theta(F_1) - \theta(F_2)). \tag{3.11}$$

For linear functionals such a representation is obvious as shown in the preceding paragraph. Other examples where assumption (3.11) is satisfied include linear regression models or the detection of relevant changes in the correlation and will be discussed in Section 4 and in Section C.2 of the online supplement.

For the construction of an estimate of $\|\theta(F_1) - \theta(F_2)\|^2$ we note that it follows by similar arguments as given in Section 2 that the function $\mathbb{T}(s) = \|\mathbb{U}(s)\|$ satisfies

$$\int_0^1 \mathbb{T}^2(s)ds = \int_0^1 (s \wedge t - st)^2 \|\theta(F_1) - \theta(F_2)\|^2 ds = \frac{(t(1-t))^2}{3} \|\theta(F_1) - \theta(F_2)\|^2. \tag{3.12}$$

Observing (3.9) and (3.12) we see that the distance

$$M^2 = M^2(F_1, F_2) = \|\theta(F_1) - \theta(F_2)\|^2 = \frac{3}{(t(1-t))^2} \int_0^1 \|\Phi(E_{F_1,F_2,t}(s))\|^2 ds \tag{3.13}$$

between the parameters $\theta(F_1)$ and $\theta(F_2)$ can be expressed as a functional of $E_{F_1,F_2,t}(\cdot,\cdot)$, which can easily be estimated by a sequential empirical process $\hat{\mathbb{F}}_n$ defined in (3.5). The null hypothesis (3.3) is then rejected for large values of this estimator. In the following discussion we will derive the asymptotic properties of this estimator, which can be used for the calculation of critical values for a test of the null hypothesis (3.3) of *no relevant change*.

3.2 ***Estimating the Distance*** $M(F_1, F_2) = \|\theta(F_1) - \theta(F_2)\|$**:** In order to estimate the distance $M^2(F_1, F_2) = \|\theta(F_1) - \theta(F_2)\|^2$ we recall the definition of the sequential empirical process in (3.5) and its asymptotic expectation $E_{F_1,F_2,t}$ defined in (3.6). Observing assumption (3.11) we consider the processes

$$\hat{\mathbb{U}}_n(s) = \Phi(\hat{\mathbb{F}}_n(s,\cdot)), \tag{3.14}$$

$$\hat{\mathbb{T}}_n^2(s) = \|\hat{\mathbb{U}}_n(s)\|^2 = \|\Phi(\hat{\mathbb{F}}_n(s,\cdot)\|^2. \tag{3.15}$$

Note that $\hat{\mathbb{U}}_n$ and $\hat{\mathbb{T}}_n$ are $\mathcal{S}$ and $\mathbb{R}$-valued processes. If $\mathcal{S} \subset \mathbb{R}^k$ we make the following assumption

$$\Big\{\sqrt{n}(\hat{\mathbb{U}}_n(s) - \mathbb{U}(s))\Big\}_{s\in[0,1]} \stackrel{\mathcal{D}}{\Longrightarrow} \Big\{\mathbb{D}_{F_1,F_2,t}(s)\Big\}_{s\in[0,1]} \tag{3.16}$$

where $\stackrel{\mathcal{D}}{\Longrightarrow}$ means weak convergence in $\ell^\infty([0,1]|\mathbb{R}^k)$ and $\{\mathbb{D}_{F_1,F_2,t}(s)\}_{s\in[0,1]}$ is a centered, k-dimensional Gaussian process with covariance kernel

$$d_{F_1,F_2,t}(s_1,s_2) = \mathbb{E}[\mathbb{D}_{F_1,F_2,t}(s_1)\mathbb{D}^T_{F_1,F_2,t}(s_2)] \in \mathbb{R}^{k\times k}.$$

Remark 3.1 Note that the weak convergence results of the type (3.16) have been investigated for numerous types of stationary stochastic processes [see Horváth et al. (1999), Aue et al. (2009) or Dehling et al. (2013)]. However, the detection of relevant change-points by testing hypothesis of the form (3.3) requires weak convergence results in the non-stationary situation (3.1), for which – to our best knowledge – no results are available. In particular, as it will be demonstrated in Section A.1 of the online supplement, the distribution of the limiting processes $\mathbb{D}_{F_1,F_2,t}$ depends on the distribution functions F_1, F_2 and the change-point t in a complicated way. Only in the case $F_1 = F_2$ it simplifies to the standard situation, which is usually considered in change-point analysis. Intuitively many results for stationary processes mentioned in the cited references should also be available in the non-standard situation (3.1), but the limiting distribution is more complicated and this has to be worked out for each case under consideration. In Section A.1 of the online supplement we illustrate the arguments for this generalization in the case of a strong mixing process (satisfying (3.1)).
In the same section similar results will be established for the sequential process $\hat{\mathbb{F}}_n$, that is

$$\Big\{\sqrt{n}(\hat{\mathbb{F}}_n(s,z) - E_{F_1,F_2,t}(s,z))\Big\}_{s\in[0,1],z\in\mathbb{R}^d} \stackrel{\mathcal{D}}{\Longrightarrow} \Big\{\mathbb{G}_{F_1,F_2,t}(s,z)\Big\}_{s\in[0,1],z\in\mathbb{R}^d}, \tag{3.17}$$

where $\mathbb{G}_{F_1,F_2,t}$ denotes a centered $(d+1)$-dimensional Gaussian process on $[0,1]\times\mathbb{R}^d$ with covariance kernel

$$g_{F_1,F_2,t}(s_1,z_1,s_2,z_2) = \mathbb{E}[\mathbb{G}_{F_1,F_2,t}(s_1,z_1)\mathbb{G}_{F_1,F_2,t}(s_2,z_2)] = k_t(s_1,s_2,z_1,z_2).$$

Consequently, if the functional Φ in (3.10) is (for example) Hadamard differentiable, weak convergence of the process $\{\sqrt{n}(\hat{\mathbb{U}}_n(s) - \mathbb{U}(s))\}_{s\in[0,1]}$ is a consequence of the representation (3.14) and (3.17). Some details are given in Remark 3.2 below. However, many functionals of interest in change-point analysis (such as the mean or variance) do not satisfy this property, and for this reason we also state (3.16) as a basic assumption, which has to be checked in concrete applications. An example where (3.17) can be used directly consists in the problem of detecting a relevant change in the distribution function and will be given in Section 4.

Theorem 3.1 *If $\mathcal{S} \subset \mathbb{R}^k$ and the assumptions* (3.16) *and* (3.4) *are satisfied, then*

$$\sqrt{n}\,\Big(\int_0^1 \hat{\mathbb{T}}_n^2(s)ds - \int_0^1 \mathbb{T}^2(s)ds\Big) \stackrel{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0,\sigma^2_{F_1,F_2,t}),$$

where $\mathbb{T}^2(s)$ *and* $\int_0^1 \mathbb{T}^2(s)ds$ *are given in* (3.9) *and* (3.12), *respectively. Here the asymptotic variance is given by*

$$\sigma^2_{F_1,F_2,t} \;=\; 4(\theta(F_1)-\theta(F_2))^T \cdot \Gamma(t,F_1,F_2)\cdot(\theta(F_1)-\theta(F_2)), \tag{3.18}$$

where the matrix $\Gamma \in \mathbb{R}^{k\times k}$ *is defined by*

$$\Gamma(t,F_1,F_2) = \int_0^1\int_0^1 (s_1\wedge t - s_1 t)(s_2\wedge t - s_2 t)d_{F_1,F_2,t}(s_1,s_2)ds_1ds_2. \tag{3.19}$$

Proof. Let $\langle\cdot,\cdot\rangle$ denote the inner product on $\mathbb{R}^k$. Observing the representation

$$\hat{\mathbb{T}}^2_n(s) - \mathbb{T}^2_n(s) = \|\hat{\mathbb{U}}_n(s)-\mathbb{U}(s)\|^2 + 2\langle \mathbb{U}(s), \hat{\mathbb{U}}_n(s)-\mathbb{U}(s)\rangle$$

it follows from assumption (3.16) that

$$\Big\{\sqrt{n}(\hat{\mathbb{T}}^2_n(s)-\mathbb{T}^2(s))\Big\}_{s\in[0,1]} \stackrel{\mathcal{D}}{\Longrightarrow} \Big\{2\langle \mathbb{U}(s), \mathbb{D}_{F_1,F_2,t}(s)\rangle\Big\}_{s\in[0,1]}.$$

Now the continuous mapping theorem yields

$$\sqrt{n}\,\Big(\int_0^1 \hat{\mathbb{T}}^2_n(s)ds - \int_0^1 \mathbb{T}^2(s)ds\Big) \stackrel{\mathcal{D}}{\Longrightarrow} 2\int_0^1 \langle \mathbb{U}(s), \mathbb{D}_{F_1,F_2,t}(s)\rangle ds,$$

and standard arguments show that the random variable on the right hand side is normally distributed with mean 0 and variance

$$\begin{aligned}\sigma^2_{F_1,F_2,t} &= 4\int_0^1\int_0^1 \mathbb{E}[\langle \mathbb{U}(s_1), \mathbb{D}_{F_1,F_2,t}(s_1)\rangle\langle \mathbb{U}(s_2), \mathbb{D}_{F_1,F_2,t}(s_2)\rangle]ds_1ds_2\\ &= 4(\theta(F_1)-\theta(F_2))^T\cdot\Gamma(t,F_1,F_2)\cdot(\theta(F_1)-\theta(F_2)).\end{aligned}$$

□

Remark 3.2 A similar statement can be derived under the assumption (3.17) if the function Φ in (3.11) is Hadamard differentiable at the point $E_{F_1,F_2,t}$ (tangentially to an appropriate subset, if necessary). In this case it follows from (3.17) and the same arguments as given in the proof of Theorem 3.1 that

$$\sqrt{n}\,\Big(\int_0^1 \hat{\mathbb{T}}^2_n(s)ds - \int_0^1 \mathbb{T}^2(s)ds\Big) \stackrel{\mathcal{D}}{\Longrightarrow} 2\int_0^1 \big\langle \Phi'(\mathbb{G}_{F_1,F_2,t}(s,\cdot)), \Phi(E_{F_1,F_2,t}(s,\cdot))\big\rangle ds$$

where Φ' denotes the Hadamard derivative of Φ and $\langle,\rangle$ is the inner product on the (not necessarily finite dimensional) Hilbert space $\mathcal{S}$. The details are omitted for the sake of brevity.

3.3 ***Testing for Relevant Changes*****:** It follows from Theorem 3.1 and (3.12) that

$$\sqrt{n}\,\Big(\frac{3}{(t(1-t))^2}\int_0^1 \hat{\mathbb{T}}^2_n(s)ds - \|\theta(F_1)-\theta(F_2)\|^2\Big) \stackrel{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0,\tau^2_{F_1,F_2,t}), \tag{3.20}$$

where the asymptotic variance is given by

$$\tau^2_{F_1,F_2,t} = \frac{9\sigma^2_{F_1,F_2,t}}{(t(1-t))^4} = \frac{36(\theta(F_1)-\theta(F_2))^T \cdot \Gamma(t,F_1,F_2)\cdot(\theta(F_1)-\theta(F_2))}{(t(1-t))^4} \tag{3.21}$$

and $\sigma^2_{F_1,F_2,t}$ is defined in (3.18). In the following discussion let $\hat{t}$ denote a consistent estimator of the change-point t, such that $|\hat{t}-t| = o_p(1/\sqrt{n})$, whenever $\theta(F_1) \neq \theta(F_2)$ and $\hat{t} \xrightarrow{\mathcal{D}} T_{\max}$ whenever $\theta(F_1) = \theta(F_2)$, where $T_{\max}$ denotes a $[0,1]$-valued random variable. Typically, the estimator $\hat{t} = \operatorname{argmax}_{s\in[0,1]} \|\hat{\mathbb{U}}_n(s)\|$ satisfies these assumptions with $T_{\max} = \operatorname{argmax}_{s\in[0,1]} \|\mathbb{G}(s)\|$ for some Gaussian process $\mathbb{G}$ [for a recent review on the relevant literature see Jandhyala et al. (2013)]. Consequently, if $\hat{\sigma}^2$ is an estimator of $\sigma^2_{F_1,F_2,t}$, we obtain by $\hat{\tau}^2 = \frac{9\hat{\sigma}^2}{(\hat{t}(1-\hat{t}))^4}$ an estimate of the asymptotic variance in (3.21). This yields for the statistic

$$\hat{\mathbb{M}}^2_n = \frac{3}{(\hat{t}(1-\hat{t}))^2}\int_0^1 \hat{\mathbb{T}}^2_n(s)ds$$

the weak convergence

$$\frac{\sqrt{n}}{\hat{\tau}}\left(\hat{\mathbb{M}}^2_n - \|\theta(F_1)-\theta(F_2)\|^2\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \tag{3.22}$$

whenever $\theta(F_1) \neq \theta(F_2)$. On the other hand, if $\theta(F_1) = \theta(F_2)$ we have

$$\sqrt{n}\int_0^1 \hat{\mathbb{T}}^2_n(s)ds \xrightarrow{P} 0. \tag{3.23}$$

Theorem 3.2 *If assumption* (3.22) *is satisfied, then the test, which rejects the null hypothesis* (3.3) *of no relevant change, whenever*

$$\hat{\mathbb{M}}^2_n \geq \Delta^2 + u_{1-\alpha}\frac{\hat{\tau}}{\sqrt{n}}, \tag{3.24}$$

is a consistent asymptotic level α test.

Proof. Define $\delta = \|\theta(F_1)-\theta(F_2)\|$ and assume that the null hypothesis $\delta \leq \Delta$ holds. If $\delta > 0$ it follows from (3.22) that the probability of rejection by the decision rule (3.24) is given by

$$\begin{aligned}\beta_n(\delta) = \mathbb{P}_\delta\Big(\hat{\mathbb{M}}^2_n \geq \Delta^2 + u_{1-\alpha}\frac{\hat{\tau}}{\sqrt{n}}\Big) &= \mathbb{P}_\delta\Big(\frac{\sqrt{n}(\hat{\mathbb{M}}^2_n-\delta^2)}{\hat{\tau}} \geq \frac{\sqrt{n}(\Delta^2-\delta^2)}{\hat{\tau}} + u_{1-\alpha}\Big) \\ &\leq \mathbb{P}_\delta\Big(\frac{\sqrt{n}(\hat{\mathbb{M}}^2_n-\delta^2)}{\hat{\tau}} \geq u_{1-\alpha}\Big) \xrightarrow[n\to\infty]{} \alpha.\end{aligned} \tag{3.25}$$

Similarly, if $\delta = 0$ (which implies ($\mathbb{U}(s) \equiv 0$), we obtain from (3.23)

$$\beta(0) = \mathbb{P}\Big(\sqrt{n}\int_0^1 \hat{\mathbb{T}}^2_n(s)ds \geq \frac{\hat{t}^2(1-\hat{t}^2)^2}{3}(\sqrt{n}\Delta^2 + u_{1-\alpha}\hat{\tau})\Big) \xrightarrow[n\to\infty]{} 0. \tag{3.26}$$

Consequently, the test, which rejects the null hypothesis whenever (3.24) is satisfied, is an asymptotic level α-test. On the other hand, under the alternative $\delta > \Delta$, a similar argument shows that $\beta_n(\delta) \xrightarrow[n\to\infty]{} 1$, which proves consistency. □

The choice of the estimators $\hat{\tau}^2$ and $\hat{t}$ depends on specific examples under consideration and will be discussed in more detail in Section 5, where we illustrate the methodology by several examples.

Remark 3.3
(a) It is worthwhile to mention that for the problem of testing the "classical" hypothesis $H_0 : \theta_1 = \theta_2$ the test (3.24) proposed in this paper is usually less powerful than the classical CUSUM test independently of the size of Δ^2. The reason for this consists in the fact that it follows from assumption (3.16) and the continuous mapping theorem that under the null hypothesis $H_0 : \theta_1 = \theta_2$ the statistic $n \int_0^1 \mathbb{T}_n^2(s)ds$ converges weakly to a non-degenerate random variable. Consequently, we observe from (3.26) that for reasonable sample sizes the level of the test is practically zero, where the classical CUSUM test has approximately level α. As a consequence, the power of the classical test for the hypothesis $H_0 : \theta_1 = \theta_2$ is usually larger than the power of the test (3.24). On the other hand, the new test (3.24) has a substantially smaller type I error. Therefore (without any adjustment of the nominal level) both test are not comparable and the test for a relevant change should not be used for the "classical" hypothesis $H_0 : \theta_1 = \theta_2$.
(b) If the true change point is very close to 0 or 1 and the sample size is fixed it is intuitively clear that the new test (3.24) has a similar behavior as in the case where there is no change point in the process. Observing (3.23) we therefore expect that the new test is conservative in this case. These observations have been confirmed in a simulation study, which is not displayed for the sake of brevity. Moreover these findings are in line with results in classical change point analysis. For example, Andrews (1993) recommended to restrict the interval $[0, 1]$ to $[\varepsilon, 1 - \varepsilon]$ for a small constant $\varepsilon > 0$ in order to gain power of the CUSUM test for the hypothesis $H_0 : \theta_1 = \theta_2$, and a similar strategy could be applied in the problem of testing for relevant changes in the process.

3.3 *Further Discussion*: In this section we briefly mention two further applications of the new approach. First we note that testing hypotheses of the form (1.1) and (1.2) does not allow to control the type II error if the null hypothesis of no (relevant) change point cannot be rejected and subsequent data analysis is performed under the assumption of no change point. If the statistician is interested in controlling the error of erroneously deciding for a non relevant change point, we propose testing hypotheses of the form (1.4) for the similarity of the parameters. The corresponding test is easily obtained form the previous discussion and rejects the null hypothesis $H_0 : \|\theta_1 - \theta_2\| > \Delta$ in favor of $H_1 : \|\theta_1 - \theta_2\| \leq \Delta$, whenever

$$\hat{\mathbb{M}}_n^2 \leq \Delta^2 + u_\alpha \frac{\hat{\tau}}{\sqrt{n}}. \qquad (3.27)$$

Secondly, $\hat{\mathbb{M}}_n^2$ provides an estimate of the magnitude of the change and it is of particular importance to quantify the uncertainty of the estimate. This can easily be achieved using the result on weak convergence specified in (3.22). For example a two-sided confidence interval for the squared distance $\|\theta_1 - \theta_2\|^2$ between the parameters θ_1 and θ_2 is given by

$$\big[\hat{\mathbb{M}}_n^2 - u_{1-\alpha/2}\frac{\hat{\tau}}{\sqrt{n}}, \hat{\mathbb{M}}_n^2 + u_{1-\alpha/2}\frac{\hat{\tau}}{\sqrt{n}}\big]. \tag{3.28}$$

The coverage probabilities of this interval are investigated by means of a simulation study in Section 5.

4 Applications: Detecting Relevant Change-Points

In this section we discuss several examples to illustrate the theory developed in Section 3. In particular, we concentrate on the detection of relevant changes in the mean, coefficients in a linear regression and a relevant change in the distribution itself. Further examples discussing changes in the variance and correlation are presented in Section C of the online supplement. In order to be precise we assume that the assumptions of Theorem A.1 and A.2 in Section A.1 are satisfied. Similar results can be derived for alternative dependency concepts.

4.1 *Relevant Changes in the Mean*: The most prominent example of change-point analysis in model (3.1) consists in the investigation of structural breaks in the mean $\mu = \theta_{\text{mean}}(F) = \int_{\mathbb{R}^d} zF(dz)$. While the "classical" change-point problem H_0: $\mu_1 = \mu_2$ versus H_1: $\mu_1 \neq \mu_2$ has been investigated by numerous authors [see Csörgő and Horváth (1997) for a survey of methods for the independent case and Aue and Horváth (2013) for an extension to dependent data], we did not find any references on testing the hypotheses (2.1) of relevant change-points in the mean. Note that in contrast to the discussion of Section 2 and to most of the literature, we do not assume that the stochastic features of the process besides the mean coincide before and after the breakpoint. In particular, the variances or more generally the dependency structures before and after the change-point can be different, although the means μ_1 and μ_2 are "close", i.e. $\|\mu_1 - \mu_2\| \leq \Delta$. Theorem A.2 in the online supplement establishes condition (3.16), where the covariance kernel of the limiting process is defined in (A.14) and (A.15) with $\theta_{\text{lin}}(I\{W_k(\ell) \leq \cdot\}) = \theta_{\text{mean}}(I\{W_k(\ell) \leq \cdot\}) = W_k(\ell)$. Consequently, the corresponding asymptotic variance in (3.20) is given by

$$\begin{aligned} \tau^2_{F_1,F_2,t} \;=\; & \frac{4}{5(t(1-t))^2}(\mu_1-\mu_2)^T\Big\{t(5-10t+6t^2)V_1^{\text{mean}} \\ & + (1-3t+8t^2-6t^3)V_2^{\text{mean}}\Big\}(\mu_1-\mu_2) \end{aligned} \tag{4.1}$$

where V_1^{mean} and V_2^{mean} are defined in Theorem A.2 in the online supplement with $\theta_{\text{lin}}(I\{W_k(\ell) \leq \cdot) = W_k(\ell)$ $(\ell = 1, 2)$. Then, for a d-dimensional sample $Z_1, \ldots, Z_n$ with $Z_i = (Z_{i,1}, \ldots, Z_{i,d})$, $i = 1, \ldots, n$, the test statistic is obtained as

$$\hat{\mathbb{M}}_n^2 = \frac{3}{(\hat{t}(1-\hat{t}))^2}\frac{1}{n}\sum_{i=1}^n T_n^2(i), \tag{4.2}$$

where $T_n(i) = \sum_{k=1}^d T_k^2(i)$, $T_k(i) = \frac{1}{n}\sum_{j=1}^i Z_{j,k} - \frac{i}{n^2}\sum_{j=1}^n Z_{j,k}$ and $\hat{t} = \frac{1}{n}\text{argmax}_{1\le i\le n}|T_n(i)|$. The null hypothesis of no relevant change in the mean with potentially different variances before and after the change-point is rejected whenever (3.24) holds. The estimator $\hat{\tau}^2_{F_1,F_2,t}$ of the asymptotic variance is obtained from formula (4.1) by replacing the unknown quantities t, μ_1, and μ_2 by their empirical counterparts $\hat{t}$, $\hat{\mu}_1$ and $\hat{\mu}_2$ [see formula (2.6) and (2.7) in Section 2]. For the estimation of the long-run variances V_1^{mean} and V_2^{mean} in (4.1) one has to account for potential serial dependence, and we propose a kernel-based estimator as described in Andrews (1991) in the two different subsamples. More precisely we choose the Bartlett kernel and a data-adaptive bandwidth $\gamma_n = 1.1477(4\hat{\rho}^2\lfloor n\hat{t}\rfloor/(1-\hat{\rho}^2)^2)^{1/3}$. Here, $\hat{\rho}$ is the mean of the estimated AR(1) parameters for the k univariate series $\{Z_{i,k} \mid i = 1,\dots,n\}$ $(k = 1,\dots,d)$ for the sample before the estimated break point, respectively (note that this choice of $\hat{\rho}$ is optimal for an $AR(1)$ process in a univariate context). The estimator of V_1^{mean} is then defined by

$$\hat{V}_1^{\text{mean}} = \frac{1}{\lfloor n\hat{t}\rfloor}\sum_{i=1}^{\lfloor n\hat{t}\rfloor}(Z_i-\hat{\mu}_1)(Z_i-\hat{\mu}_1)^T + \frac{2}{\lfloor n\hat{t}\rfloor}\sum_{j=1}^{\lfloor n\hat{t}\rfloor-1} k\left(\frac{j}{\gamma_n}\right)\sum_{i=1}^{\lfloor n\hat{t}\rfloor-j}(Z_i-\hat{\mu}_1)(Z_{i+j}-\hat{\mu}_1)^T$$

with $k(x) = (1-|x|)\mathbf{1}_{\{|x|\le 1\}}$ and an analogue expression is used for the estimation of the quantity V_2^{mean} in (4.1). The choice of the bandwidth has no big impact in the case of serial independence, but reduces size distortions if there is high serial dependence.

4.2 *Relevant Changes in the Parameters of a Regression*: Early results on change-point inference in linear regression models can be found in Kim and Siegmund (1989), Hansen (1992), Andrews (1993), Kim and Cai (1993) and Andrews et al. (1996). More recent work on this problem has been done by Chen et al. (2013) and Nosek and Szkutnik (2014), among others. In this section we introduce the problem of testing for relevant changes in the parameters of a regression model. To be precise, we consider the common linear regression model

$$Y_{n,i} = g^T(X_i)\beta_{(i)} + \varepsilon_i \qquad i = 1,\dots,n$$

where $\beta_{(1)} = \ldots = \beta_{(\lfloor nt\rfloor)} \neq \beta_{(\lfloor nt\rfloor+1)} = \ldots\beta_{(n)} = \beta_2$ and $(X_i)_{i=1,\dots,n}$ and $(\varepsilon_i)_{i=1,\dots,n}$ are independent strictly stationary processes. In the notation of Section 3 and Section A.1 in the online supplement we have $Z_{n,1},\dots,Z_{n,\lfloor nt\rfloor} = (X_1,Y_{n,1}),\dots,(X_{\lfloor nt\rfloor},Y_{n,\lfloor nt\rfloor}) \sim F_1$, and $Z_{n,\lfloor nt\rfloor+1},\dots,Z_{n,n} = (X_{\lfloor nt\rfloor+1},Y_{n,\lfloor nt\rfloor+1}),\dots,(X_n,Y_{n,n}) \sim F_2$, where F_1 and F_2 are the joint distribution functions before and after the change-point, respectively. Note that the marginal distribution F_X of the predictor X satisfies $F_X = F_1(\cdot,\infty) = F_2(\cdot,\infty)$ by these assumptions.
In order to construct tests for the null hypothesis of no relevant change

$$H_0: \|\beta_1-\beta_2\| \le \Delta \quad \text{versus} \quad H_1: \|\beta_1-\beta_2\| > \Delta \tag{4.3}$$

we assume that the $k\times k$ matrix

$$B := \int_{\mathbb{R}^{d+1}} g(x)g^T(x)F(dx,dy) = \int_{\mathbb{R}^d} g(x)g^T(x)F_X(dx) \tag{4.4}$$

is non-singular and note that the parameter β_i can be represented as

$$\beta_i = \theta(F_i) = \Big(\int_{\mathbb{R}^{d+1}} g(x)g^T(x)F_i(dx,dy)\Big)^{-1}\Big\{\int_{\mathbb{R}^{d+1}} yg(x)F_i(dx,dy)\Big\} \qquad i=1,2. \tag{4.5}$$

For an illustration we consider the case $k=1$, $g(x)=x$ that is $Y_i = \beta X_i + \varepsilon_i$ $(i=1,\ldots,n)$. The test statistic is defined by (4.2) where $T_n(i) = \frac{1}{\hat{B}_n}(\frac{1}{n}\sum_{j=1}^i X_jY_j - \frac{i}{n^2}\sum_{j=1}^n X_jY_j)$, $\hat{B}_n = \frac{1}{n}\sum_{i=1}^n X_i^2$ and $\hat{t} = \frac{1}{n}\text{argmax}_{1\le i\le n}|T_n(i)|$. The null hypothesis (4.3) of no relevant change in the parameter β is rejected whenever (3.24) is satisfied, where

$$\begin{aligned}\tau^2_{F_1,F_2,t} \;=\; & \frac{4(\beta_1-\beta_2)^2}{5B^2t^2(1-t)^2}\Big\{V_1(1+2(1-t)t) + V_0\Big[5t(1-t)((1-t)\beta_1+t\beta_2)^2 \\ & +(t^3\beta_1^2+(1-t)^3\beta_2^2)\Big]\Big\}.\end{aligned} \tag{4.6}$$

See Section B.1 in the online supplement for a derivation of these formulas and the definition of V_1 and V_2. We replace the unknown quantities $t, B, \beta_1, \beta_2, V_0$ and V_1 by $\hat{t}$, $\hat{B}_n$, the OLS-estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ from the two subsamples before and after the estimated change-point $\lfloor n\hat{t}\rfloor$ and the estimators

$$\begin{aligned}\hat{V}_0 &= \frac{1}{n}\sum_{i=1}^n\big(X_i^2 - \frac{1}{n}\sum_{j=1}^n X_j^2\big)^2, \\ \hat{V}_1 &= \frac{1}{n}\sum_{i=1}^{\lfloor n\hat{t}\rfloor}\big(X_i\hat{\varepsilon}_i^{(1)} - \frac{1}{\lfloor n\hat{t}\rfloor}\sum_{j=1}^{\lfloor n\hat{t}\rfloor}X_j\hat{\varepsilon}_j^{(1)}\big)^2 + \frac{1}{n}\sum_{i=\lfloor n\hat{t}\rfloor+1}^{n}\Big(X_i\hat{\varepsilon}_i^{(2)} - \frac{1}{n-\lfloor n\hat{t}\rfloor}\sum_{j=\lfloor n\hat{t}\rfloor+1}^{n}X_j\hat{\varepsilon}_j^{(2)}\Big)^2,\end{aligned}$$

where $\hat{\varepsilon}_i^{(1)}$ and $\hat{\varepsilon}_i^{(2)}$ are the least squares residuals form the sample before and after the estimated change-point. In the case of serial dependence the estimators $\hat{V}_0$ and $\hat{V}_1$ have to be modified appropriately and the details are omitted for the sake of brevity. We finally mention that the results of this section can generalized to error processes $(\varepsilon_i)_{i=1}^n$ with different strictly stationary phases before and after the change-point.

4.3 *Relevant Changes in the Distribution*: In order to investigate the problem of a relevant change with respect to the distribution in a univariate sequence of the form (3.1) we consider the distance

$$\|F_1 - F_2\| = \Big(\int_{\mathbb{R}}(F_1(z)-F_2(z))^2dz\Big)^{1/2} \tag{4.7}$$

on the set of all distribution functions with existing first moment. In this case the null hypothesis of no relevant change in the distribution function is formulated as

$$H_0: \|F_1-F_2\| \le \Delta \qquad\qquad H_1: \|F_1-F_2\| > \Delta. \tag{4.8}$$

For a given sample $Z_1,\ldots,Z_n$ of independent random variables the test statistic for the null hypothesis (4.7) of no relevant change in the distribution function is defined by (4.2), where

$$T_n(i) = \sum_{k=1}^{n-1}(Z_{(k+1)} - Z_{(k)})\Big(\frac{1}{n}\sum_{j=1}^{i}\mathbf{1}_{\{Z_j\le Z_{(k)}\}} - \frac{i}{n^2}\sum_{j=1}^{n}\mathbf{1}_{\{Z_j\le Z_{(k)}\}}\Big)^2. \tag{4.9}$$

Here $Z_{(1)}, \ldots, Z_{(n)}$ denotes the order statistic of $Z_1, \ldots, Z_n$. The null hypothesis of no relevant change in the distribution function is rejected whenever (3.24) holds. For the definition of an estimator of the asymptotic variance we note that we assumed independent observations such that we have

$$\tau^2_{F_1,F_2,t} = \frac{4}{5t^2(1-t)^2}\Big[t(5-10t+6t^2)\int_{\mathbb{R}^2}\Delta(z_1,z_2)(F_1(z_1)\wedge F_1(z_2) - F_1(z_1)F_1(z_2))dz_1dz_2 \\ +(1-3t+8t^2-6t^3)\int_{\mathbb{R}^2}\Delta(z_1,z_2)(F_2(z_1)\wedge F_2(z_2) - F_2(z_1)F_2(z_2))dz_1dz_2\Big] .$$

where we use the notation $\Delta(z_1, z_2) = (F_1(z_1) - F_2(z_1))(F_1(z_2) - F_2(z_2))$. The estimator $\hat{\tau}^2_{F_1,F_2,t}$ is now obtained by plugging in $\hat{t} = \frac{1}{n}\text{argmax}_{1\leq i\leq n}|T_n(i)|$ and replacing the unknown distribution functions by F_1 and F_2 by the empirical distribution functions calculated from the subsample before and after the estimated change-point.

5 Finite Sample Properties

In this section, we illustrate the application of the new testing procedure and provide some finite sample evidence. For the sake of brevity we investigate three cases: the detection of relevant changes in the mean, parameter of a linear regression model and a relevant change in the distribution function. Similar results can be obtained for the other testing problems considered in the online supplement but are not displayed here for the sake of brevity. In all examples under consideration, we performed 5000 replications of the test (3.24) at significance level $\alpha = 0.05$. We also note that it follows from the proof of Theorem 3.2 that the power of the test (3.24) is approximately given by

$$\beta_n(\delta) \approx 1 \; - \; \Phi\Big(\frac{\sqrt{n}(\Delta^2 - \delta^2)}{\hat{\tau}_{F_1,F_2,t}} + u_{1-\alpha}\Big). \tag{5.1}$$

Similarly, we obtain a formula for the p-value of the test, that is

$$1 - \Phi\Big(\sqrt{\frac{n}{\hat{\tau}^2_{F_1,F_2,t}}}\,\Big(\hat{\mathbb{M}}^2_n - \Delta^2\Big)\Big), \tag{5.2}$$

where Φ is the distribution function of the standard normal distribution. These formulas will be helpful to understand some properties of the test (3.24).

5.1 *Relevant Changes in the Mean*: At first, we look at the test for changes in the mean as discussed at the beginning of in Section 4 focussing on a one-dimensional sample $Z_1, \ldots, Z_n$. In Figure 1 we display the rejection probabilities of the test (3.24) for sample sizes $n = 200, 500, 1000$ and independent normally distributed random variables with mean $\mu_1 = 0$ in the first half and mean $\mu_2 = 1$ in the second half of the sample, i.e. $t = 0.5$. The variance is constant and equal to 1. The left part of Figure 1 presents the empirical rejection probabilities of the test (3.24) for

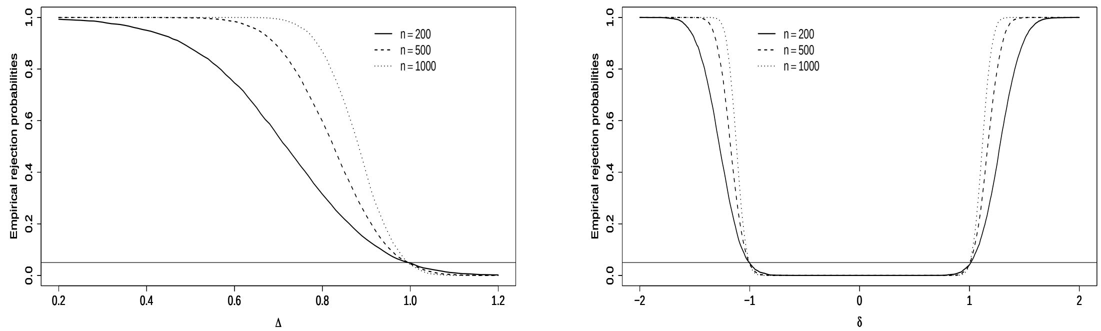

Figure 1: *Empirical rejection probabilities of the test* (3.24) *for the null hypothesis of no relevant change in the mean, where* $\mu_1 = 0$*,* $\mu_2 = 1$*,* $t = 0.5$*. Left panel: constant* $\delta = 1$*, varying* Δ*. Right panel: constant* $\Delta = 1$*, varying* δ*.*

fixed $\delta = 1$, where the parameter Δ, which defines the size of a relevant change in the hypothesis (2.1), varies in the interval $[0.2, 1.2]$. We observe that the power of the test decreases in Δ as predicted by formula (5.1). For $\Delta = 1$, the power is approximately 0.05, which shows that the test keeps its nominal level.

The right part of Figure 1 displays the power curve of the test (3.24) for the same sample sizes and $\Delta = 1$, where the "true" difference $\delta = \mu_1 - \mu_2$ varies in the interval $[-2, 2]$. As expected, the power curve is U-shaped with a minimum at $\delta = 0$ [note that the power of the test converges to zero in this case - see formula (3.26) in the proof of Theorem 3.2]. Again the nominal level is well approximated at the boundary of the null hypothesis, that is $\delta = \pm 1$. We also observe that the type I error is much smaller inside the interval $\{\delta \in \mathbb{R} \mid |\delta| < \Delta\}$.

Figures as displayed in the left part of Figure 1 are useful to obtain the minimal size of the parameter Δ in (2.1) such that the null hypothesis of no relevant change of size Δ is rejected at controlled type I error, while the figure in the right part directly displays the power function of the test (3.24). Both types essentially provide the same information and for the sake of brevity we focus in the following discussion only on the power function. Moreover, due to the obvious symmetry, we just present the values for $\delta \geq 0$.

In Figure 2 we analyze the effect of changes in the variances on the testing procedure, where the sample size is fixed as $n = 200$ and the setting is the same as in Figure 1. The left part of the figure shows the power of the test (3.24) for the null hypothesis of no relevant change in the mean, where the variances are the same before and after the change-point and given by $\sigma^2 = 0.2^2, 0.5^2, 1, 2^2, 5^2$. We observe that the approximation of the nominal level is rather accurate at the point $\delta = 1$. Moreover, the rejection probabilities decrease in σ^2. Note that there is essentially no power for $\sigma^2 = 5^2$ because in this case the variance is dominating the mean. Moreover, in this case the level of the test is not very well approximated, which is due to the fact that it is difficult to estimate the change-point t accurately under a large signal to noise ratio. In the right part of Figure 2 we display the effect of changing variances in the same setting as

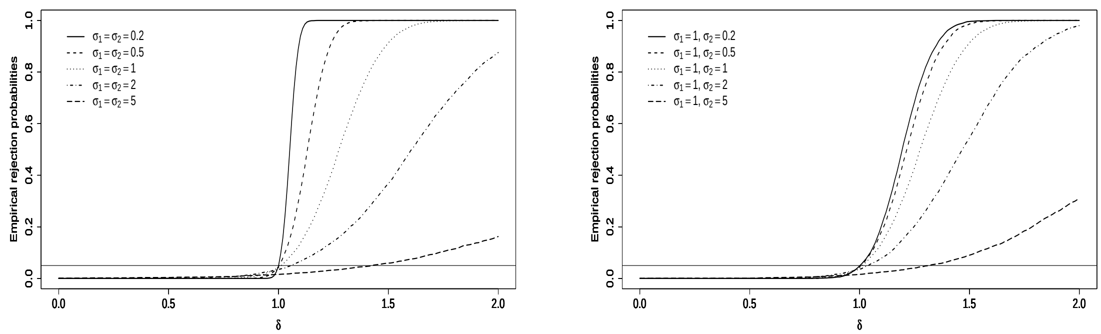

Figure 2: *Empirical rejection probabilities of the test* (3.24) *for the null hypothesis of no relevant change in the mean, where* $\Delta = 1$. *Left panel: constant variances. Right panel: different variances before and after the change-point. The sample size is* $n = 200$ *and the horizontal line marks the significance level* 0.05.

in the left part where the variance in the first half is equal to 1 and in the second half given by $\sigma^2 = 0.2^2, 0.5^2, 1, 2^2, 5^2$. We do not observe substantial differences with respect to the quality of approximation of the nominal level. Compared to the case of constant variances the power is in general lower for $\sigma_2^2 > 1$ and higher for $\sigma_2^2 < 1$. These empirical findings reflect the asymptotic theory, because the asymptotic variance of the estimator $\hat{\mathbb{M}}_n^2$ is an increasing function of σ_1^2 and σ_2^2 [see formula (4.1)] and it follows from (5.1) that the power of the test (3.24) is decreasing with this variance.

Finally, we investigate the effect of serial dependence on the test (3.24) for the null hypothesis of no relevant change in the the mean. For this purpose we generate $n = 200$ and $n = 500$ realizations of an $AR(1)$ process with AR parameter $\rho = 0, 0.4, 0.8$, mean zero and standard normal distributed innovations using the R-function *arima.sim*. Note that such a process fulfills a strong mixing condition with mixing coefficients that decay exponentially [see for example Doukhan (1994), Theorem 6, p. 99.]. After that, we add δ to the last 100 realizations. Figure 3 shows the serial dependence has an impact on the quality of the approximation of the nominal level if the sample size is $n = 200$. Moreover, the power decreases with increasing correlation. These properties have also been observed by other authors in the context of CUSUM-type testing procedures for "classical" hypotheses [see Xiao and Phillips (2002) and Aue et al. (2009b)]. Moreover, using the asymptotic theory from Section 4 we can also give a precise explanation of these observations. For the AR(1) model under consideration the quantities V_i^{mean} in (4.1) are given by $V_1^{\text{mean}} = V_2^{\text{mean}} = \rho^2/(1-\rho^2)$. Consequently the asymptotic variance $\tau^2_{F_1,F_2,t}$ is increasing with $|\rho|$ and by formula (5.1) the power is decreasing.

5.2 *Relevant Changes in the Parameters of a Regression*: In this section we investigate the finite sample properties of the test for a relevant change in the slope parameter of the regression model

$$Y_i = \beta X_i + \varepsilon_i \ , \quad i = 1, \dots, n$$

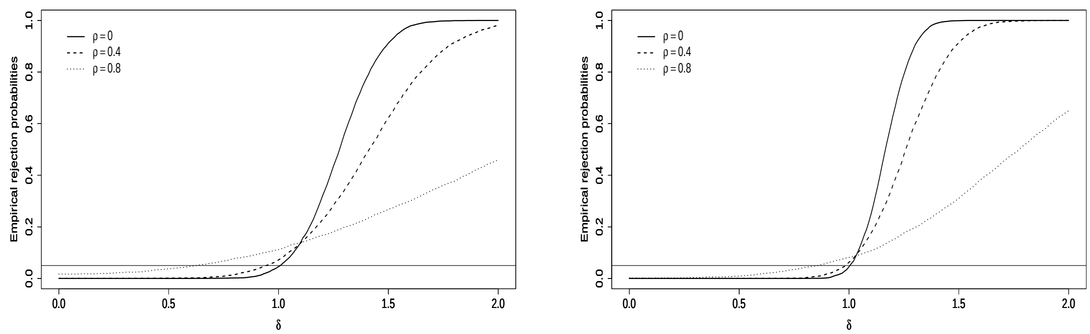

Figure 3: *Empirical rejection probabilities of the test* (3.24) *for the null hypothesis of no relevant change in the mean under serial dependence, where* $\Delta = 1$. *Left panel: sample size* $n = 200$. *Right panel: sample size* $n = 500$. *The horizontal line marks the significance level* 0.05.

based on a bivariate sample. In the left part of Figure 4 we display the power of the test (3.24) for the null hypothesis of no relevant change in the parameter β, where $\beta_1 = 0$ in the first half and $\beta_2 = \delta \geq 0$ of the sample and the explanatory variables X_i and errors ε_i in the linear regression model are independent identically standard normal distributed. The approximation of the nominal level is rather accurate and the power is increasing with the sample size. On the other hand, the power of the test for a change in the slope is lower than the power for the test for change of the same size in the mean as considered in Figure 1[1].This observation can be easily explained by the asymptotic representation of the probability of rejection in (5.1) which is a decreasing function of the asymptotic variance $\tau^2_{F_1,F_2,t}$. For the test of the null hypothesis of no relevant change in the mean and slope these variances are given by 307.2 and 576, respectively [see (4.1) and (4.6)]. In the right part of Figure 4 we display the results for heavy-tailed predictors X_i, that is $X_i \sim \sqrt{\frac{3}{5}} t_5$, where t_f denotes a t-distribution with f degrees of freedom. Note that the t-distribution is standardized with $\mathrm{Var}(X_i) = 1$. We observe a less accurate approximation of the nominal level if the sample size is $n = 200$. Moreover, t_5 distributed regressors yield also a loss in power. This observation can also be explained by formula (5.1), where the asymptotic variance $\tau^2_{F_1,F_2,0.5}$ is given by 576 and 1024 for the normal and t_5-distribution, respectively.

5.3 *Relevant Changes in the Distribution*: We continue with a brief finite sample study of the test for the null hypothesis of no relevant change in the distribution function, which was discussed in Section 4. We choose sample sizes $n = 200, 500, 1000$ with serially independent random variables, $\mathcal{N}(0,1)$-distributed in the first half and χ^2-distributed with different degrees of freedom $f = 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4$ in the second half of the sample. The χ^2-distributed random variables are standardized such that they have mean 0 and variance 1. The distance $||F_1 - F_2||$ for $f = 1$ is approximately equal to 0.2254 and this value was chosen as Δ in the test (3.24). Table 1 shows the rejection probabilities of the test (3.24). Due to the different distance

[1]Additional simulations show that this power difference still exists if we do not account for serial dependence in the mean test, that means if we consider $\hat{V}_1^{\mathrm{mean}} = \frac{1}{\lfloor n\hat{t} \rfloor} \sum_{i=1}^{\lfloor n\hat{t} \rfloor} (X_i - \hat{\mu}_1)^2$ and the analogue formula for $\hat{V}_2^{\mathrm{mean}}$.

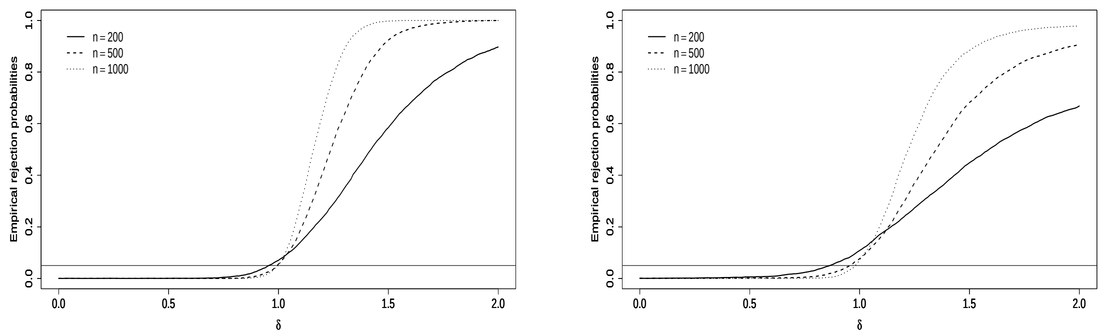

Figure 4: *Empirical rejection probabilities of the test* (3.24) *for the null hypothesis of no relevant change in the parameter of a linear regression model. Left panel: normal distributed regressors. Right panel:* t_5*-distributed regressors. The horizontal line marks the significance level* 0.05.

df	$f = 0.2$	$f = 0.4$	$f = 0.6$	$f = 0.8$	$f = 1$	$f = 1.2$	$f = 1.4$
n / δ	0.3730	0.3154	0.2764	0.2476	0.2254	0.2077	0.1932
200	0.995	0.784	0.404	0.174	0.078	0.042	0.021
500	1.000	0.978	0.614	0.221	0.069	0.023	0.006
1000	1.000	1.000	0.846	0.313	0.064	0.011	0.001

Table 1: *Empirical rejection probabilities of the test* (3.24) *for the null hypothesis of no relevant change in the distribution function. The first half of the sample is generated from a* $N(0,1)$ *and the second half from a (standardized)* χ^2*-distribution with different degrees of freedom. The size of a relevant change is defined by* $\Delta = 0.2254$ *and corresponds to* $f = 1$.

measure, they are somewhat difficult to compare with the other figures, but apparently, the test does work well. The power decreases in f as the χ^2_f-distribution, standardized such that it has mean zero and variance one, converges to the $\mathcal{N}(0,1)$-distribution for $f \to \infty$.

5.4 *Confidence Intervals*: As discussed in Section 3, the method proposed in this paper can also be used for constructing confidence intervals for the magnitude of the change between the parameters θ_1 and θ_2. We illustrate this for the case of a change in the mean considering the setting from the right panel of Figure 1. In Table 2 we display the simulated coverage probabilities of the confidence interval (3.28), where $\delta \in \{0.4, 0.6, 0.8, 1, 1.2, 1.4\}$. We observe that the empirical coverage probabilities are close to the theoretical ones. If δ approaches 0 the corresponding confidence intervals become conservative if the sample size is not too large.

5.5 *Empirical Illustration*: We illustrate the application of the new test in an example from portfolio management. For this purpose we consider continuous returns from the closing prices of BASF and Sanofi. The time period for the prices is 1st June 2009 to 1st June 2012, that is $n = 784$, and we observe the bivariate vectors (X_i, Y_i) $(i = 1, \ldots, n)$, where X represents

n / δ	0.4	0.6	0.8	1	1.2	1.4	0.4	0.6	0.8	1	1.2	1.4
	$\alpha = 0.05$						$\alpha = 0.1$					
200	.973	.955	.949	.946	.944	.944	.939	.911	.905	.903	.900	.899
500	.956	.953	.951	.952	.951	.951	.914	.906	.903	.902	.901	.901
1000	.957	.954	.953	.952	.952	.952	.907	.902	.900	.899	.899	.900

Table 2: *Empirical coverage probabilities of the confidence interval* (3.28) *for the null hypothesis of no relevant change in the mean. The first half of the sample is generated from a* $N(0,1)$*- and the second half from a* $N(\delta,1)$*-distribution.*

BASF and Y Sanofi. We first used the procedure described in Section 4 to test the hypothesis of a relevant change point in the variance Σ, where Σ is the covariance matrix of (X,Y). More precisely, we test the series $\{Z_i = (X_i^2, Y_i^2, X_iY_i)^T | \ i = 1,\dots,n\}$ for a relevant change in the mean. Assuming constant means, which is standard with financial returns, this is equivalent to testing for a relevant change in the variance-covariance-matrix [see Section C.1 in the online supplement for more details]. On a 5% significance level the test (3.24) does not reject the null hypothesis of a relevant change of size $\Delta = 10^{-4}$ in the norm of the variance matrix. The test statistic is given by $\mathbb{M}_n^2 = 4.67 \cdot 10^{-8}$ and the right-end of the critical region is given by $4.72 \cdot 10^{-8}$. Moreover, the 95%-confidence interval for the squared norm of the vector of differences in the second order moments is given by $[-1.9 \cdot 10^{-9}, 9.5 \cdot 10^{-8}]$.
Next we consider the "classical" testing problem $H_0 : \Sigma_1 = \Sigma_2$ to detect changes in the variance structure using the test proposed in Aue et al. (2009). More precisely, we consider the test statistic Ω_n defined in this paper and kernel/bandwidth chosen as described in Section 4.1, where the AR parameters are estimated from the full sample. The value of the test statistic is 1.41, while the critical value is 1.00 [see Page (1959), p. 444]. Consequently this test rejects the null $H_0 : \Sigma_1 = \Sigma_2$. The estimated break point $\hat{t}$ is 0.717.
Finally, we illustrate some consequences of the different decisions. For this purpose we consider two simplified investment strategies which are both based on the global minimum variance portfolio given by $\frac{\Sigma^{-1}(1,1)^T}{(1,1)\Sigma^{-1}(1,1)^T}$. The first strategy is a consequence of the test for the hypothesis of no relevant change points. As pointed out in the previous paragraph the method proposed in this paper does not reject this hypothesis and the matrix Σ is estimated by the empirical covariance matrix from the full sample. The second strategy is based on the test of Aue et al. (2009), which rejects the "classical" hypothesis of a change in the variance. Here the empirical covariance matrix is estimated from the sample $(X_{\lfloor n\hat{t}\rfloor+1}, Y_{\lfloor n\hat{t}\rfloor+1}), \dots (X_n, Y_n)$ (with $\hat{t} = 0.717$). We calculate the daily returns for both strategies and the profit/loss at the end of the time period: With a start amount of 10000 pounds, one would have a loss of 388.04 pounds for the first and of only 101.19 pounds for the second strategy. Therefore, in a world with only fixed but no variable transaction the second strategy would be useful if the fixed transaction costs are not higher than $388.04 - 101.19 = 286.85$ pounds. In other words, if the costs are higher, one

should not reject the null hypothesis and the break in the covariance matrix should be regarded as an irrelevant change. These results can be used for the choice of the threshold Δ in future investment strategies. More precisely, if an investor assumes that these retrospective calculations are valid for future decisions and if the transaction costs are higher than 286.85 pounds, the new test developed in this paper should be used with threshold $\Delta = 8 \cdot 10^{-5}$. If the null hypothesis of a no relevant change in the variance is not rejected, the investor should not shift his portfolio even if the test by Aue et al. (2009) rejects the "classical" hypothesis $H_0 : \Sigma_1 = \Sigma_2$.

6 Conclusions and Future Research

In this paper we have investigated the problem of testing for a relevant change in the parameters of a time series. Our work was motivated by the observation that in many cases data analysts are only interested in changes where the difference between the parameters before and after the change point exceeds a minimum threshold, say Δ. An important ingredient in our approach is the appropriate choice of this threshold, which has to be defined carefully in every particular application in order distinguish between statistical and scientific significance. The classical approach avoids this choice by simply putting $\Delta = 0$, and we strongly argue to choose this threshold thoroughly considering the scientific background of the testing problem. Statistical methodology is developed for testing the hypothesis of no relevant change points. Moreover, if a reasonable choice of the threshold based on the scientific background is not possible, our approach still provides a confidence interval for the norm of the difference between the parameters of the process before and after the change point.

The method proposed in this paper is based on an L_2-distance of the classical CUSUM process and can be applied in numerous change point problems. Asymptotic analysis is performed in order to derive (asymptotic) critical values under the assumption that the process before and after the change point is strictly stationary and a simulation study demonstrates good finite sample properties of the new test. An important and interesting topic for future research is to extend these results to non-stationary processes using a similar approach as in Zhou (2013). For example, consider the model $Y_i = \mu_i + e_i$ $(i = 1, \ldots, n)$, where $(e_i)_{i\in\mathbb{Z}}$ is a piecewise locally stationary process (PLS) in the sense of Definition 1 of Zhou (2013), and $\mu_i = \mu_1$ if $1 \leq i \leq \lfloor nt \rfloor$ and $\mu_i = \mu_2$ if $\lfloor nt \rfloor + 1 \leq i \leq n$. It then follows using similar results as presented in this paper that

$$\sqrt{n}\{\mathbb{U}_n(s) - (s \wedge t - st)(\mu_1 - \mu_2)\}_{s\in[0,1]} \stackrel{\mathcal{D}}{\Longrightarrow} \{U(s) - sU(1)\}_{s\in[0,1]}$$

where $\{U(s)\}_{s\in[0,1]}$ is a centered Gaussian process with covariance kernel

$$k(s,t) = \sum_{i=0}^{\tau} \int_{b_i}^{b_{i+1}} \sum_{k\in\mathbb{R}} \mathrm{Cov}[G_i(u, \mathcal{F}_k), G_i(u, \mathcal{F}_0)]du$$

and the constants b_i, the σ-fields $\mathcal{F}_k$ and the nonlinear filters G_i appear in the definition of the PLS $(e_i)_{i\in\mathbb{Z}}$ [see Zhou (2013)]. From this result an analogue of Theorem 2.1 could be derived.

However, the asymptotic variance depends in a complicated way on the covariance kernel k and critical values cannot be derived from the asymptotic theory. For the implementation of a test for the hypothesis of no relevant changes in the mean bootstrap methods have to be developed, which address the particular problems appearing with tests of interval hypotheses adequately. A further challenging direction for future research is the extension of the methodology for detecting relevant changes in processes with multiple break points, which have found considerable attention in the recent literature. One obvious idea is to use binary segmentation methods for this problem [see for example Cho and Fryzlewicz (2014)], but multi-scale inference [see Frick et al. (2014)] or methods based on spectral analysis [see for example Preuss et al. (2014)] might yield to other powerful methods.

Acknowledgements The authors thank Martina Stein, who typed numerous versions of this manuscript with considerable technical expertise, and Axel Bücher and Stanislav Volgushev for helpful discussions. This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Teilprojekt A1 and C1) of the German Research Foundation (DFG). Parts of this paper were written while H. Dette was visiting the Isaac Newton Institute, Cambridge, UK, in 2014 ("Inference for change-point and related processes") and the authors would like to thank the institute for its hospitality. The authors are also grateful to three referees and the associate editor. Their constructive comments on an earlier version yielded to substantial improvement of an earlier version of this paper.

References

Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):128–156.

Andrews, D. W. K., Lee, I., and Ploberg, W. (1996). Optimal change-point tests for normal linear regression. *Journal of Econometrics*, 70(1):9–36.

Aue, A., Hörmann, S., Horváth, L., and Reimherr, M. (2009a). Break detection in the covariance structure of multivariate time series models. *Annals of Statistics*, 37(6B):4046–4087.

Aue, A. and Horváth, L. (2013). Structural breaks in time series. *Journal of Time Series Analysis*, 34(1):1–16.

Aue, A., Horváth, L., Hušková, M., and Ling, S. (2009b). On distinguishing between random walk and change in the mean alternative. *Econometric Theory*, 25(2):411–441.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78.

Berger, J. O. and Delampady, M. (1987). Testing precise hypotheses. *Statistical Science*, 2(3):317–335.

Berkes, I., Gombay, E., and Horváth, L. (2009). Testing for changes in the covariance structure of linear processes. *Journal of Statistical Planning and Inference*, 139(6):2044–2063.

Berkson, J. (1938). Some difficulties of interpretation encountered in the application of the chi-square test. *Journal of the American Statistical Association*, 33(203):526–536.

Brown, R., Durbin, J., and Evans, J. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society Series B*, 37(2):149–163.

Carlstein, E. (1988). Nonparametric change-point estimation. *Annals of Statistics*, 16(1):188–197.

Chen, C., Chan, J., Gerlach, R., and Hsieh, W. (2013). A comparison of estimators for regression models with change points. *Statistics and Computing*, 21(3):395–414.

Cho, H. and Fryzlewicz, P. (2014). Multiple change-point detection for high-dimensional time series via Sparsified Binary Segmentation. *Journal of the Royal Statistical Society, Ser. B, DOI: 10.1111/rssb.12079*.

Chow, G. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 28(3):591–605.

Chow, S.-C. and Liu, P.-J. (1992). *Design and Analysis of Bioavailability and Bioequivalence Studies.* Marcel Dekker, New York.

Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis.* Wiley, New York.

Dehling, H., Durieu, O., and Tusche, M. (2013). A sequential empirical central limit theorem for multiple mixing processes with application to B-geometrically ergodic Markov chains. *arXiv:1303.4537*.

Doukhan, P. (1994). *Mixing: Properties and Examples (Lecture Notes in Statistics 85).* Springer, Berlin.

Frick, K., Munk, A., and Sieling, H. (2014). Multiscale change point inference. *Journal of the Royal Statistical Society, Series B*, 76(3):495–580.

Hansen, B. E. (1992). Tests for parameter instablity in regression with I(1) processes. *Journal of Business and Economic Statistics*, 10(3):321–335.

Horváth, L., Kokoszka, P., and Steinebach, J. (1999). Testing for changes in multivariate dependent observations with an application to temperature changes. *Journal of Multivariate Analysis*, 68(1):96–119.

Jandhyala, V., Fotopoulos, S., MacNeill, I., and Liu, P. (2013). Inference for single and multiple change-points in time series. *Journal of Time Series Analysis*, forthcoming, doi: 10.1111/jtsa12035.

Kim, H.-J. and Cai, L. (1993). Robustness of the likelihood ratio test for a change in simple linear regression. *Journal of the American Statistical Association*, 88(423):864–871.

Kim, H.-J. and Siegmund, D. (1989). The likelihood ratio test for a change-point simple linear regression. *Biometrika*, 76(3):409–423.

Krämer, W., Ploberger, W., and Alt, R. (1988). Testing for structural change in dynamic models. *Econometrica*, 56(6):1355–1369.

Lehmann, E. L. (1986). *Testing Statistical Hypotheses.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, second edition.

Lorden, G. (1971). Procedures for reacting to a change in distribution. *Annals of Mathematical Statistics*, 42(6):1897–1908.

McBride, G. B. (1999). Equivalence tests can enhance environmental science and management. *Australian & New Zealand Journal of Statistics*, 41(1):19–29.

Moustakides, G. V. (2004). Optimality of the CUSUM procedure in continuous time. *Annals of*

Statistics, 32(1):302–315.

Nosek, K. and Szkutnik, Z. (2014). Change-point detection in a shape-restricted regression model. *Statistics*, 48:641–656.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41(1-2):100–115.

Page, E. S. (1955). Control charts with warning lines. *Biometrika*, 42(1-2):243–257.

Page, E. S. (1959). K-sample analogues of the kolmogorov-smirnov and cramer-v. mises tests. *Annals of Mathematical Statistics*, 30(2):420–447.

Preuss, P., Puchstein, R., and Dette, H. (2014). Detection of multiple structural breaks in multivariate time series. *Journal of the American Statistical Association, DOI: 10.1080/01621459.2014.920613*.

Wied, D. (2013). CUSUM-type testing for changing parameters in a spatial autoregressive model for stock returns. *Journal of Time Series Analysis*, 34(1):221–229.

Wied, D., Krämer, W., and Dehling, H. (2012). Testing for a change in correlation at an unknown point in time using an extended functional delta method. *Econometric Theory*, 28(3):570–589.

Xiao, Z. and Phillips, P. (2002). A cusum test for cointegration using regression residuals. *Journal of Econometrics*, 108(1):43–61.

Zhou, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *Journal of the American Statistical Association*, 108(502):726–740.

A Online Appendix

A.1 Strong Mixing Processes

Assumptions (3.16) and (3.17) are crucial for the asymptotic analysis presented in Sections 3. If $F_1 = F_2$ (i.e. there exists no structural break) they have been verified in several situations. For example, Deo (1973) proved that

$$\text{(A.1)} \qquad \sqrt{n}\Big\{(\hat{\mathbb{F}}_n(1,z) - F(1,z))\Big\}_{z\in\mathbb{R}^d} \overset{\mathcal{D}}{\Longrightarrow} \{\mathbb{G}(1,z)\}_{z\in\mathbb{R}^d}$$

if the process $\{Z_k\}_{k=1}^n$ is stationary and strong mixing with mixing coefficients α_n converging sufficiently fast to 0, that is $\sum_{n=1}^{\infty} \alpha_n^{1/2-\tau} n^2 < \infty$ for some $\tau \in (0, 1/2)$. Here $\{\mathbb{G}(1,z)\}_{z\in\mathbb{R}^d}$ denotes a Gaussian process with covariance structure

$$k(z_1, z_2) = \sum_{k\in\mathbb{Z}} \mathbb{E}[(I\{Z_k \le z_1\} - F(z_1))(I\{Z_k \le z_2\} - F(z_2))].$$

These results can be extended to other concepts of dependency and to the sequential empirical process defined in (3.5), and for some recent results in this direction we refer to the work of Berkes et al. (2009) and Dehling et al. (2013).

However, these results are not applicable anymore in the problem of detecting relevant change-points by means of a test for the hypothesis (1.3), because statements of the form (3.16) or (3.17) are required for the case $F_1 \neq F_2$ in order to obtain the asymptotic distribution in Theorem 3.1. In this case the process under consideration is not stationary anymore. Additional difficulties appear because one has to work under the assumption of a triangular array, and it will be necessary to reflect this fact in our notation throughout this section, that is

$$\text{(A.2)} \qquad Z_{n,1}, \dots, Z_{n,\lfloor nt\rfloor} \sim F_1 \; ; \qquad Z_{n,\lfloor nt\rfloor+1}, \dots, Z_{n,n} \sim F_2,$$

where F_1 and F_2 are the distribution functions before and after the change-point. We also assume that F_1 and F_2 are continuous. In principle, it should be possible to extend the results for the case $F_1 = F_2$ to the case $F_1 \neq F_2$ for most of the commonly used dependency concepts, but a general discussion is very complicated and beyond the scope of the present paper. For these reasons we restrict ourselves to the case of strong mixing triangular arrays in the subsequent discussion and investigate assumptions (3.16) and (3.17) in this case. Other concepts of dependency could be treated similarly.

To be precise, consider the triangular array $\{Z_{n,k} \mid k = 1, \dots, n\}_{n\in\mathbb{N}}$ in (A.2) and define for $1 \le s \le t$ the σ-field $\mathcal{F}_s^t(n) = \sigma(\{Z_{n,j} \mid s \le j \le t\})$ generated by the random variable $\{Z_{n,j} \mid s \le j \le t\}$. We denote by

$$\alpha(m) = \sup_{n\in\mathbb{N}} \sup_{1\le k\le n-m} \sup\big\{|\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)| \mid A \in \mathcal{F}_{m+k}^n(n),\; B \in \mathcal{F}_1^k(n)\big\},\; m \in \mathbb{N}$$

the strong mixing coefficients of the triangular array $\{Z_{n,1}, \dots, Z_{n,n}\}_{n\in\mathbb{N}}$ and assume that for some $\eta > 0$

$$\text{(A.3)} \qquad \alpha(n) = O(n^{-(1+\eta)})$$

as $n \to \infty$. Moreover, for $\ell = 1, 2$ let $\{W_t(\ell)\}_{t\in\mathbb{Z}}$ denote strictly stationary processes, such that for each $n \in \mathbb{N}$

$$(Z_{n,1}, \dots, Z_{n,\lfloor nt \rfloor}) \stackrel{\mathcal{D}}{=} (W_1(1), \dots, W_{\lfloor nt \rfloor}(1)) \tag{A.4}$$

$$(Z_{n,\lfloor nt \rfloor +1}, \dots, Z_{n,n}) \stackrel{\mathcal{D}}{=} (W_1(2), \dots, W_{n-\lfloor nt \rfloor}(2)) \tag{A.5}$$

The interpretation of this assumption is as follows: there exist two regimes $\{W_t(1)\}_{t\in\mathbb{Z}}$ and $\{W_t(2)\}_{t\in\mathbb{Z}}$ and the process under consideration switches from one regime to the other. The following statement specifies the weak convergence of the sequential empirical process

$$\hat{\mathbb{F}}_n(s, x) = \frac{1}{n} \sum_{i=1}^{\lfloor ns \rfloor} I\{Z_{n,i} \le z\}. \tag{A.6}$$

Theorem A.1 *Let $\{Z_{n,1}, \dots, Z_{n,n}\}_{n\in\mathbb{N}}$ denote a triangular array of strong mixing random variables of the form* (A.2)*, such that* (A.3)*,* (A.4) *and* (A.5) *hold, then a standardized version of the process $\{\hat{\mathbb{F}}_n(s, z)\}_{s\in[0,1], z\in\mathbb{R}^d}$ converges weakly in $\ell([0, 1] \times \mathbb{R}^d|\mathbb{R})$, that is*

$$\Big\{\sqrt{n}(\hat{\mathbb{F}}_n(s, z) - E_{F_1,F_2,t}(s, z))\Big\}_{s\in[0,1], z\in\mathbb{R}^d} \stackrel{\mathcal{D}}{\Longrightarrow} \Big\{\mathbb{G}_{F_1,F_2,t}(s, z)\Big\}_{s\in[0,1], z\in\mathbb{R}^d}.$$

Here $E_{F_1,F_2,t}$ is defined in (3.6)*, $\mathbb{G}_{F_1,F_2,t}$ denotes a centered Gaussian process with covariance kernel*

$$\mathbb{E}[\mathbb{G}_{F_1,F_2,t}(s_1, z_1)\mathbb{G}_{F_1,F_2,t}(s_2, z_2)] = (s_1 \wedge s_2 \wedge t)k_1(z_1, z_2) + (s_1 \wedge s_2 - t)_+ k_2(z_1, z_2), \tag{A.7}$$

and the kernels k_1 and k_2 are defined by

$$k_\ell(z_1, z_2) = \sum_{i\in\mathbb{Z}} \mathrm{Cov}(I\{W_0(\ell) \le z_1\}, I\{W_i(\ell) \le z_2\}); \qquad \ell = 1, 2. \tag{A.8}$$

Proof: Recalling the definition of $\hat{\mathbb{F}}_n$ and $E_{F_1,F_2,t}$ in (A.6) and (3.6), respectively, we obtain the decomposition

$$\hat{\mathbb{F}}_n(s, z) - E_{F_1,F_2,t}(s, z) = \mathbb{X}_n^{(1)}(s, z) + \mathbb{X}_n^{(2)}(s, z) + o_p(\frac{1}{\sqrt{n}}), \tag{A.9}$$

uniformly with respect to $(s, z) \in [0, 1] \times \mathbb{R}^d$, where the processes $\mathbb{X}_n^{(1)}$ and $\mathbb{X}_n^{(2)}$ are defined by

$$\begin{aligned}
\mathbb{X}_n^{(1)}(s, z) &= \frac{1}{n} \sum_{j=1}^{\lfloor n(s\wedge t)\rfloor} (I\{Z_{n,j} \le z\} - F_1(z)) = \sum_{j=1}^{\lfloor n(s\wedge t)\rfloor} Y_{n.j}, \\
\mathbb{X}_n^{(2)}(s, z) &= \frac{1}{n} I\{s > t\} \sum_{j=\lfloor n(s\wedge t)\rfloor+1}^{\lfloor ns \rfloor} (I\{Z_{n,j} \le z\} - F_2(z)) = I\{s > t\} \sum_{j=\lfloor n(s\wedge t)\rfloor+1}^{\lfloor ns \rfloor} Y_{n.j},
\end{aligned}$$

and the random variables $Y_{n,j}$ are defined by

$$\text{(A.10)}\quad Y_{n,j}(z) = I\{j \le \lfloor nt \rfloor\} \frac{I\{Z_{n,j} \le z\} - F_1(z)}{n} + I\{j > \lfloor nt \rfloor\} \frac{I\{Z_{n,j} \le z\} - F_2(z)}{n}.$$

Observing (A.4) and (A.5) it then follows from Bücher (2014) that

$$\Big\{ \sqrt{n}\, \mathbb{X}_n^{(\ell)}(s,z) \Big\}_{s\in[0,1], z\in\mathbb{R}^d} \stackrel{\mathcal{D}}{\Longrightarrow} \Big\{ \mathbb{G}^{(\ell)}(s,z) \Big\}_{s\in[0,1], z\in\mathbb{R}^d},$$

where $\mathbb{G}^{(1)}$ and $\mathbb{G}^{(2)}$ are two centered independent Gaussian processes with covariance structure

$$\text{(A.11)}\qquad \mathbb{E}[\mathbb{G}^{(\ell)}(s_1,z_1)\mathbb{G}^{(\ell)}(s_2,z_2)] = \begin{cases} (s_1 \wedge s_2 \wedge t) k_1(z_1,z_2) & \text{if } \ell = 1 \\ (s_1 \wedge s_2 - t)_+ k_2(z_1,z_2) & \text{if } \ell = 2 \end{cases}.$$

Consequently, the processes $\sqrt{n}\, \mathbb{X}_n^{(1)}$, $\sqrt{n}\, \mathbb{X}_n^{(2)}$ and its sum $\sqrt{n}\, \mathbb{X}_n = \sqrt{n}\, (\mathbb{X}_n^{(1)} + \mathbb{X}_n^{(2)})$ are asymptotically tight [see Section 1.5 in van der Vaart and Wellner (1996)], and in order to prove weak convergence of the process $\sqrt{n}\, \mathbb{X}_n$ it remains to establish the weak convergence of the finite dimensional distributions. For this purpose we use the Crámer-Wold device and show for all $(s_1,z_1),\dots,(s_k,z_k) \in [0,1]\times\mathbb{R}^d,\ \alpha_1,\dots,\alpha_k \in \mathbb{R}$

$$\text{(A.12)}\qquad \sqrt{n}\Big\{ \sum_{j=1}^k \alpha_j \mathbb{X}_n(s_j,z_j) \Big\} \stackrel{\mathcal{D}}{\Longrightarrow} \sum_{j=1}^k \alpha_j \mathbb{G}_{F_1,F_2,t}(s_j,z_j),$$

where $\mathbb{G}_{F_1,F_2,t}$ is the Gaussian process defined in Theorem A.1. For the sake of a clear exposition we restrict ourselves to the case $k=2$ and begin with a calculation of the covariance of $X_n^{(1)}(s_1,z_1)$ and $X_n^{(2)}(s_2,z_2)$. If $s_1 \le s_2 \le t$ we can use the same arguments as in Bücher (2014) and obtain

$$n \operatorname{Cov}\Big(\mathbb{X}_n^{(\ell)}(s_1,z_1), \mathbb{X}_n^{(\ell)}(s_2,z_2) \Big) \stackrel{n\to\infty}{\longrightarrow} \begin{cases} (s_1 \wedge s_2 \wedge t) k_1(z_1,z_2) & \text{if } \ell = 1 \\ 0 & \text{if } \ell = 2 \end{cases}.$$

Similarly, if $t \le s_1 \le s_2 \le 1$ we have

$$n \operatorname{Cov}\Big(\mathbb{X}_n^{(\ell)}(s_1,z_1), \mathbb{X}_n^{(\ell)}(s_2,z_2) \Big) \stackrel{n\to\infty}{\longrightarrow} \begin{cases} t k_1(z_1,z_2) & \text{if } \ell = 1 \\ (s_1 \wedge s_2 - t)_+ k_2(z_1,z_2) & \text{if } \ell = 2 \end{cases}.$$

Finally, if $s_1 < t \le s_2$ we have by assumption (A.3)

$$n\Big| \operatorname{Cov}(\mathbb{X}_n^{(1)}(s_1,z_1), \mathbb{X}_n^{(2)}(s_2,z_2)) \Big| \ = \ n\Big| \operatorname{Cov}\Big(\sum_{j=1}^{\lfloor ns_1 \rfloor} Y_{n,j}(z_1), \sum_{j=\lfloor nt \rfloor+1}^{\lfloor ns_2 \rfloor} Y_{nj}(z_2) \Big) \Big| = O(\frac{1}{n^\eta}) = o(1),$$

where the random variables $Y_{n,j}$ are defined in (A.10). If $s_1 = t \le s_2$ we use a sequence ε_n satisfying $\varepsilon_n n \to \infty$ and $n\varepsilon_n^2 \to 0$ and obtain by the same arguments

$$\begin{aligned} & n\Big| \operatorname{Cov}(\mathbb{X}_n^{(1)}(t_1,z_1), \mathbb{X}_n^{(2)}(s_2,z_2)) \Big| \\ &= n\Big| \operatorname{Cov}\Big(\sum_{j=1}^{\lfloor n(t-\varepsilon_n) \rfloor} Y_{n,j} + \sum_{j=\lfloor n(t-\varepsilon_n) \rfloor+1}^{\lfloor nt \rfloor} Y_{n,j}, \sum_{j=\lfloor nt \rfloor+1}^{\lfloor n(t+\varepsilon_n) \rfloor} Y_{n,j} + \sum_{j=\lfloor n(t+\varepsilon_n) Y_{n,j} \rfloor+1}^{\lfloor ns_2 \rfloor} \Big) \Big| \\ &= O(\frac{1}{(n\varepsilon_n)^\eta}) + O(n\varepsilon_n^2) = o(1). \end{aligned}$$

Using similar arguments for the remaining cases it follows from assumptions (A.4) and (A.5) that

(A.13)

$$
\begin{aligned}
\sigma^2 &= \lim_{n\to\infty} \operatorname{Var}(\sqrt{n}\sum_{j=1}^{2} \alpha_j \mathbb{X}_n(s_j, z_j)) \\
&= \lim_{n\to\infty} n\Big\{\alpha_1^2 \mathrm{Cov}(\mathbb{X}_n^{(1)}(s_1,z_1), \mathbb{X}_n^{(1)}(s_1,z_1)) + 2\alpha_1\alpha_2 \mathrm{Cov}(\mathbb{X}_n^{(1)}(s_1,z_1), \mathbb{X}_n^{(1)}(s_2,z_2)) \\
&\quad +\alpha_1^2 \mathrm{Cov}(\mathbb{X}_n^{(2)}(s_1,z_1), \mathbb{X}_n^{(2)}(s_1,z_1)) + 2\alpha_1\alpha_2 \mathrm{Cov}(\mathbb{X}_n^{(2)}(s_1,z_1), \mathbb{X}_n^{(2)}(s_2,z_2)) \\
&\quad +\alpha_2^2 \mathrm{Cov}(\mathbb{X}_n^{(1)}(s_2,z_2), \mathbb{X}_n^{(1)}(s_2,z_2)) + \alpha_2^2 \mathrm{Cov}(\mathbb{X}_n^{(2)}(s_2,z_2), \mathbb{X}_n^{(2)}(s_2,z_2))\Big\} \\
&= \alpha_1^2\Big\{(s_1\wedge t)k_1(z_1,z_1) + (s_1-t)_+ k_2(z_1,z_1)\Big\} + \alpha_2^2\Big\{(s_2\wedge t)k_1(z_2,z_2) + (s_2-t)_+ k_2(z_2,z_2)\Big\} \\
&\quad +2\alpha_1\alpha_2\Big\{(s_1\wedge s_2\wedge t)k_1(z_1,z_2) + (s_1\wedge s_2 - t)_+ k_2(z_1,z_2)\Big\} \\
&= \operatorname{Var}\Big(\alpha_1 \mathbb{G}_{F_1,F_2,t}(s_1,z_1) + \alpha_2 \mathbb{G}_{F_1,F_2,t}(s_2,z_2)\Big),
\end{aligned}
$$

where $\mathbb{G}_{F_1,F_2,t}$ denotes the centered Gaussian process defined in Theorem A.1.
In order to prove asymptotic normality of the statistic $\sqrt{n}\sum_{j=1}^{2}\alpha_j \mathbb{X}_n(s_j,z_j)$ we introduce the notation

$$
T_n = \frac{\sqrt{n}}{\sigma}\sum_{j=1}^{2}\alpha_j \mathbb{X}_n(s_j,z_j) = \sum_{j=1}^{n} S_{n,j} + o_p(1),
$$

where the random variably $S_{n,j}$ are defined by

$$
\begin{aligned}
S_{n,j} &= \frac{\alpha_1 I\{j \le \lfloor ns_1\rfloor\}}{\sigma\sqrt{n}}(I\{Z_{n,j}\le z_1\} - E_{F_1,F_2,t}(s_1,z_1)) \\
&\quad + \frac{\alpha_2 I\{j \le \lfloor ns_2\rfloor\}}{\sigma\sqrt{n}}(I\{Z_{n,j}\le z_2\} - E_{F_1,F_2,t}(s_2,z_2)),
\end{aligned}
$$

and we use a central limit theorem for triangular arrays of strong mixing random variables [see Theorem 2.1 in Liebscher (1996), where $p=\infty$]. For this purpose we note that it follows from the discussion in the previous paragraph that $\lim_{n\to\infty}\mathbb{E}[T_n^2]=1$ and that it is easy to see that $\lim_{n\to\infty}\sum_{j=1}^{n}\big(\operatorname{ess\,sup}_{w\in\Omega}[|S_{n,j}| I\{|S_{n,j}| > \varepsilon\}]\big)^2 = 0$ (a.s.). Similarly, it follows that the condition $\sum_{j=1}^{n}\big(\operatorname{ess\,sup}_{w\in\Omega}|S_{n,j}|\big)^2 \le$ const(a.s.) of Theorem 2.1 in Liebscher (1996) is also satisfied. Therefore this result shows that

$$
\sqrt{n}\sum_{j=1}^{2}\alpha_j X_n(s_j,z_j) = \frac{\sigma\, T_n}{\sqrt{\mathbb{E}[T_n^2]}} \stackrel{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0,\sigma^2),
$$

where the asymptotic variance σ^2 is defined in (A.13). This proves the convergence of the finite dimensional distributions and completes the proof of Theorem A.1. □

As pointed out in Section 3 , there exist many cases where assumption (3.17) (as established by Theorem A.1 for strong mixing triangular arrays) is not satisfied. In this case it is necessary to prove (3.16) for the specific functional under consideration. A general statement can be obtained if the functional of interest is linear. The proof is obtained by similar arguments as given for Theorem A.1 and therefore omitted.

Theorem A.2 *Assume that the conditions of Theorem A.1 are satisfied and that the functional in* (3.2) *is linear and* $\mathcal{S} \subset \mathbb{R}^k$*. Then a standardized version of the process* $\{\hat{\mathbb{U}}_n(s)\}_{s\in[0,1]}$ *defined by* $\hat{\mathbb{U}}_n(s) = \theta_{\text{lin}}(\hat{\mathbb{F}}_n(s,\cdot) - s\hat{\mathbb{F}}_n(1,\cdot))$ *converges weakly in* $\ell([0,1]|\mathbb{R}^k)$*, that is*

$$\Big\{\sqrt{n}(\hat{\mathbb{U}}_n(s) - \mathbb{U}(s))\Big\}_{s\in[0,1]} \stackrel{\mathcal{D}}{\Longrightarrow} \Big\{\mathbb{D}_{F_1,F_2,t}(s)\Big\}_{s\in[0,1]}.$$

Here $\mathbb{D}_{F_1,F_2,t}$ *denotes a centered Gaussian process with covariance kernel*

$$\begin{aligned} \text{(A.14)} \qquad d_{F_1,F_2,t}(s_1,s_2) &= \mathbb{E}[\mathbb{D}_{F_1,F_2,t}(s_1)\mathbb{D}^T_{F_1,F_2,t}(s_2)] \\ &= \{(s_1\wedge s_2\wedge t) + s_1s_2t - s_2(s_1\wedge t) - s_1(s_2\wedge t)\}V_1 \\ &+ \{(s_1\wedge s_2 - t)_+ + s_1s_2(1-t) - s_1(s_2-t)_+ - s_2(s_1-t)_+\}V_2 \end{aligned}$$

and the matrices $V_1, V_2 \in \mathbb{R}^{k\times k}$ *are defined by*

$$\text{(A.15)} \qquad V_\ell = \sum_{k\in\mathbb{Z}} \text{Cov}(\theta_{\text{lin}}(I\{W_0(\ell) \le \cdot\}, \theta_{\text{lin}}(I\{W_k(\ell) \le \cdot\}), \quad \ell = 1,2.$$

B Details for Section 4

B.1 Linear Regression

Due to the nonlinearity of the representation (4.5), it is more difficult to derive a representation of the form (3.11). For this purpose consider the functional

$$\Phi(F)(s) := \Big(\int_{\mathbb{R}^{d+1}} g(x)g^T(x)F(1,dx,dy)\Big)^{-1}\int_{\mathbb{R}^{d+1}} yg(x)\Big(F(s,dx,dy) - sF(1,dx,dy)\Big).$$

defined on the set $\mathcal{F} \subset \ell^\infty([0,1]\times\mathbb{R}^{d+1}|\mathbb{R})$ of all bounded functions F for which the integrals exist (for each $s \in [0,1]$) and which satisfy $|\int_{\mathbb{R}^{d+1}} g(x)g^T(x)F(1,dx,dy)| \neq 0$. The analog of the quantity (3.6) is given by

$$E_{F_1,F_2,t}(s,y,x) = (s\wedge t)F_1(x,y) + (s-t)_+\ F_2(x,y),$$

and it follows by a straightforward calculation that

$$\begin{aligned} \text{(A.16)} \quad \mathbb{U}_{\text{lin}}(s) := \Phi(E_{F_1,F_2,t})(s) &= \Big(\int_{\mathbb{R}^d} g(x)g^T(x)F_X(dx)\Big)^{-1} \\ &\quad\times\Big\{(s\wedge t - st)\int_{\mathbb{R}^{d+1}} yg(x)(F_1(dx,dy) - F_2(dx,dy))\Big\} \\ &= (s\wedge t - st)(\beta_1-\beta_2) = (s\wedge t - st)(\theta(F_1) - \theta(F_2)), \end{aligned}$$

where we used (4.4) and the representation (4.5).
Assume for a moment that the matrix B in (4.4) is known, then we obtain from Theorem A.2 that the process

$$\tilde{\mathbb{U}}_n(s) = \int_{\mathbb{R}^{d+1}} yg(x)(\hat{\mathbb{F}}_n(s,dx,dy) - s\hat{\mathbb{F}}_n(1,dx,dy))$$

converges weakly to a centered Gaussian process, that is

$$\left\{\sqrt{n}\ (\tilde{\mathbb{U}}_n(s) - \tilde{\mathbb{U}}(s))\right\}_{s\in[0,1]} \overset{\mathcal{D}}{\Longrightarrow} \left\{\tilde{\mathbb{D}}_{F_1,F_2,t}(s)\right\}_{s\in[0,1]}$$

where $\tilde{\mathbb{U}} = B\mathbb{U}_{\text{lin}}$, $\mathbb{U}_{\text{lin}}$ is defined in (A.16) and $\tilde{\mathbb{D}}_{F_1,F_2,t}$ is a centered Gaussian process with covariance kernel defined by (A.14) and (A.15), where

(A.17) $$\theta(I\{W_k(\ell) \le \cdot\}) = W_{k,2}(\ell)g(W_{k,1}(\ell))\ ; \ell = 1,2,$$

and $W_{k,1}(\ell) \in \mathbb{R}^d$, $W_{k,2}(\ell) \in \mathbb{R}$ denote the components of the process $\{W_k(\ell)\}_{k\in\mathbb{Z}}$ considered in Theorem A.1, that is $W_k(\ell) = (W_{k,1}(\ell), W_{k,2}(\ell))^T \quad (k \in \mathbb{Z},\ \ell = 1,2)$. However, the estimation of the matrix B yields an additional effect, which makes the asymptotic analysis of the process $\hat{\mathbb{U}}_n$ substantially more complicated. It is already visible in the decomposition

$$\sqrt{n}\ (\hat{\mathbb{U}}_n(s) - \mathbb{U}_{\text{lin}}(s)) = \sqrt{n}\ \hat{B}_n^{-1}(\tilde{\mathbb{U}}_n(s) - \tilde{\mathbb{U}}(s)) + \sqrt{n}\ (\hat{B}_n^{-1} - B^{-1})\tilde{\mathbb{U}}(s),$$

where $\hat{B}_n = \frac{1}{n}\sum_{i=1}^n g(X_i)g^T(X_i) = \int_{\mathbb{R}^{d+1}} g(x)g^T(x)\hat{\mathbb{F}}_n(1,dx,\infty)$ denotes the common estimate of the matrix B defined in (4.4). In order to explain this effect in more detail we restrict ourselves to one-dimensional models. The general case can be treated exactly in the same way with some extra matrix algebra. To be precise, define the empirical process $\hat{\mathbb{H}}_n(s) = (\hat{B}_n, \tilde{\mathbb{F}}_n(s))^T$, where

$$\hat{B}_n = \frac{1}{n}\sum_{i=1}^n g^2(X_i)\ ; \qquad \tilde{\mathbb{F}}_n(s) = \frac{1}{n}\sum_{i=1}^{\lfloor ns\rfloor} g(X_i)Y_{n,i}.$$

Similar arguments as given in Section A.1 show

(A.18) $$\left\{\sqrt{n}\ (\mathbb{H}_n(s) - (B, \tilde{E}_{F_1,F_2,t}(s))^T)\right\}_{s\in[0,1]} \overset{\mathcal{D}}{\Longrightarrow} \left\{\mathbb{H}(s)\right\}_{s\in[0,1]}$$

where $B = \int_{\mathbb{R}} g^2(x)F_X(dx)$, $\tilde{E}_{F_1,F_2,t}(s) = (s\wedge t)B\beta_1 + (s-t)_+B\beta_2$ and $\mathbb{H}$ denotes a two-dimensional centered Gaussian process with covariance matrix

$$\text{Cov}(\mathbb{H}(s_1),\mathbb{H}(s_2)) = \begin{pmatrix} V_0 & V_0((s_2\wedge t)\beta_1 + (s_2-t)_+\beta_2) \\ V_0((s_2\wedge t)\beta_1 + (s_2-t)_+\beta_2) & (s_1\wedge s_2\wedge t)V_{0,1} + (s_1\wedge s_2 - t)_+V_{0,1} \end{pmatrix},$$

where $V_{0,1} = (V_0\beta_1^2 + V_1)$ and the matrices V_0 and V_1 are defined by

$$V_0 = \sum_{k\in\mathbb{Z}} \text{Cov}(g^2(X_0), g^2(X_k)), \qquad V_1 = \sum_{k\in\mathbb{Z}} \text{Cov}(g(X_0)\varepsilon_0, g(X_k)\varepsilon_k).$$

Now an application of the functional Delta-method [see van der Vaart and Wellner (1996)] yields

$$\Big\{\sqrt{n}\ \hat{\mathbb{U}}_n(s)-\mathbb{U}_{\text{lin}}(s)\Big\}_{s\in[0,1]} \overset{\mathcal{D}}{\Longrightarrow} \Big\{B^{-1}\big(\mathbb{H}_2(s)-s\mathbb{H}_2(1)-(s\wedge t-st)\mathbb{H}_1(1)(\beta_1-\beta_2)\big)\Big\}_{s\in[0,1]},$$

where $\mathbb{H}_1$ and $\mathbb{H}_2$ denote the components of the limiting process in (A.18). A tedious calculation yields for the covariance structure of this process

$$\begin{aligned}
k(s_1,s_2) \;=\; & \frac{V_0}{B^2}\Big[\beta_1^2\big\{(s_1\wedge s_2\wedge t)-s_1(s_2\wedge t)-s_2(s_1\wedge t)+s_1s_2t-(s_1\wedge t-s_1t)(s_2\wedge t-s_2t)\big\} \\
&+\;\beta_2^2\big\{(s_1\wedge s_2-t)_+-s_1(s_2-t)_+-s_2(s_1-t)_++s_1s_2(1-t)+(s_1\wedge t-s_1t)(s_2\wedge t-s_2t) \\
&\quad+(s_2\wedge t-s_2t)((s_1-t)_+-s_1(1-t))+(s_1\wedge t-s_1t)((s_2-t)_+-s_2(1-t))\big\} \\
&+\;\beta_1\beta_2\big\{(s_1\wedge t-s_1t)(s_2(1-t)-(s_2-t)_+)+(s_2\wedge t-s_2t)(s_1(1-t)-(s_1-t)_+)\big\}\Big] \\
&+\;\frac{V_1}{B^2}\Big[(s_1\wedge s_2\wedge t)+(s_1\wedge s_2-t)_+-s_2(s_1\wedge t+(s_1-t)_+) \\
&\quad-s_1(s_2\wedge t+(s_2-t)_+)+(s_1\wedge t-s_1t)(s_2\wedge t-s_2t)\Big].
\end{aligned}$$

Observing (A.16) we have by similar arguments as given in Section 2 that

$$\frac{3\sqrt{n}}{t^2(1-t)^2}\Big(\int_0^1 \hat{\mathbb{U}}_n^2(s)ds-(\beta_1-\beta_2)^2\Big) \overset{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0,\tau^2_{F_1,F_2,t}),$$

where the asymptotic variance is given by (4.6). From these considerations a test for the hypothesis of a relevant change in the parameters of the linear regression model can easily be constructed as indicated in Section 2 and 3 with (4.6).

B.2 Distribution Function

In order to estimate the distance described in (4.8) we define $\mathcal{F}\subset\ell^\infty([0,1]\times\mathbb{R}|\mathbb{R})$ as the set of all functions F, such that for each $s\in[0,1]$ the integral

$$\text{(A.19)}\qquad \int_{\mathbb{R}}(F(s,z)-sF(1,z))^2dz$$

exists (for all $s\in[0,1]$). Note that this set contains the set of all functions of the form $E_{F_1,F_2,t}$ defined in (3.6), such that F_1 and F_2 have moments of order one [Székely and Rizzo (2005), p. 73]. We consider the functional $\Phi_{\text{non}}:\mathcal{F}\to\ell^\infty([0,1]|\mathbb{R})$ defined by $\Phi_{\text{non}}(F)(s)=F(s,\cdot)-sF(1,\cdot)$, then

$$\text{(A.20)}\qquad \Phi_{\text{non}}(E_{F_1,F_2,t})(s)=(s\wedge t-st)(F_1-F_2)(\cdot)=:Z_{F_1,F_2,t}(s,\cdot).$$

In this case it follows from Theorem A.1 that assumption (3.17) is satisfied, and as a consequence we obtain

$$\text{(A.21)}\qquad \Big\{\sqrt{n}(\hat{\mathbb{Z}}_n(s,z)-Z_{F_1,F_2,t}(s,z))\Big\}_{s\in[0,1],z\in\mathbb{R}^d} \overset{\mathcal{D}}{\Longrightarrow} \Big\{\mathbb{H}_{F_1,F_2,t}(s,z)\Big\}_{s\in[0,1],z\in\mathbb{R}^d},$$

where the limiting process $\mathbb{H}$ is defined by $\mathbb{H}_{F_1,F_2,t}(s,z) = \mathbb{G}_{F_1,F_2,t}(s,z) - s\ \mathbb{G}_{F_1,F_2,t}(1,z)$, and the covariance kernel of this process is given by

$$\begin{aligned} h_{F_1,F_2,t}(s_1,z_1,s_2,z_2) &= \{(s_1 \wedge s_2 \wedge t) + s_1 s_2 t - s_2(s_1 \wedge t) - s_1(s_2 \wedge t)\}k_1(z_1,z_2) \\ &+ \{(s_1 \wedge s_2 - t)_+ + s_1 s_2(1-t) - s_1(s_2 - t)_+ - s_2(s_1 - t)_+\}k_2(z_1,z_2). \end{aligned}$$

Defining $\mathbb{T}_n(s) = \|\hat{\mathbb{Z}}_n(s,\cdot)\|, \mathbb{T}(s) = \|\mathbb{Z}(s,\cdot)\|$ and observing the representation

$$\begin{aligned} \int_0^1 \mathbb{T}_n^2(s)ds - \int_0^1 \mathbb{T}^2(s)ds &= \int_0^1 (\|\hat{\mathbb{Z}}_n(s,\cdot)\|^2 - \|Z_{F_1,F_2,t}(s,\cdot)\|^2)ds \\ &= 2\int_0^1 \int_{\mathbb{R}} Z_{F_1,F_2,t}(s,z)(\hat{\mathbb{Z}}_n(s,z) - Z_{F_1,F_2,t}(s,z))dzds + o_p\left(\frac{1}{\sqrt{n}}\right) \end{aligned}$$

we have

$$\text{(A.22)} \qquad \sqrt{n}\Big(\int_0^1 \mathbb{T}_n^2(s)ds - \int_0^1 \mathbb{T}^2(s)ds\Big) \overset{\mathcal{D}}{\Longrightarrow} 2\int_0^1 \int_{\mathbb{R}} Z_{F_1,F_2,t}(s,z)\mathbb{H}_{F_1,F_2,t}(s,z)dzds\ ,$$

where $\mathbb{H}_{F_1,F_2,t}$ is the limiting process defined in (A.21). Note that the right hand side of (A.22) is normal distributed with variance

$$\begin{aligned} \sigma^2_{F_1,F_2,t} &= \frac{4}{45}(t^2(1-t)^2)\Big[t(5-10t+6t^2)\int_{\mathbb{R}^2}(F_1(z_1)-F_2(z_1))(F_1(z_2)-F_2(z_2))k_1(z_1,z_2)dz_1dz_2 \\ &+(1-3t+8t^2-6t^3)\int_{\mathbb{R}^2}(F_1(z_1)-F_2(z_1))(F_1(z_2)-F_2(z_2))k_2(z_1,z_2)dz_1dz_2\Big]\ . \end{aligned}$$

Consequently, we obtain

$$\sqrt{n}\ \Big(\int_0^1 \mathbb{T}_n^2(s)ds - \frac{t^2(1-t)^2}{3}\ \|F_1 - F_2\|\Big) \overset{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0, \sigma^2_{F_1,F_2,t}),$$

and the test for the hypotheses of relevant change-points can be constructed following the arguments given in Section 3 with $\tau^2_{F_1,F_2,t} = \frac{9\sigma^2_{F_1,F_2,t}}{t^4(1-t)^4}$.

C Further Examples

C.1 Relevant Changes in the Variance

Following Aue et al. (2009) we consider a triangular array of d-dimensional random variables $(Z_{n,t})_{t=1}^n$ with constant mean and investigate the problem of detecting a relevant change in the variance. This means that the functional of interest is given by

$$\theta_{\mathrm{var}}(F) = \int_{\mathbb{R}^d} zz^T F(dz) - \int_{\mathbb{R}^d} zF(dz)\int_{\mathbb{R}^d} z^T F(dz),$$

where F is the distribution function. Note that this functional is not linear. However, if $\mu = \mathbb{E}[Z_{n,1}] = \ldots = \mathbb{E}[Z_{n,t}]$, $\Sigma_1 = \mathrm{Var}(Z_{n,1})$ and $\Sigma_2 = \mathrm{Var}(Z_{n,n})$ denote the common mean and the variance before and after the break, respectively, a straightforward calculation yields the representation

$$\begin{aligned}\mathbb{U}_{\mathrm{var}}(s) &= \Phi_{\mathrm{var}}(E_{F_1,F_2,t})(s) := \theta_{\mathrm{var}}(E_{F_1,F_2,t}(s,\cdot) - sE_{F_1,F_2,t}(1,\cdot)) \\ &= \tilde{\theta}_{\mathrm{var}}(E_{F_1,F_2,t}(s,\cdot) - sE_{F_1,F_2,t}(1,\cdot)) = (s \wedge t - st)(\Sigma_1 - \Sigma_2),\end{aligned}$$

where $\tilde{\theta}_{\mathrm{var}}(F) = \int_{\mathbb{R}^d} zz^T F(dz)$. Consequently, assumption (3.11) is satisfied and we obtain from Theorem A.2 the weak convergence of the process

$$\hat{\mathbb{U}}_n(s) = \frac{1-s}{n} \sum_{j=1}^{\lfloor ns \rfloor} Z_{n,j} Z_{n,j}^T - \frac{s}{n} \sum_{j=\lfloor ns \rfloor+1}^{n} Z_{n,j} Z_{n,j}^T$$

to a centered Gaussian process with covariance kernel (A.14) and (A.15), where

$$\theta_{\mathrm{lin}}(I\{W_k(\ell) \le \cdot\}) = \tilde{\theta}_{\mathrm{var}}(I\{W_k(\ell) \le \cdot\}) = W_k(\ell) W_k^T(\ell).$$

A straightforward but tedious calculation shows that the limiting variance in (3.20) is given by

$$\begin{aligned}\tau^2_{F_1,F_2,t} &= \frac{4}{5(t(1-t))^2} \mathrm{tr}\Big\{ \big[(t(5 - 10t + 6t^2) V_1^{\mathrm{var}} \\ &\quad + (1 - 3t + 8t^2 - 6t^3) V_2^{\mathrm{var}} \big] (\Sigma_1 - \Sigma_2)(\Sigma_1 - \Sigma_2)^T \Big\},\end{aligned}$$

where V_1^{var} and V_2^{var} are defined in Theorem A.2 with $\theta_{\mathrm{lin}}(I\{W_k(\ell) \le \cdot\}) = W_k(\ell)W_k^T(\ell) \quad (\ell = 1,2)$.

C.2 Relevant Changes in the Correlation

Let $Z_{n,1}, \ldots, Z_{n,n} = (X_{n,1}, Y_{n,1}), \ldots, (X_{n,n}, Y_{n,n})$ denote two-dimensional random variables satisfying (A.2). Following Wied et al. (2012) we assume that $\mathbb{E}[X_i] = \mu_1,\ \mathbb{E}[X_i^2] = \mu_2,\ \mathbb{E}[Y_i] = \nu_1,\ \mathbb{E}[Y_i^2] = \nu_2$ and we are interested in a relevant change-point in the correlation, that is

$$H_0 : |\rho_1 - \rho_2| \le \Delta \quad \text{versus} \quad H_1 : |\rho_1 - \rho_2| > \Delta,$$

where

$$\rho_i = \theta(F_i) = \frac{\int_{\mathbb{R}^2} (x - \int_{\mathbb{R}^2} u F_i(du,dv))(y - \int_{\mathbb{R}} v F_i(du,dv)) F_i(dx,dy)}{\big\{ \int_{\mathbb{R}^2} (x - \int_{\mathbb{R}^2} u F_i(du,dv))^2 F_i(dx,dy) \int_{\mathbb{R}^2} (y - \int_{\mathbb{R}^2} v F_i(du,dv))^2 F_i(dx,dy) \big\}^{1/2}}$$

denotes the correlation of the distribution function F_i $(i = 1,2)$. Consider the functional

$$\Phi_{\mathrm{corr}}(F)(s) = \frac{\int_{\mathbb{R}^2} xy(F(s,dx,dy) - sF(1,dx,dy))}{\big\{ \int_{\mathbb{R}^2} (x - \int_{\mathbb{R}^2} uF(1,du,dv))^2 F(1,dx,dy) \int_{\mathbb{R}^2} (y - \int_{\mathbb{R}^2} vF(1,du,dv))^2 F(1,dx,dy) \big\}^{1/2}}$$

defined on the set $\mathcal{F} \subset \ell^\infty([0,1] \times \mathbb{R}^2|\mathbb{R})$ of functions such that all integrals exist. If F_1 and F_2 denote the distributions of the two-dimensional vector (X, Y) before and after the change-point and

$$E_{F_1,F_2,t}(s,x,y) = (s \wedge t)F_1(x,y) + (s-t)_+ \; F_2(x,y),$$

then a straightforward but tedious calculation yields a representation of the form (3.11), that is

$$\mathbb{U}_{\mathrm{corr}}(s) \;:=\; \Phi_{\mathrm{corr}}(E_{F_1,F_2,t})(s) = \frac{(s \wedge t - st)\int_{\mathbb{R}^2} xy(F_1 - F_2)(dx,dy)}{\left\{(\mu_2 - \mu_1^2)(\nu_2 - \nu_1^2)\right\}^{1/2}} = (s \wedge t - st)(\theta(F_1) - \theta(F_2)).$$

Recall the definition $\hat{\mathbb{F}}_n(s,x,y) = \frac{1}{n}\sum_{i=1}^{\lfloor ns \rfloor} I\{X_{n,i} \le x, Y_{n,i} \le y\}$, and define $\hat{\mathbb{U}}_n(s) = \Phi_{\mathrm{corr}}(\hat{\mathbb{F}}_n)(s)$, then it follows that

$$\hat{\mathbb{U}}_n(s) = \frac{\overline{\mathbb{F}}_n(s) - s\overline{\mathbb{F}}_n(1)}{\sqrt{(\hat{\mu}_2 - (\hat{\mu}_1)^2)(\hat{\nu}_2^2 - (\hat{\nu}_1)^2)}}, \tag{A.23}$$

where $\overline{\mathbb{F}}_n(s) = \frac{1}{n}\sum_{j=1}^{\lfloor ns \rfloor} X_{n,j}Y_{n,j}$, and $\hat{\mu}_i = \frac{1}{n}\sum_{j=1}^{n} X_{n,j}^i$; $\hat{\nu}_i = \frac{1}{n}\sum_{j=1}^{n} Y_{n,j}^i (i = 1,2)$ denote the common estimators of the moments $\mathbb{E}[X^i]$ and $\mathbb{E}[Y^i]$ $(i = 1,2)$. Similar arguments as given in Section A.1 yield

$$\left\{\sqrt{n}\,\left((\hat{\mu}_1, \hat{\mu}_2, \hat{\nu}_1, \hat{\nu}_2, \overline{\mathbb{F}}_n(s))^T - (\mu_1, \mu_2, \nu_1, \nu_2, E_{F_1,F_2,t})^T\right)\right\}_{s\in[0,1]} \overset{\mathcal{D}}{\Longrightarrow} \left\{\mathbb{H}(s)\right\}_{s\in[0,1]}$$

where $\{\mathbb{H}(s)\}_{s\in[0,1]}$ is a centered Gaussian process. Now weak convergence of the process $\{\hat{\mathbb{U}}_n(s)\}_{s\in[0,1]}$ follows by the functional Delta-method and a tedious calculation. The details are omitted for the sake of brevity.

References

Aue, A., Hörmann, S., Horváth, L., and Reimherr, M. (2009). Break detection in the covariance structure of multivariate time series models. *Annals of Statistics*, 37(6B):4046–4087.

Berkes, I., Hörmann, S., and Schauer, J. (2009). Asymptotic results for the empirical process of stationary sequences. *Stochastic Processes and their Applications*, 119(4):1298–1324.

Bücher, A. (2014). A note on weak convergence of the sequential multivariate empirical process under strong mixing. *Journal of Theoretical Probability, to appear.*

Dehling, H., Durieu, O., and Tusche, M. (2013). A sequential empirical central limit theorem for multiple mixing processes with application to B-geometrically ergodic Markov chains. *arXiv:1303.4537.*

Deo, C. M. (1973). A note on empirical processes of strong-mixing sequences. *The Annals of Probability*, 1(5):870–875.

Liebscher, E. (1996). Central limit theorems for the sums of α-mixing random variables. *Stochastics and Stochastic Reports*, 59(3-4):241–258.

Székely, G. J. and Rizzo, M. L. (2005). A new test for multivariate normality. *Journal of Multivariate Analysis*, 93(1):58–80.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes. Springer Series in Statistics*. Springer, New York.

Wied, D., Krämer, W., and Dehling, H. (2012). Testing for a change in correlation at an unknown point in time using an extended functional delta method. *Econometric Theory*, 28(3):570–589.