A new approach for open-end sequential change point monitoring

Josua Gösmann

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

josua.goesmann@ruhr-uni-bochum.de

(corresponding author)

Tobias Kley

University of Bristol

School of Mathematics

Bristol BS8 1TH, United Kingdom

tobias.kley@bristol.ac.uk

Holger Dette

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

holger.dette@ruhr-uni-bochum.de

June 3, 2019

Abstract

We propose a new sequential monitoring scheme for changes in the parameters of a multivariate time series. In contrast to procedures proposed in the literature which compare an estimator from the training sample with an estimator calculated from the remaining data, we suggest to divide the sample at each time point after the training sample. Estimators from the sample before and after all separation points are then continuously compared calculating a maximum of norms of their differences. For open-end scenarios our approach yields an asymptotic level α procedure, which is consistent under the alternative of a change in the parameter.

1 Introduction

Nowadays, nearly all fields of applications require sophisticated statistical modelling and statistical inference to draw scientific conclusions from the observed data. In many cases data is time dependent and the involved model parameters or the model itself may not be necessarily stable. In such situations it is of particular importance to detect changes in the processed data as soon as possible and to adapt the statistical analysis accordingly. These changes are usually called *change points* or *structural breaks* in the literature. Due to its universality, methods for change point analysis have a vast field of possible applications - ranging from natural sciences [for example biology and meteorology] to humanities [economics, finance, social sciences]. Since the seminal papers of Page (1954, 1955) the problem of detecting change points in time series has received substantial attention in the statistical literature. The contributions to this field can be roughly divided into the areas of *retrospective* and *sequential* change point analysis.

In the retrospective case, historical data sets are examined with the aim to test for changes and identify their position within the data. In this setup, the data is assumed to be completely available before the statistical analysis is started (a-posteriori analysis). A comprehensive overview of retrospective change point analysis can be found in Aue and Horváth (2013). In many practical applications, however, data arrives consecutively and breaks can occur at any new data point. In such cases the statistical analysis for changes in the processed data has to start immediately with the target to detect changes as soon as possible. This field of statistics is called sequential change point detection [sometimes also: online change point detection].

In the major part of the 20th century the problem of sequential change point detection was tackled using procedures [mostly called control charts, see Lai (1995, 2001) for comprehensive reviews], which are optimized to have a minimal detection delay but do not control the probability of a false alarm (type I error). A new paradigm was then introduced by Chu et al. (1996), who use initial data sets and therefrom employ invariance principles to also control the type I error. The methods developed under this paradigm [see below] can again be subdivided into *closed-end* and *open-end* approaches. In closed-end scenarios monitoring is stopped at a fixed pre-defined point of time, while in open-end scenarios monitoring can - in principle - continue forever if no change point is detected.

In the paper at hand we develop a new approach for sequential change point detection in an open-end scenario. To be more precise let $\{X_t\}_{t\in\mathbb{Z}}$ denote a d-dimensional time series and let F_t be the distribution function of the random variable X_t at time t. We are studying monitoring procedures for detecting changes of a parameter $\theta_t = \theta(F_t)$, where $\theta = \theta(F)$ is a p-dimensional parameter of a distribution function F on $\mathbb{R}^d$ (such as the mean, variance, correlation etc.). In particular we will develop a decision rule for the hypothesis

of a constant parameter, that is

$$H_0: \ \theta_1 = \cdots = \theta_m = \theta_{m+1} = \theta_{m+2} = \dots \ , \tag{1.1}$$

against the alternative that the parameter changes (once) at some time $m+k^\star$ with $k^\star \geq 1$, that is

$$H_1: \ \exists k^\star \in \mathbb{N}: \ \ \theta_1 = \cdots = \theta_{m+k^\star-1} \neq \theta_{m+k^\star} = \theta_{m+k^\star+1} = \dots. \tag{1.2}$$

In this setup, which was originally introduced by Chu et al. (1996), the first m observations are assumed to be stable and will serve as an initial training set. The problem of sequential change point detection in the hypotheses paradigm as pictured above has received substantial interest in the literature. Since the seminal paper of Chu et al. (1996) several authors have worked in this area. Aue et al. (2006), Aue et al. (2009), Fremdt (2014b) and Aue et al. (2014) developed methodology for detecting changes in the coefficients of a linear model, while Wied and Galeano (2013) and Pape et al. (2016) considered sequential monitoring schemes for changes in special functionals such as the correlation or variance. A Mosum-approach was employed by Leisch et al. (2000), Horváth et al. (2008) or Chen and Tian (2010) to monitor the mean and linear models, respectively. Recently, Hoga (2017) used an ℓ_1-norm to detect changes in the mean and variance of a multivariate time series, Kirch and Weber (2018) defined a unifying framework for detecting changes in different parameters with the help of several statistics and Otto and Breitung (2019) considered a Backward CUSUM, which monitors changes based on recursive residuals in a linear model. A helpful but not exhaustive overview of different sequential procedures can be found in Section 1, in particular Table 1, of Anatolyev and Kosenok (2018). A common feature of all procedures in the cited literature consists in the comparison of estimators from different subsamples of the data. To be precise, let $X_1, \dots, X_m$ denote an initial training sample and $X_1, \dots, X_m, \dots, X_{m+k}$ be the available data at time $m + k$. Several authors propose to investigate the differences

$$\hat{\theta}_1^m - \hat{\theta}_{m+1}^{m+k} \ , \tag{1.3}$$

(in dependence of k), where $\hat{\theta}_i^j$ denotes the estimator of the parameter from the sample $X_i, \dots, X_j$. In the sequential change point literature monitoring schemes based on the differences (1.3) are usually called (ordinary) CUSUM procedures and have been considered by Horváth et al. (2004), Aue et al. (2006, 2009, 2014), Schmitz and Steinebach (2010) or Hoga (2017). Other authors suggest using a function of the differences

$$\left\{ \hat{\theta}_1^m - \hat{\theta}_{m+j+1}^{m+k} \right\}_{j=0,\dots,k-1} \tag{1.4}$$

3

(in dependence of k) and the corresponding procedures are usually called Page-CUSUM tests [see Fremdt (2014b), Aue et al. (2015), or Kirch and Weber (2018) among others]. As an alternative we propose - following ideas of Dette and Gösmann (2018) - a monitoring scheme based on a function of the differences

$$\left\{ \hat{\theta}_1^{m+j} - \hat{\theta}_{m+j+1}^{m+k} \right\}_{j=0,\ldots,k-1} . \tag{1.5}$$

The intuitive advantage of (1.5) over (1.3) is the screening for all possible positions of the change point, which takes into account that the change point not necessarily comes with observation X_{m+1} and so $\hat{\theta}_{m+1}^{m+k}$ maybe 'corrupted' by pre-change observations. This issue is also partially addressed by (1.4), where different positions are examined and compared with the estimator of the parameter from the training sample. We will demonstrate in Section 4 that sequential monitoring schemes based on the differences (1.5) yield a substantial improvement in power compared to the commonly used methods based on (1.3) and (1.4). To avoid misunderstandings, the reader should note that a (total) comparison based on differences of the form (1.5), is typically also called a CUSUM-approach in the retrospective change point analysis [see Aue and Horváth (2013) for a comprehensive overview of (retrospective) change point analysis].

The present paper is devoted to a rigorous statistical analysis of a sequential monitoring based on the differences defined in (1.5) in the context of an open-end scenario. In Section 2 we introduce the new procedure and develop a corresponding asymptotic theory to obtain critical values such that monitoring can be performed at a controlled type I error. The theory is broadly applicable to detect changes in a general parameter θ of a multivariate time series. As all monitoring schemes in this context the method depends on a threshold function and we also discuss the choice of this function. In particular we establish an interesting result regarding this choice and establish a connection to corresponding ideas made by Horváth et al. (2004) and Fremdt (2014b), which may also be of interest in closed-end scenarios. In Section 3 we discuss several special cases and demonstrate that the new methodology is applicable to detect changes in the mean and the parameters of a linear model. Finally, we present a small simulation study in Section 4, where we compare our approach to those developed by Horváth et al. (2004) and Fremdt (2014b). In particular we demonstrate that the monitoring scheme based on the differences (1.5) yields a test with a controlled type I error and a smaller type II error than the procedures in the cited references.

2 Asymptotic properties

Throughout this paper let F denote a d-dimensional distribution function and $\theta = \theta(F)$ a p-dimensional parameter of F. We will denote by

$$\hat{F}_i^j(z) = \frac{1}{j-i+1} \sum_{t=i}^{j} I\{X_t \leq z\} \tag{2.1}$$

the empirical distribution function of observations $X_i, \ldots, X_j$ (here the inequality is understood component-wise) and consider the canonical estimator $\hat{\theta}_i^j = \theta(\hat{F}_i^j)$ of the parameter θ from the sample $X_i, \ldots, X_j$.

To test the hypotheses (1.1) and (1.2) in the described online setting in a open-end scenario we propose a monitoring scheme defined by

$$\hat{E}_m(k) = m^{-1/2} \max_{j=0}^{k-1} (k-j) \left\| \hat{\theta}_1^{m+j} - \hat{\theta}_{m+j+1}^{m+k} \right\|_{\hat{\Sigma}^{-1}} , \tag{2.2}$$

where the statistic $\hat{\Sigma}$ denotes an estimator of the long-run variance matrix Σ (defined in Assumption 2.2) and the symbol $\|v\|_A^2 = v^\top A v$ denotes a weighted norm of the vector v induced by the positive definite matrix A. The monitoring is then performed as follows. With observation X_{m+k} arriving, one computes $\hat{E}_m(k)$ and compares it to an appropriate threshold function, which is also called *weighting function* in the literature, say w. If

$$\hat{E}_m(k) > c_\alpha w(k/m) \tag{2.3}$$

monitoring is stopped and the null hypothesis (1.1) is rejected in favor of the alternative (1.2). If the inequality (2.3) does not hold, monitoring is continued with the next observation X_{m+k+1}. We will derive the limiting distribution of $\sup_{k=1}^{\infty} \hat{E}_m(k)/w(k/m)$ in Theorem 2.6 below to determine the constant c_α involved in (2.3), such that the test keeps a nominal level of α (asymptotically as $m \to \infty$).

Remark 2.1 The statistic (2.2) is related to a detection scheme, which was recently proposed by Dette and Gösmann (2018) for the closed-end case, where monitoring ends with observation mT, for some $T \in \mathbb{N}$. These authors considered the statistic

$$\hat{D}_m(k) = m^{-3/2} \max_{j=0}^{k-1} (m+j)(k-j) \|\hat{\theta}_1^{m+j} - \hat{\theta}_{m+j+1}^{m+k}\|_{\hat{\Sigma}^{-1}} , \tag{2.4}$$

and showed

$$\max_{k=1}^{mT} \frac{\hat{D}_m(k)}{w(k/m)} \xrightarrow{\mathcal{D}} \max_{t \in [0,T]} \max_{s \in [0,T]} \frac{|(s+1)W(t+1) - (t+1)W(s+1)|}{w(t)} , \tag{2.5}$$

where W denotes a p-dimensional Brownian motion and throughout this paper the symbol

$\overset{\mathcal{D}}{\Longrightarrow}$ denotes weak convergence (in the space under consideration). However, this statistic cannot be considered in an open-end scenario for the typical threshold functions considered in the literature satisfying $\limsup_{t\to\infty} t/w(t) < \infty$ (in this case the limit on the right-hand side of (2.5) would be almost surely infinite for $T = \infty$). As threshold functions satisfying $\limsup_{t\to\infty} t^2/w(t) < \infty$ will cause a loss in power as demonstrated in an unpublished simulation study, we propose to replace the factor $(m+j)$ in (2.4) by the size of the initial sample m, which leads to the monitoring scheme defined by (2.2).

To discuss the asymptotic properties of our approach, we require the following notation. The symbol $\overset{\mathbb{P}}{\Longrightarrow}$ denotes convergence in probability. The process $\{W(s)\}_{s\in[0,\infty)}$ will usually represent a standard p-dimensional Brownian motion. For a vector $v \in \mathbb{R}^d$, we denote by $|v| = \left(\sum_{i=1}^{d} v_i^2\right)^{1/2}$ its Euclidean norm. For the sake of a clear distinction we will employ

$$\overset{n}{\underset{i=1}{\sup}}\, a(i)$$

for discrete indexing (with integer arguments) and

$$\sup_{0\leq x\leq 1} a(x)$$

for continuous indexing (with arguments taken from the interval $[0,1]$ or another subset of $\mathbb{R}$).

Next, we define the influence function (assuming its existence) by

$$\mathcal{IF}(x, F, \theta) = \lim_{\varepsilon \searrow 0} \frac{\theta((1-\varepsilon)F + \varepsilon\delta_x) - \theta(F)}{\varepsilon}\ , \tag{2.6}$$

where $\delta_x(z) = I\{x \leq z\}$ is the distribution function of the Dirac measure at the point $x \in \mathbb{R}^d$ and the inequality in the indicator is again understood component-wise. We will focus on functionals that allow for an asymptotic linearization in terms of the influence function, that is

$$\hat{\theta}_i^j - \theta = \theta(\hat{F}_i^j) - \theta(F) = \frac{1}{j-i+1}\sum_{t=i}^{j} \mathcal{IF}(X_t, F, \theta) + R_{i,j} \tag{2.7}$$

with asymptotically negligible remainder terms $R_{i,j}$. Finally, for the sake of readability we introduce the following abbreviation

$$\mathcal{IF}_t = \mathcal{IF}(X_t, F_t, \theta)\ ,$$

where F_t is again the distribution function of X_t. Under the null hypothesis (1.1) we will impose the following assumptions on the underlying time series.

Assumption 2.2 (Approximation) *The time series $\{X_t\}_{t\in\mathbb{Z}}$ is (strictly) stationary, such that $F_t = F$ for all $t \in \mathbb{Z}$. Further, for each $m \in \mathbb{N}$ there exist two independent, p-dimensional standard Brownian motions $W_{m,1}$ and $W_{m,2}$, such that for some positive constant $\xi < 1/2$ the following approximations hold*

$$\sup_{k=1}^{\infty} \frac{1}{k^\xi} \left| \sum_{t=m+1}^{m+k} \mathcal{IF}_t - \sqrt{\Sigma} W_{m,1}(k) \right| = \mathcal{O}_{\mathbb{P}}(1) \tag{2.8}$$

and

$$\frac{1}{m^\xi} \left| \sum_{t=1}^{m} \mathcal{IF}_t - \sqrt{\Sigma} W_{m,2}(k) \right| = \mathcal{O}_{\mathbb{P}}(1) \tag{2.9}$$

as $m \to \infty$, where $\Sigma = \sum_{t\in\mathbb{Z}} \mathrm{Cov}\left(\mathcal{IF}_0,\ \mathcal{IF}_t\right) \in \mathbb{R}^{p\times p}$ denotes the long-run variance matrix of the process $\{\mathcal{IF}_t\}_{t\in\mathbb{Z}}$, which we assume to exist and to be non-singular.

Assumption 2.3 (Threshold function) *The threshold function $w : \mathbb{R}_{\geq 0} \to \mathbb{R}_+$ is uniformly continuous, has a positive lower bound, say $\ell_w > 0$, and satisfies*

$$\limsup_{t\to\infty} \frac{t}{w(t)} < \infty \ .$$

Assumption 2.4 (Linearization) *The remainder terms in the linearization* (2.7) *satisfy*

$$\max_{\substack{i,j=1 \\ i<j}}^{k} \frac{(j-i+1)}{\sqrt{k}} |R_{i,j}| = o(1) \tag{2.10}$$

as $k \to \infty$ with probability one.

Remark 2.5 Let us give a brief explanation on the assumption stated above.

(i) Assumption 2.2 is a uniform invariance principle and frequently used in the (sequential) change point literature [see for example Aue et al. (2006) or Fremdt (2014b) among others]. In the one-dimensional case, Assumption (2.8) was verified by Aue and Horváth (2004) for different classes of time series including GARCH or strongly mixing processes and can be easily extended to the multivariate case considered here. Assumption 2.2 is stronger than a functional central limit theorem, which is usually sufficient to work in a closed-end setup [see for example Wied and Galeano (2013), Pape et al. (2016) or Dette and Gösmann (2018)]

(ii) Assumption 2.3 gives necessary restrictions on the feasible set of threshold functions, which are required for the existence of a (weak) limit derived in Theorem 2.6. It is also worth mentioning that the assumption of a lower bounded threshold can be relaxed

to

$$\lim_{t \to 0} \frac{t^\gamma}{w(t)} = 0$$

for a constant $0 \leq \gamma < 1/2$. In this case, the assumption for the remainders in (2.10) has to be replaced by

$$\max_{\substack{1=i,j \\ i<j}}^{m+k} \frac{(j-i+1)}{(m+k)^{1/2-\gamma}} |R_{i,j}| = o(1) \ \ a.s.$$

For the sake of a transparent presentation we use the assumption of a lower bound here as this also simplifies the technical arguments in the proofs later on.

(iii) Assumption 2.4 is crucial for the proof of our main theorem and directly implies

$$\sup_{k=1}^{\infty} \max_{\substack{1=i,j \\ i<j}}^{m+k} \frac{(j-i+1)}{(m+k)^{1/2}} |R_{i,j}| = \sup_{k=m+1}^{\infty} \max_{\substack{i,j=1 \\ i<j}}^{k} \frac{(j-i+1)}{k^{1/2}} |R_{i,j}| = o(1) \ \ a.s.$$

Note that in the location model $\theta(F) = \mathbb{E}_F[X]$ we have $R_{i,j} = 0$ and (2.10) obviously holds. In general however, Assumption 2.4 is highly non-trivial and crucially depends on the structure of the functional θ and the time series $\{X_t\}_{t \in \mathbb{Z}}$. For a comprehensive discussion the reader is referred to Dette and Gösmann (2018), where the estimate (2.10) has been verified in probability for different functionals including quantiles and variance.

The following result is the main theorem of this section.

Theorem 2.6 *Assume that the null hypothesis* (1.1) *and Assumptions* 2.2 - 2.4 *hold. If further* $\hat{\Sigma}_m$ *is a consistent and non-singular estimator of the long-run variance matrix* Σ *it holds that*

$$\sup_{k=1}^{\infty} \frac{\hat{E}_m(k)}{w(k/m)} \xrightarrow{\mathcal{D}} \sup_{0 \leq t < \infty} \max_{0 \leq s \leq t} \frac{t+1}{w(t)} \left| W\left(\frac{s}{s+1}\right) - W\left(\frac{t}{t+1}\right) \right|, \qquad (2.11)$$

where W *is a* p-*dimensional Brownian motion with independent components.*

For the sake of completeness, the reader should note that due to Assumption 2.3 the asymptotic behaviour of the threshold guarantees that the random variable on the right-hand side of (2.11) is finite (with probability one).

In light of Theorem 2.6 one can choose a constant $c(\alpha)$, such that

$$\mathbb{P}\left(\sup_{0 \leq t < \infty} \max_{0 \leq s \leq t} \frac{t+1}{w(t)} \left| W\left(\frac{s}{s+1}\right) - W\left(\frac{t}{t+1}\right) \right| > c(\alpha) \right) \leq \alpha \ . \qquad (2.12)$$

8

The following corollary then states that our approach leads to a level α detection scheme.

Corollary 2.7 *Grant the Assumptions of Theorem 2.6 and further let $c(\alpha)$ satisfy inequality (2.12), then*

$$\limsup_{m \to \infty} \mathbb{P}\left(\sup_{k=1}^{\infty} \frac{\hat{E}_m(k)}{w(k/m)} > c(\alpha) \right) \leq \alpha \ .$$

The limit distribution obtained in Theorem 2.6 strongly depends on the considered threshold. A special family of thresholds that has received considerable attention in the literature [see Horváth et al. (2004), Fremdt (2014b), Kirch and Weber (2018) among many others] is given by

$$w_\gamma(t) = (1+t) \max\left\{ \left(\frac{t}{1+t} \right)^\gamma, \varepsilon \right\} \qquad \text{with} \qquad 0 \leq \gamma < 1/2 \ , \tag{2.13}$$

where the cutoff $\varepsilon > 0$ can be chosen arbitrary small in applications and only serves to reduce the assumptions and technical arguments in the proof [see also Wied and Galeano (2013)]. With these functions the limit distribution in (2.11), with the threshold function as the denominator, can be simplified to an expression that is more easily tractable via simulations. Straightforward calculations yield that Assumption 2.3 is satisfied by the function w_γ and the limit distribution in Theorem 2.6 simplifies as follows.

Corollary 2.8 *For a p-dimensional Brownian motion W with independent components it holds that*

$$\sup_{0 \leq t < \infty} \max_{0 \leq s \leq t} \frac{t+1}{w_\gamma(t)} \left| W\left(\frac{s}{s+1} \right) - W\left(\frac{t}{t+1} \right) \right| \stackrel{\mathcal{D}}{=} \sup_{0 \leq t < 1} \max_{0 \leq s \leq t} \frac{1}{\max\{t^\gamma, \varepsilon\}} \left| W(t) - W(s) \right| := L_{1,\gamma} \ .$$

For the investigation of the consistency of the monitoring scheme (2.2) we require the following assumption.

Assumption 2.9 *Under the alternative H_1 defined in (1.2) let*

$$\theta^{(1)} := \theta(F_1) = \theta(F_2) = \cdots = \theta(F_{m+k^*}) \neq \theta^{(2)} := \theta(F_{m+k^*+1}) = \theta(F_{m+k^*+1}) = \cdots$$

Further assume that k^* is independent of m and that the process $\{\mathcal{IF}_t\}$ is of the following order before and after the change, respectively,

$$\frac{1}{\sqrt{m}} \left| \sum_{t=1}^{m+k^*} \mathcal{IF}_t \right| = \mathcal{O}_\mathbb{P}(1) \quad and \quad \frac{1}{\sqrt{m}} \left| \sum_{t=m+k^*+1}^{2m} \mathcal{IF}_t \right| = \mathcal{O}_\mathbb{P}(1) \ . \tag{2.14}$$

Additionally assume that the remainders defined in (2.4) *satisfy*

$$|R_{1,m+k^*}| = \mathcal{O}_{\mathbb{P}}\Big(\frac{1}{\sqrt{m}}\Big) \quad and \quad |R_{m+k^*+1,2m}| = \mathcal{O}_{\mathbb{P}}\Big(\frac{1}{\sqrt{m}}\Big) \ . \tag{2.15}$$

Remark 2.10 The assumptions stated above are substantially weaker than those used to investigate the asymptotic properties of $\sup_{k=1}^{\infty} \hat{E}_m(k)/w(\frac{k}{m})$ under the null hypothesis. Basically, we only assume reasonable behavior of the time series before and after the change point and can drop the uniform approximation in Assumption 2.2 and the uniform negligibility of the remainders in Assumption 2.4. It is easy to see, that (2.14) is already satisfied if both, the phases before and after the change fulfill a central limit theorem. Finally, it is worth mentioning that one can also derive the subsequent results when replacing the $2m$ by cm for an arbitrary constant $c > 1$, however - for the sake of better readability - we will work with this (minimally stricter) assumption.

The next Theorem yields consistency under the alternative hypothesis.

Theorem 2.11 *Assume that the alternative hypothesis* (1.2) *and Assumptions* 2.3 *and* 2.9 *hold. If further* $\hat{\Sigma}_m$ *is a consistent and non-singular estimator of the long-run variance matrix* Σ *it holds that*

$$\sup_{k=1}^{\infty} \frac{\hat{E}_m(k)}{w(k/m)} \stackrel{\mathbb{P}}{\Longrightarrow} \infty \ .$$

Consequently,

$$\lim_{m \to \infty} \mathbb{P}\bigg(\sup_{k=1}^{\infty} \frac{\hat{E}_m(k)}{w(k/m)} > c \bigg) = 1$$

for any constant $c \in \mathbb{R}$.

3   Some specific change point problems

In this section we briefly illustrate how the theory developed in Section 2 can be employed to construct monitoring schemes for a specific parameter of the distribution function. For the sake of brevity we restrict ourselves to the mean and the parameters in a linear model. Other examples such as the variance or quantiles can be found in Dette and Gösmann (2018).

3.1 Changes in the mean

The sequential detection of changes in the mean

$$\mu(F) = \mathbb{E}_F[X] = \int_{\mathbb{R}^d} x dF(x) \ .$$

has been extensively discussed in the literature [see Aue and Horváth (2004), Fremdt (2014b) or Hoga (2017) among many others].

Is is easy to verify (and well known), that the influence function for the mean is given by

$$\mathcal{IF}(x, F, \mu) = x - \mathbb{E}_F[X] \ ,$$

and Assumption 2.4 and the second part of Assumption 2.9 are obviously satisfied in this case since we have $R_{i,j} = 0$ for all i, j. For the remaining assumptions in Section 2 it now suffices that the centered time series $\{X_t - \mathbb{E}[X_t]\}_{t \in \mathbb{Z}}$ fulfills Assumption 2.2, which also implies the remaining part of Assumption 2.9 [see also the discussion in Remark 2.5]. In this situation both, Theorem 2.6 and Theorem 2.11 are valid provided that the chosen threshold fulfills Assumption 2.3.

3.2 Changes in linear models

Consider the time-dependent linear model

$$Y_t = P_t^\top \beta_t + \varepsilon_t \ , \tag{3.1}$$

where the random variables $\{P_t\}_{t \in \mathbb{N}}$ are the $\mathbb{R}^p$-valued predictors, $\beta_t \in \mathbb{R}^p$ is a p-dimensional parameter and $\{\varepsilon_t\}_{t \in \mathbb{N}}$ is a centered random sequence independent of $\{P_t\}_{t \in \mathbb{N}}$. The identification of changes in the vector of parameters in the linear model represents the prototype problem in sequential change point detection as it has been extensively studied in the literature [see Chu et al. (1996), Horváth et al. (2004), Aue et al. (2009), Fremdt (2014b), among many others].

This situation is covered by the general theory developed in Section 2 and 3. To be precise let

$$X_t = (P_t^\top, Y_t)^\top \ \in \mathbb{R}^d \ , \ \ d = p + 1 \ \text{ and } \ t = 1, 2 \dots \tag{3.2}$$

be the the joint vectors of predictor and response with (joint) distribution function F_t, such that the marginal distributions of Y_t and P_t are given by

$$F_{t,Y} = F_t(\infty, \dots, \infty, \cdot) \quad \text{and} \quad F_{t,P} = F_t(\cdot, \dots, \cdot, \infty) \ ,$$

respectively, where we will assume that the predictor sequence is stationary, that is $F_{t,P} = F_P$. In a first step we will consider the case, where the moment matrix

$$M := \mathbb{E}[P_1 P_1^\top] = \int_{\mathbb{R}^d} p \cdot p^\top dF_P(p)$$

is known (we will discuss later on why this assumption is non-restrictive) and non-singular. In this setup, the parameter β_t can be represented as a functional of the distribution function F_t, that is

$$\beta_t = \beta(F_t) := M^{-1} \cdot \int_{R^d} p \cdot y dF_t(y, p) = M^{-1} \cdot \mathbb{E}\big[P_t Y_t\big] \ ,$$

which leads to the estimators

$$\hat{\beta}_i^j = \beta(\hat{F}_i^j) = \frac{M^{-1}}{j - i + 1} \sum_{t=i}^{j} P_t Y_t \tag{3.3}$$

from the sample $(P_i, Y_i), \dots, (P_j, Y_j)$. To compute the influence function, let $(p, y) \in \mathbb{R}^p \times \mathbb{R}$, then

$$
\begin{aligned}
\mathcal{IF}\big((p, y), F_t, \beta\big) &= \lim_{\eta \searrow 0} \frac{\beta\big((1 - \eta)F_t + \varepsilon \delta_{(p,y)}\big) - \beta(F_t)}{\eta} \\
&= \lim_{\eta \searrow 0} M^{-1} \left[\frac{(1 - \eta)\mathbb{E}[P_t Y_t] + \eta p y}{\eta} - \frac{\beta_t}{\eta} \right] \\
&= M^{-1}\big(p y - \mathbb{E}[P_t Y_t]\big) \ ,
\end{aligned}
$$

which is the influence function (for β) in the linear model stated above [see for example Hampel et al. (1986) for a comprehensive discussion of influence functions]. In the following, we will use the notation $\mathcal{IF}_t = \mathcal{IF}\big(X_t, F_t, \beta\big)$ again. Note that

$$\mathcal{IF}_t = M^{-1}\big(P_t Y_t - \mathbb{E}[P_t Y_t]\big) = M^{-1} P_t Y_t - \beta_t \ , \tag{3.4}$$

which directly gives $\mathbb{E}[\mathcal{IF}_t] = 0$. Under the null hypothesis the random sequence $(X_t)_{t\in\mathbb{N}}$ is stationary and the linearization defined in (2.7) simplifies to

$$
\begin{aligned}
\hat{\beta}_i^j - \beta_1 = \beta(\hat{F}_i^j) - \beta_1 &= \frac{M^{-1}}{j-i+1}\sum_{t=i}^{j} P_t Y_t - \beta_1 = \frac{1}{j-i+1}\sum_{t=i}^{j}\left(M^{-1}P_t Y_t - \beta_t\right) \\
&= \frac{1}{j-i+1}\sum_{t=i}^{j}\mathcal{IF}_t\ .
\end{aligned}
\tag{3.5}
$$

Consequently, the remainders in (2.7) vanish and Assumption 2.4 is obviously satisfied. Next, note that the long-run variance matrix is given by

$$
\Sigma = \sum_{t\in\mathbb{Z}}\mathrm{Cov}\left(\mathcal{IF}_0,\ \mathcal{IF}_t\right) = M^{-1}\Gamma M^{-1}
\tag{3.6}
$$

with $\Gamma = \sum_{t\in\mathbb{Z}}\mathrm{Cov}\left(Y_0 P_0,\ Y_t P_t\right)$, which can be estimated by $\hat{\Sigma} = M^{-1}\hat{\Gamma}M^{-1}$ where $\hat{\Gamma}$ is an estimator for Γ. Observing (3.5) it is now easy to see that in the resulting statistic $\hat{E}_m$ the matrix M cancels out, that is

$$
\begin{aligned}
\hat{E}_m(k) &= m^{-1/2}\max_{j=0}^{k-1}(k-j)\left\|\hat{\beta}_1^{m+j} - \hat{\beta}_{m+j+1}^{m+k}\right\|_{\hat{\Sigma}^{-1}} \\
&= m^{-1/2}\max_{j=0}^{k-1}(k-j)\left\|\frac{1}{m+j}\sum_{t=1}^{m+j}Y_t P_t - \frac{1}{k-j}\sum_{t=m+j+1}^{m+k}Y_t P_t\right\|_{\hat{\Gamma}^{-1}}
\end{aligned}
\tag{3.7}
$$

and therefore does not depend on the matrix M. We therefore obtain the following result, which describes the asymptotic properties of the monitoring scheme based on the statistic $\hat{E}_m$ for a change in the parameter in the linear regression model (3.1). The proof is a direct consequence of Theorem 2.6 and 2.11.

Corollary 3.1 *Assume that the predictor sequence $\{P_t\}_{t\in\mathbb{N}}$ is strictly stationary with a non-singular second moment matrix $M = \mathbb{E}[P_1 P_1^\top]$. Let $\hat{\Gamma}_m$ denote a non-singular, consistent estimator of the non-singular long-run variance matrix Γ defined in (4.8). Further suppose that the sequences $\{P_t\}_{t\in\mathbb{N}}$ and $\{\varepsilon_t\}_{t\in\mathbb{N}}$ are independent and let the threshold function under consideration fulfill Assumption 2.3.*

(i) Under the null hypothesis H_0 of no change assume additionally that the sequence $\{\mathcal{IF}_t\}_{t\in\mathbb{N}}$ defined in (3.4) admits the approximation in Assumption 2.2. Then monitoring based on the statistic $\hat{E}$ in (3.7) is an asymptotic level α procedure.

(ii) Under the alternative hypothesis H_1 assume that $\{\mathcal{IF}_t\}_{t\in\mathbb{N}}$ fulfills (2.14) of Assumption 2.9. Then the monitoring based on the statistic $\hat{E}$ in (3.7) is consistent.

4   Finite sample properties

In this section we investigate the finite sample properties of our monitoring procedure and demonstrate its superiority with respect to the available methodology. We choose the following two statistics as our benchmark

$$
\begin{aligned}
\hat{Q}_m(k) &:= \frac{k}{m^{1/2}} \left\| \hat{\theta}_1^m - \hat{\theta}_m^{m+k} \right\|_{\hat{\Sigma}^{-1}}, \\
\hat{P}_m(k) &:= \max_{j=0}^{k-1} \frac{k-j}{m^{1/2}} \left\| \hat{\theta}_1^m - \hat{\theta}_{m+j+1}^{m+k} \right\|_{\hat{\Sigma}^{-1}}.
\end{aligned}
\tag{4.1}
$$

The procedure based on $\hat{Q}$ was originally proposed by Horváth et al. (2004) for detecting changes in the parameters of linear models and since then reconsidered for example by Aue et al. (2012), Wied and Galeano (2013) and Pape et al. (2016) for the detection of changes in the CAPM-model, correlation and variances, respectively. A statistic of the type $\hat{P}$ was recently proposed by Fremdt (2014b) and has been already reconsidered by Kirch and Weber (2018). In the simulation study we will restrict ourselves to the commonly used class of threshold functions w_γ defined in (2.13), where we set the involved, technical constant $\varepsilon = 0$.

Under the assumptions made in Section 2, it can be shown by similar arguments as given in Appendix A that

$$
\sup_{k=1}^{\infty} \frac{\hat{Q}_m(k)}{w_\gamma(k/m)} \xrightarrow{\mathcal{D}} \sup_{0 \le t < 1} \frac{|W(t)|}{\max\{t^\gamma, \varepsilon\}} =: L_{2,\gamma}
\tag{4.2}
$$

and

$$
\widetilde{\sup}_{k=1}^{\infty} \frac{\hat{P}_m(k)}{w_\gamma(k/m)} \xrightarrow{\mathcal{D}} \sup_{0 \le t < 1} \max_{0 \le s \le t} \frac{1}{\max\{t^\gamma, \varepsilon\}} \left| W(t) - \frac{1-t}{1-s} W(s) \right| =: L_{3,\gamma},
\tag{4.3}
$$

where W denotes a p-dimensional Brownian motion. For detailed proofs (under slightly different assumptions) of (4.2) and (4.3), the reader is relegated to Horváth et al. (2004) and Fremdt (2014b), where procedures of these types are considered in the special case of a linear model.

Recall the notation of $L_{1,\gamma}$ introduced in Corollary 2.8. By (4.2), (4.3) and Corollary 2.7 the necessary critical values for the procedures $\hat{E}$, $\hat{Q}$ and $\hat{P}$ combined with threshold w_γ are given as the $(1-\alpha)$-quantiles of the distributions $L_{1,\gamma}$, $L_{2,\gamma}$ and $L_{3,\gamma}$, respectively and can easily be obtained by Monte Carlo simulations. The quantiles are listed in Table 1 for dimensions $p = 1$ and $p = 2$ and have been calculated by 10000 runs simulating the corresponding distributions where the underlying Brownian motions have been approximated on a grid of 5000 points. In Sections 4.1 and 4.2 below, we will examine the finite sample properties of the three statistics for the detection of changes in the mean and in the regres-

sion coefficients of a linear model, respectively. All subsequent results presented in these sections are based on 1000 independent simulation runs and a fixed test level of $\alpha = 0.05$.

p	$\gamma \setminus \alpha$	$L_{1,\gamma}$			$L_{2,\gamma}$			$L_{3,\gamma}$		
		0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
	0	2.9762	2.4721	2.2175	2.8262	2.2599	1.9914	2.7912	2.2365	1.9497
1	0.25	3.1050	2.5975	2.3542	2.9638	2.4296	2.1758	2.9445	2.3860	2.1060
	0.45	3.4269	2.9701	2.7398	3.3817	2.9241	2.7002	3.3015	2.7992	2.5437
	0	3.4022	2.8943	2.6562	3.2272	2.6794	2.4008	3.2461	2.6957	2.4266
2	0.25	3.5279	3.0948	2.7781	3.3322	2.7981	2.5481	3.3630	2.8433	2.5911
	0.45	3.8502	3.3912	3.1509	3.7010	3.2046	2.9543	3.7467	3.2966	3.0620

Table 1: *(1-α)-quantiles of the distributions $L_{1,\gamma}$, $L_{2,\gamma}$ and $L_{3,\gamma}$ for different choices of γ and different dimensions of the parameter. The cutoff constant was set to $\varepsilon = 0$. The results for $L_{2,\gamma}$ and $L_{3,\gamma}$ for $p = 1$ are taken from Fremdt (2014b) and Horváth et al. (2004), respectively.*

4.1 Changes in the mean

In this section we will compare the finite sample properties of the procedures based on the statistics $\hat{E}, \hat{P}$ and $\hat{Q}$ for changes in the mean as outlaid in Section 3.1. Here we test the null hypothesis of no change which is given by

$$H_0 : \ \mu_1 = \cdots = \mu_m = \mu_{m+1} = \mu_{m+2} = \dots \ , \tag{4.4}$$

while the alternative, that the parameter μ_t changes beyond the initial data set, is defined as

$$H_1 : \ \exists k^\star \in \mathbb{N} : \ \mu_1 = \cdots = \mu_{m+k^\star-1} \neq \mu_{m+k^\star} = \mu_{m+k^\star+1} = \dots \ . \tag{4.5}$$

We will consider two different data generating models, a white noise process and an autoregressive process given by

(M1) X_t i.i.d. $\sim \mathcal{N}(0, 1)$,

(M2) $X_t = 0.1 X_{t-1} + \varepsilon_t$ with ε_t i.i.d. $\sim \mathcal{N}(0, 1)$.

For the AR(1)-process specified in model (M2), we create a burn-in sample of 100 observations in the first place. To simulate the alternative hypotheses, changes in the mean are added to the data, that is

$$X_t^\delta = \begin{cases} X_t & \text{if } t < m + k^* \ , \\ X_t + \delta & \text{if } t \geq m + k^* \ , \end{cases}$$

where $\delta = \mathbb{E}[X_{m+k^*}] - \mathbb{E}[X_{m+k^*-1}]$ denotes the desired change amount. For the necessary covariance estimation we employ the well known quadratic spectral estimator [see (Andrews, 1991)] with its implementation in the R-package 'sandwich' [see Zeileis (2004)]. To take into account the possible appearance of changes only the initial stable segment $X_1, \ldots, X_m$ is used for this estimate, which is standard in the literature [see for example Horváth et al. (2004), Wied and Galeano (2013), or Dette and Gösmann (2018) among many others].

In Table 2 we display the type 1-Error for both time series models and different choices of γ in the threshold functions. The principle observation is, that all three statistical procedures offer a reasonable approximation of the desired nominal level of $\alpha = 0.05$. The results for the dependent model (M2) are slightly worse than those for the white noise model (M1). This effect may be caused by a less precise estimation of the long-run variance for small sample sizes. Accordingly, this effect is weaker for the case $m = 100$.

In Figures 1, 2, 3 and 4 we illustrate the power of the procedures under the alternative hypothesis for increasing values of the change and different change positions for combinations of $\gamma = 0$, $\gamma = 0.45$ and $m = 50$, $m = 100$. The basic tendency in all four plots is similar: While the procedures behave similar for a change close to the initial data set (first row), the method based on $\hat{E}$ is clearly superior to the others the more the distance to the initial set grows. To give an example, consider the left plot of the last row in Figure 1. Here the test based on the statistic $\hat{E}$ already has a power of 32.9% for a change of $\mu = 1$ whereas the tests based on the statistics $\hat{P}$ and $\hat{Q}$ have power of 24.4% and 22.7%, respectively. The superior performance of $\hat{E}$ can most likely be explained by the more accurate estimate of the pre-change parameter by $\hat{\theta}_1^{m+j}$, while the the other statistics only involve the estimator $\hat{\theta}_1^m$ [see formulas (2.2) and (4.1)].

For the sake of an appropriate understanding of our findings, the reader should be aware of the fact, that - although we consider open-end procedures here - simulations have to be stopped eventually. Here we chose this stopping point as 1000 ($m = 50$) or 3000 ($m = 100$) observations and it is expectable that the testing power of all procedures increases with a later stopping point. Therefore the observed superiority of $\hat{E}$ refers to *the type 2-Error until the specified stopping point*.

		(M1)			(M2)		
m	γ	$\hat{E}$	$\hat{Q}$	$\hat{P}$	$\hat{E}$	$\hat{Q}$	$\hat{P}$
	0	5.4%	5.2%	5.5%	8.1%	7.1%	8.2%
50	0.25	5.0%	4.9%	5.4%	8.3%	7.0%	9.5%
	0.45	4.5%	3.6%	4.8%	7.6%	5.6%	9.2%
	0	4.2%	4.3%	4.9%	6.9%	6.5%	6.9%
100	0.25	5.0%	4.9%	5.9%	7.6%	6.5%	7.0%
	0.45	6.0%	4.9%	7.0%	6.5%	4.8%	7.7%

Table 2: *Type 1-Error for the open-end procedures based on $\hat{E}$, $\hat{Q}$ and $\hat{P}$. The size of the known stable data was set to $m = 50$ (upper part), $m = 100$ (lower part).*
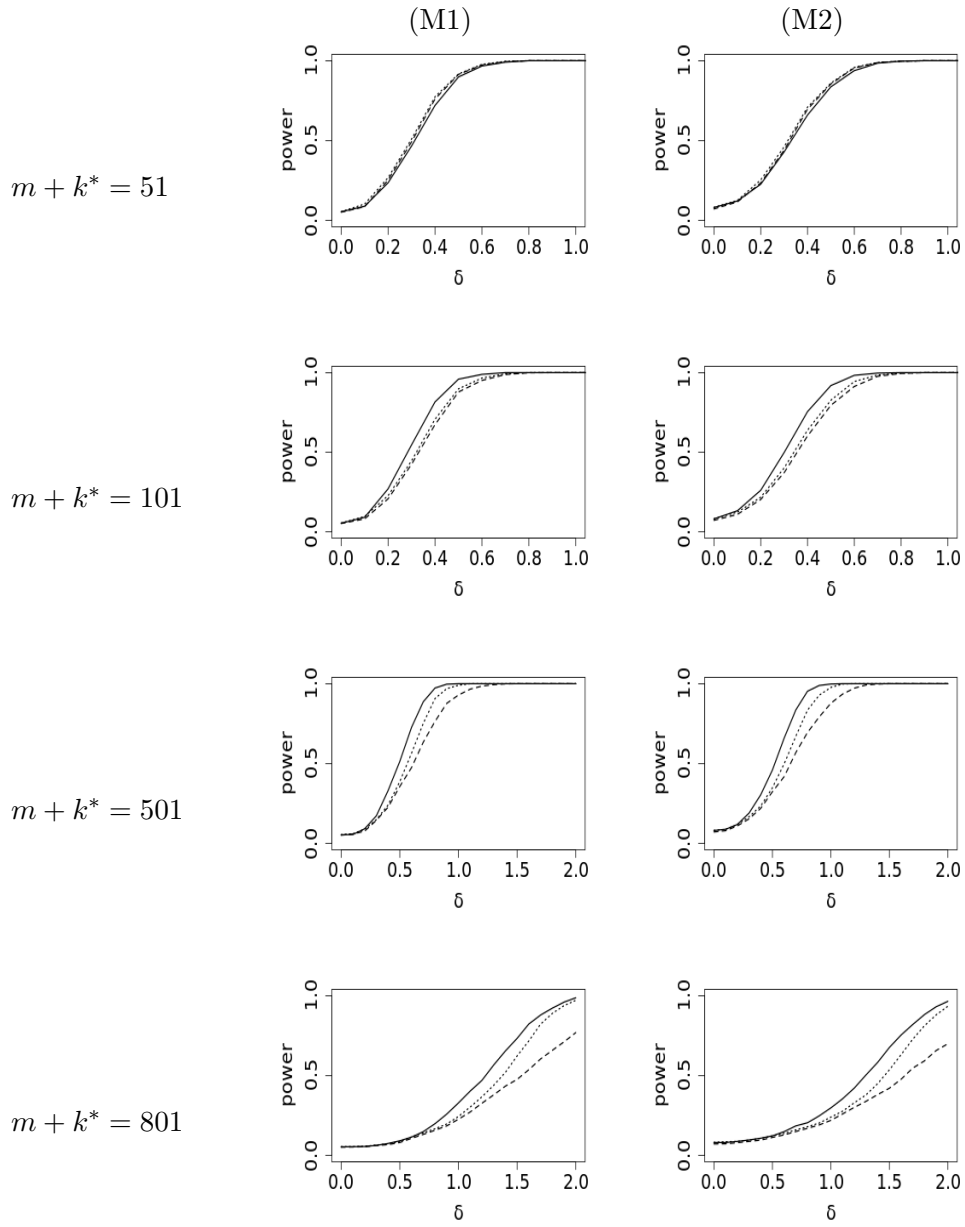
Figure 1: *Power of the monitoring procedures for a change in the mean based on the statistics* $\hat{E}$ *(solid line),* $\hat{Q}$ *(dashed line) and* $\hat{P}$ *(dotted line) with* $\gamma = 0$ *and* $m = 50$.

Figure 2: *Power of the monitoring procedures for a change in the mean based on the statistics $\hat{E}$ (solid line), $\hat{Q}$ (dashed line) and $\hat{P}$ (dotted line) with $\gamma = 0.45$ and $m = 50$.*
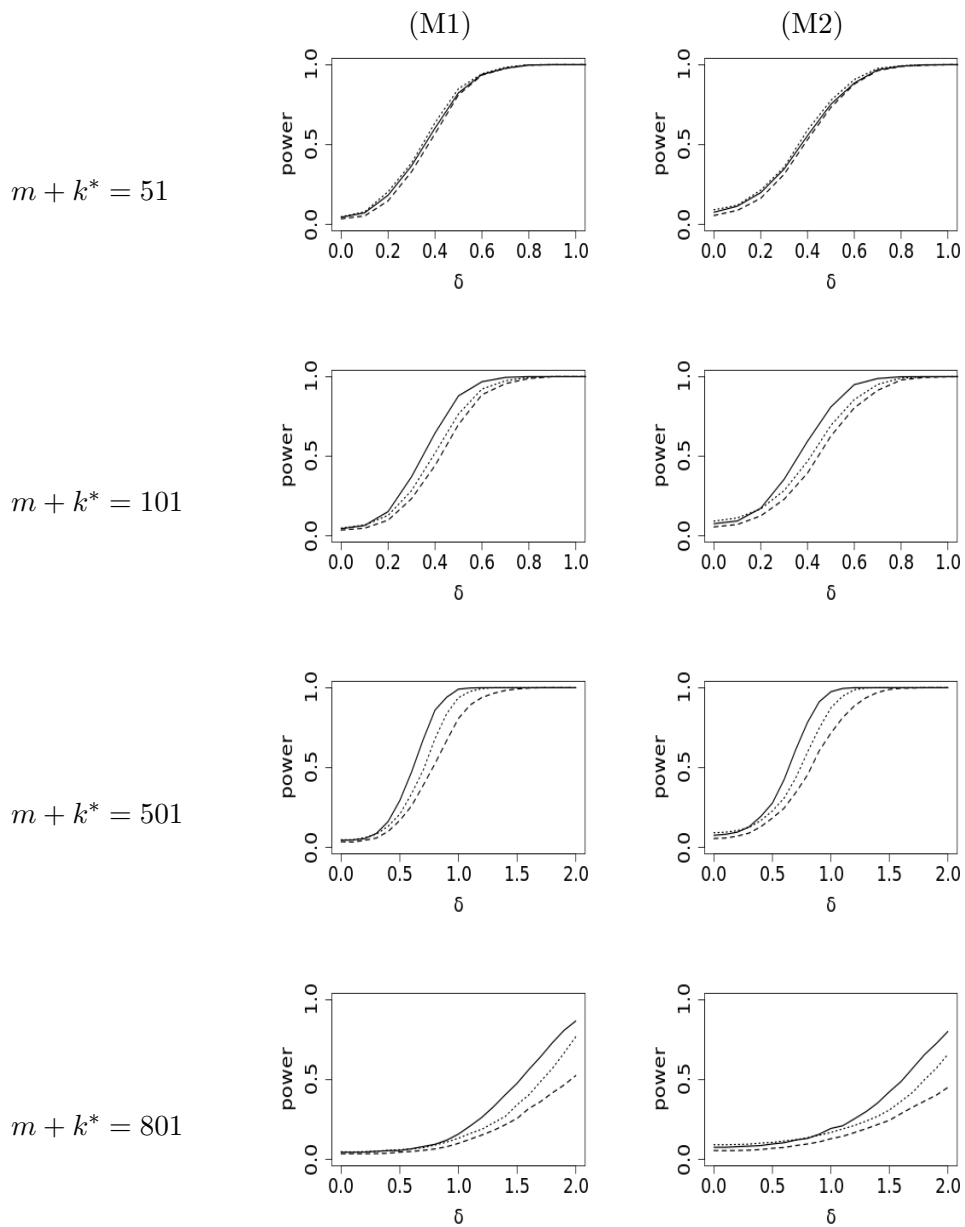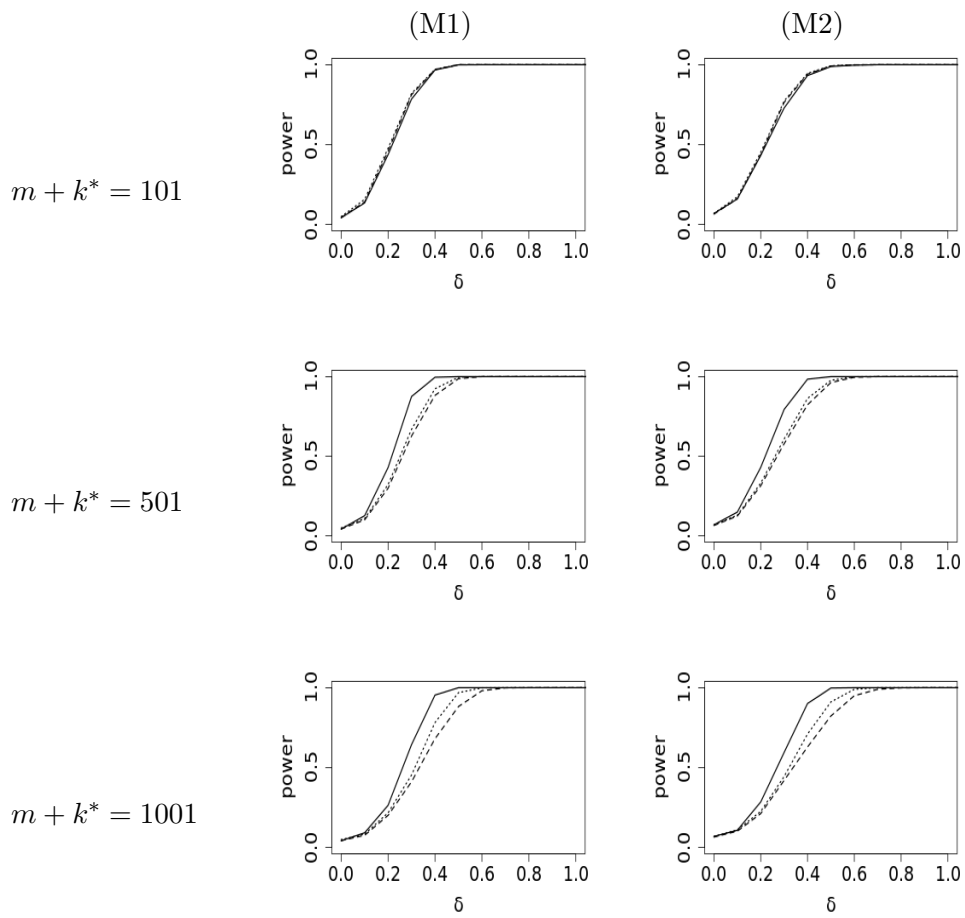
Figure 3: *Power of the monitoring procedures for a change in the mean based on the statistics $\hat{E}$ (solid line), $\hat{Q}$ (dashed line) and $\hat{P}$ (dotted line) with $\gamma = 0$ and $m = 100$.*
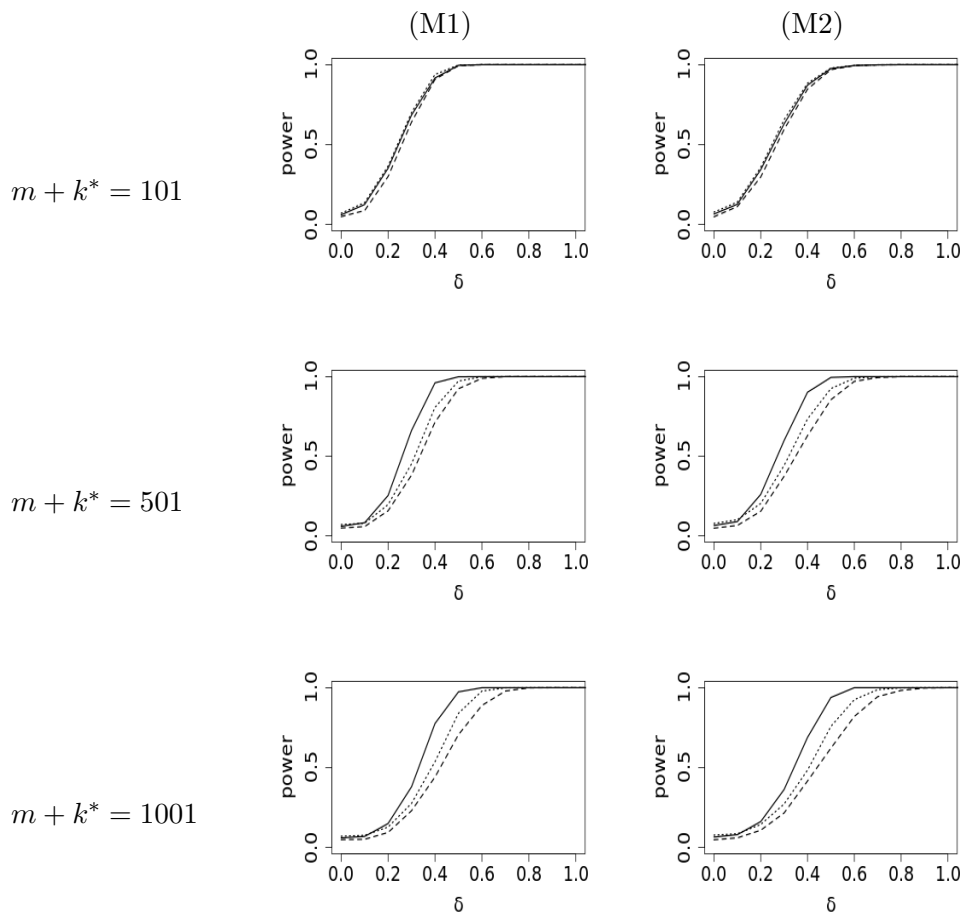
Figure 4: *Power of the monitoring procedures for a change in the mean based on the statistics $\hat{E}$ (solid line), $\hat{Q}$ (dashed line) and $\hat{P}$ (dotted line) with $\gamma = 0.45$ and $m = 100$.*

4.2 Changes in linear models

In this section we present some simulation results for the detection of changes in the linear model (3.1). We aim to detect changes in the unknown parameter vector $\beta_t \in \mathbb{R}^p$ by testing the null hypothesis

$$H_0 : \ \beta_1 = \cdots = \beta_m = \beta_{m+1} = \beta_{m+2} = \ldots \, , \qquad (4.6)$$

against the alternative that the parameter β_t changes beyond the initial data set, that is

$$H_1 : \ \exists k^\star \in \mathbb{N} : \ \ \beta_1 = \cdots = \beta_{m+k^\star-1} \neq \beta_{m+k^\star} = \beta_{m+k^\star+1} = \ldots \, . \qquad (4.7)$$

To be precise, we consider the model (3.1) with $p = 2$ and the following choice of predictors

(LM1) $P_t = (1, Z_t)^\top$,

(LM2) $P_t = (1, 1 + G_t)^\top$ with $G_t = \bar{\sigma}_t Z_t$ and $\bar{\sigma}_t^2 = 0.5 + 0.2 Z_{t-1} + 0.3 \bar{\sigma}_{t-1}^2$,

where Z_t denotes an i.i.d. sequence of $\mathcal{N}(0, 0.5)$ random variables in both models. The parameter vector is fixed at $\beta_t = (1, 1)$ under the null hypothesis and to examine the alternative hypothesis, changes are added to its second component, that is

$$\beta_t^\delta = \begin{cases} (1, 1)^\top & \text{if } t < m + k^* \, , \\ (1, 1 + \delta)^\top & \text{if } t \geq m + k^* \, . \end{cases}$$

For both scenarios we simulated the residuals ε_t in model (3.1) as i.i.d. $\mathcal{N}(0, 0.5)$ sequences. Note that the models specified above have been already considered by Fremdt (2014b). As pointed out in Section 3.2 the asymptotic variance that needs to be estimated within our procedures is given by

$$\Gamma = \sum_{t \in \mathbb{Z}} \text{Cov} \left(P_0 Y_0, \ P_t Y_t \right) . \qquad (4.8)$$

We estimate this quantity based on the stable segment of observations $(Y_1, P_1), \ldots, (Y_m, P_m)$ using the well known quadratic spectral estimator [see Andrews (1991)] with its implementation in the R-package 'sandwich' [see Zeileis (2004)].

The problem of detecting changes in the parameter of the linear model has also been addressed using partial sums of the residuals $\hat{\varepsilon}_t = Y_t - P_t^\top \hat{\beta}_I$ in statistics similar to (4.1), where $\hat{\beta}_I$ is an initial estimate of β computed from the initial stable segment [see for example Chu et al. (1996), Horváth et al. (2004), Fremdt (2014a) among many others]. Our approach directly compares estimators for the vector β_t, which are derived using the

general methodology introduced in Section 2 and 3. The resulting statistics are obtained replacing $\hat{\theta}$ by $\hat{\beta}$ in equation (4.1). We also refer to Leisch et al. (2000) for a comparison of residual based methods with methods using the estimators directly (these authors consider a statistic similar to $\hat{Q}$).

In Table 3 we display the approximation of the nominal level for the three statistics with different values of the parameter γ in the threshold function, where monitoring was stopped after 1500 observations. We observe a reasonable approximation of the nominal level 5% in the case $\gamma = 0$, while the rejection probabilities for for $\gamma = 0.25$ or $\gamma = 0.45$ slightly exceed the desired level of 5%. The fact that larger values of $\gamma \in [0, 1/2)$ can lead to a worse approximation of the desired type 1-Error has also been observed by other authors [see, for example, Wied and Galeano (2013)] and can be explained by a more sensitive threshold function at the monitoring start if γ is chosen close to $1/2$. Overall, the approximation is slightly better for the independent case in model (LM1).

In Figure 5 we compare the power with respect to the change amount for different change positions, where we restrict ourselves to the case $\gamma = 0$ for the sake of brevity. The results are very similar to those provided for the mean functional in Section 4.1. Again the monitoring scheme based on $\hat{E}$ outperforms the procedures based on $\hat{Q}$ and $\hat{P}$, and the superiority is larger for a later change. We omit a detailed discussion and summarize that the empirical findings have indicated superiority (w.r.t. testing power) of the monitoring scheme based on the statistic $\hat{E}$.

	(LM1)			(LM2)		
γ	$\hat{E}$	$\hat{Q}$	$\hat{P}$	$\hat{E}$	$\hat{Q}$	$\hat{P}$
0	6.4%	6.5%	6.7%	7.2%	6.7%	7.2%
0.25	7.6%	8.8%	9.1%	8.5%	9.6%	9.5%
0.45	12.0%	12.2%	12.1%	12.6%	12.2%	12.6%

Table 3: *Type 1-Error for the open-end procedures based on $\hat{E}$, $\hat{Q}$ and $\hat{P}$. The size of the known stable data was set to $m = 100$.*
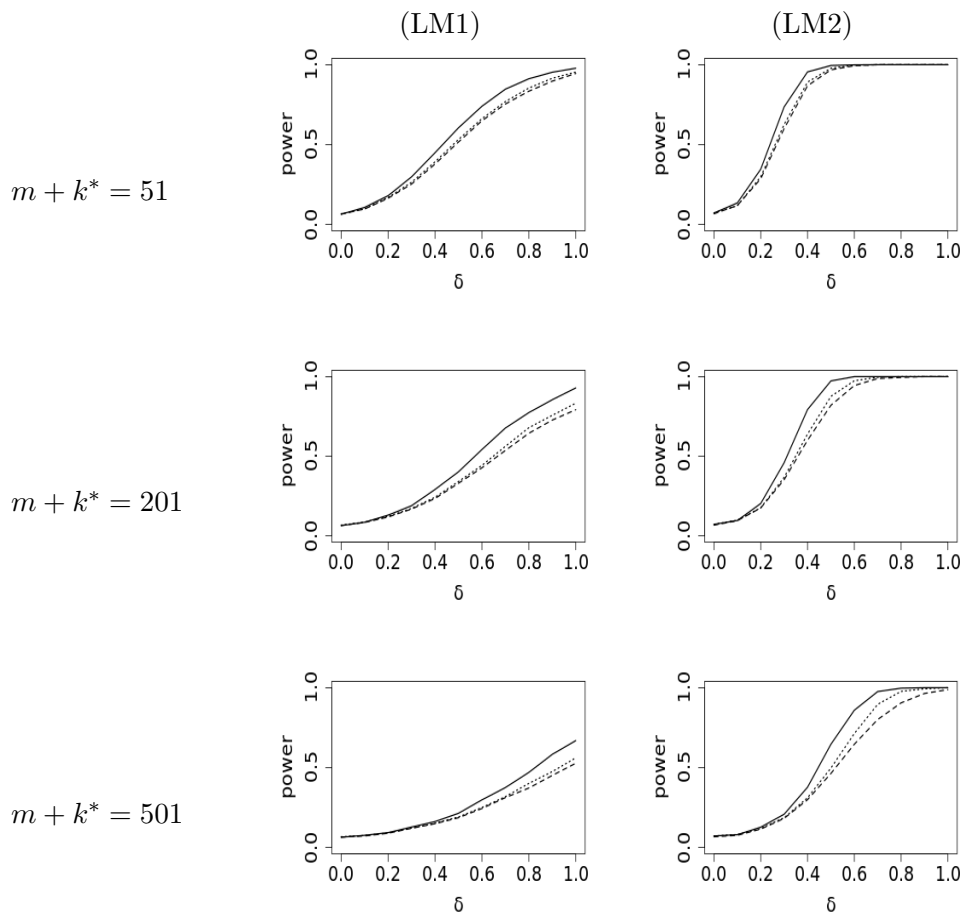
Figure 5: *Power of the monitoring procedures for a change in the regression parameters for the open-end procedures based on the statistics $\hat{E}$ (solid line), $\hat{Q}$ (dashed line) and $\hat{P}$ (dotted line) with $\gamma = 0$ and $m = 100$.*

5 Closed-end scenarios

It is worthwhile to mention that the theory developed so far also covers the case of closed-end scenarios [sometimes also called finite time horizon in the literature]. In this section, we will very briefly discuss this situation and present a small batch of simulation results, which also indicate the superiority of the statistic $\hat{E}$ for closed-end scenarios. Note that the null hypothesis in this setup is given by

$$H_0 : \ \theta_1 = \cdots = \theta_m = \theta_{m+1} = \theta_{m+2} = \ldots = \theta_{Tm} \ , \tag{5.1}$$

which is tested against the alternative that the parameters changes (once) at some time $m + 1 \leq m + k^\star \leq Tm$, that is

$$H_1 : \ \exists k^\star \in \mathbb{N} : \ \theta_1 = \cdots = \theta_{m+k^\star-1} \neq \theta_{m+k^\star} = \theta_{m+k^\star+1} = \ldots = \theta_{Tm} \ . \tag{5.2}$$

Here the factor $T \in \mathbb{N}$ controls the length of the monitoring period compared to the size of the initial data set. Under the assumptions stated in Section 2, we can prove a corresponding statement of Theorem 2.6 and Corollary 2.8.

Theorem 5.1 *Assume that the null hypothesis* (5.1) *and Assumptions* 2.2 - 2.4 *hold. If further* $\hat{\Sigma}_m$ *is a consistent and non-singular estimator of the long-run variance matrix* Σ *it holds that*

$$\sup_{k=1}^{Tm} \frac{\hat{E}_m(k)}{w_\gamma(k/m)} \overset{\mathcal{D}}{\Longrightarrow} \max_{0 \leq t < T} \max_{0 \leq s \leq t} \frac{1}{w_\gamma(t)} \left| \frac{1+t}{1+s} W(1+s) - W(1+t) \right|$$

$$\overset{\mathcal{D}}{=} \max_{0 < t < T/(T+1)} \max_{0 \leq s \leq t} \frac{1}{\max\{t^\gamma, \varepsilon\}} \left| W(t) - W(s) \right| =: L_{1,\gamma}(T) \ , \tag{5.3}$$

where W *is a p-dimensional Brownian motion with independent components.*

The proof of Theorem 5.1 follows by a straightforward adaption of the proofs of Theorem 2.6 and Corollary 2.8 given in Appndix A. The corresponding results for the tests based on statistics $\hat{Q}$ and $\hat{P}$ defined in (4.1) read as follows

$$\max_{k=1}^{Tm} \frac{\hat{Q}_m(k)}{w_\gamma(k/m)} \overset{\mathcal{D}}{\Longrightarrow} \max_{0 < t < T/(T+1)} \frac{|W(t)|}{\max\{t^\gamma, \varepsilon\}} =: L_{2,\gamma}(T) \tag{5.4}$$

and

$$\max_{k=1}^{Tm} \frac{\hat{P}_m(k)}{w_\gamma(k/m)} \overset{\mathcal{D}}{\Longrightarrow} \max_{0 < t < T/(T+1)} \max_{0 \leq s \leq t} \frac{1}{\max\{t^\gamma, \varepsilon\}} \left| W(t) - \frac{1-t}{1-s} W(s) \right| =: L_{3,\gamma}(T) \ . \tag{5.5}$$

To complete the discussion on closed-end scenarios we will display a small batch of simulation results for the detection of changes in the mean as described in Section 3.1. For the sake of brevity, only the choice $T = 4$ is examined here [unpublished simulation results show similar outcomes for other choices of T]. The remaining simulation settings are the same as used for the simulation study presented in Section 4.1 and in Table 4 we display the necessary critical values defining the rejection regions for the different procedures.

The approximation of the nominal level under the null hypothesis is displayed in Table 5 and in Figures 6 and 7 the power of the different procedures with respect to change amount and change position for $\gamma = 0$ and $\gamma = 0.45$ is illustrated. The results are very similar to the open-end scenario discussed in Section 4 and confirm the superiority of tests based on the statistic $\hat{E}$.

$\gamma \setminus \alpha$	$L_{1,\gamma}(4)$			$L_{2,\gamma}(4)$			$L_{3,\gamma}(4)$		
	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
0	2.7080	2.1995	1.9811	2.5145	1.9826	1.7380	2.5572	2.0435	1.8019
0.25	2.9558	2.4345	2.2220	2.7602	2.2223	1.9799	2.8210	2.2986	2.0750
0.45	3.3850	2.9371	2.6994	3.2238	2.7398	2.4952	3.3156	2.8626	2.6274

Table 4: *(1-α)-quantiles of the distributions $L_{1,\gamma}(4)$, $L_{2,\gamma}(4)$ and $L_{3,\gamma}(4)$ for different choices of γ. The cutoff constant was set to $\varepsilon = 0$ and the dimension is $p = 1$.*

	(M1)			(M2)		
γ	$\hat{E}$	$\hat{Q}$	$\hat{P}$	$\hat{E}$	$\hat{Q}$	$\hat{P}$
0	5.2%	5.3%	5.3%	8.0%	7.3%	7.4%
0.25	5.4%	6.0%	5.8%	8.6%	7.5%	8.1%
0.45	4.9%	5.4%	4.5%	6.1%	6.4%	5.9%

Table 5: Type 1-Error for the closed-end procedures for a change in the mean based on the statistics $\hat{E}$, $\hat{Q}$ and $\hat{P}$ with a training data set of size $m = 200$ and a monitoring window of $T = 4$.
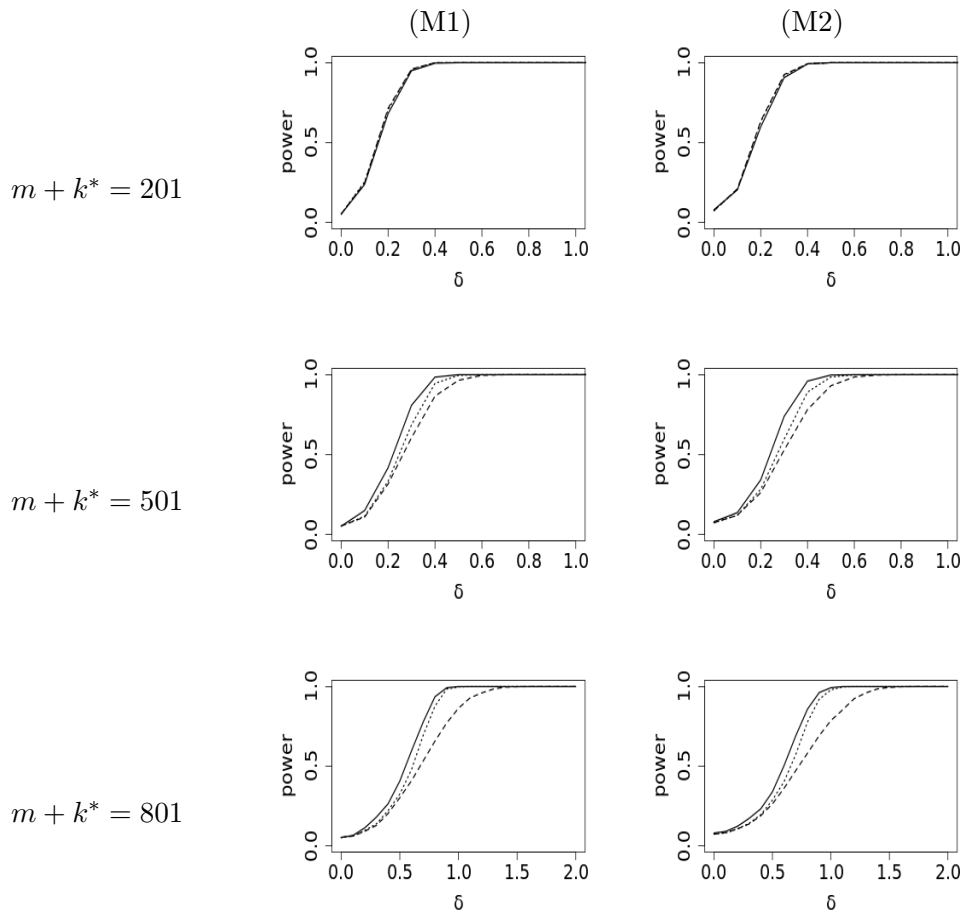
Figure 6: *Power of the (closed-end) monitoring procedures for a change in the mean based on the statistics $\hat{E}$ (solid line), $\hat{Q}$ (dashed line) and $\hat{P}$ (dotted line) with $\gamma = 0$, $m = 200$ and $T = 4$.*
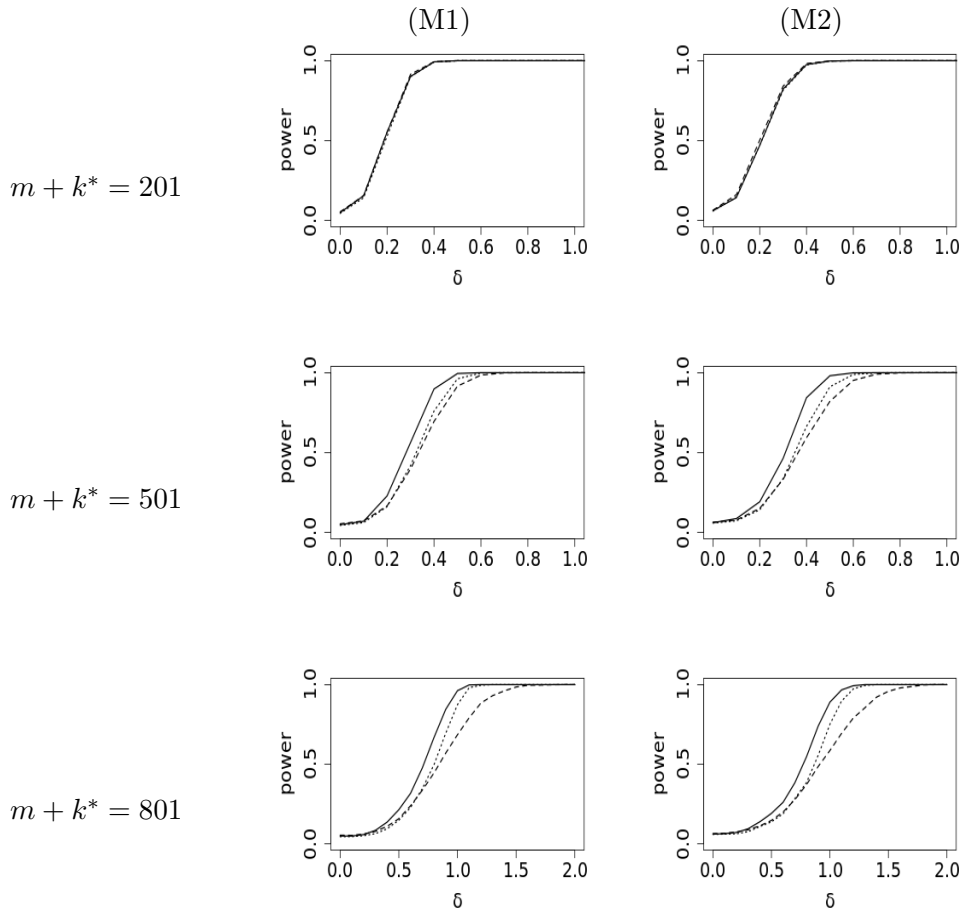
Figure 7: *Power of the (closed-end) monitoring procedures for a change in the mean based on the statistics $\hat{E}$ (solid line), $\hat{Q}$ (dashed line) and $\hat{P}$ (dotted line) with $\gamma = 0.45$, $m = 200$ and $T = 4$.*

6 Two applications

In this section, we discuss some practical issues that arise when our methodology is applied to monitor for changes in linear models. As a summary of our recommendations, we provide Algorithm 1 and two elaborated examples related to the United Kingdom European Union membership referendum 2016. We consider the linear model

$$Y_t = \beta_1 P_{1,t} + \beta_2 P_{2,t} + \varepsilon_t , \tag{6.1}$$

where Y_t is a real-valued response and $(P_{1,t}, P_{2,t})$ is a two-dimensional predictor, which is a special case of the linear model considered in Section 3.2.

As most monitoring procedures, our approach requires a stable segment of m observations in which no changes have yet happened. To ensure that a candidate for a stable segment is suitable, we suggest to make use of well known retrospective change point test methodology. A comprehensive introduction and summary for this topic can, for example, be found in Csörgő and Horváth (1997). For the sake of completeness details of the retrospective test that is employed as part of our analysis are provided in Section B of the appendix.

In light of our asymptotic arguments in Section 2, we require that m will not be chosen smaller than a minimum segment size, say $m_{\min}$, to be specified by the user. In practice, the observations of times $t = 1, \ldots, m$ are then used as (a first candidate for) a stable segment and monitoring starts not earlier than at time $m + 1$; see line 2 of Algorithm 1.

Before monitoring begins, we apply the retrospective test to the candidate for a stable segment and use it as such only if the test does not reject the null hypothesis of no change. Otherwise, if there is evidence of a change, we record the change point as 'found during setup', and continue in this manner with the next $m_{\min}$ observations onwards from the change that we identified until a segment is found where the retrospective test does not reject. Obtaining the stable segment for monitoring is described in lines 4–10 of Algorithm 1.

Once a stable segment is available monitoring starts: The detector is updated for every incoming observation, namely $(Y_t, P_{1,t}, P_{2,t})$, and a decision is made whether to reject the null hypothesis and stop the procedure or to continue monitoring with the subsequent observation. This is described in lines 11–13 of Algorithm 1.

After a change has been detected, the monitoring phase ends and an estimate for the time of the change is obtained via the formula

$$\hat{\ell} = m + \arg\max_{j=0}^{k-1} m^{-3/2} j(k-j) \left\| \frac{1}{m+j} \sum_{t=s}^{s+m+j-1} Y_t P_t - \frac{1}{k-j} \sum_{t=s+m+j}^{s+m+k-1} Y_t P_t \right\|_{\hat{\Gamma}^{-1}}, \tag{6.2}$$

where $P_t = (P_{1,t}, P_{2,t})^\top$. The objective function in (6.2) resembles the one in the detector

29

defined in (3.7), but we use the weights of the retrospective tests defined in (B.1). Note that - once monitoring has stopped - the problem of estimating the change time is of retrospective nature with the constraint that the first m observations are *known* as stable. The estimated time of change $\hat{\ell}$ is then labeled as 'found during monitoring' and recorded.

For the next monitoring phase the observations from the change point that was just identified to the last observation previously monitored are considered as a candidate for the next stable segment. In the case where this set would have fewer than $m_{\min}$ elements it is enlarged by adding incoming observations until $m_{\min}$ observations are available. This is described in lines 14–16 of Algorithm 1. The procedure then goes back to line 4 where the retrospective test is applied to assess the suitability of the candidate to be the next stable segment. Note that, if a change is identified in the candidate for the next stable segment, a new candidate is chosen to be the observations starting with that identified change point to the time where monitoring has last stopped or $m_{\min}$ observations onwards from the identified change in case that the candidate would otherwise have fewer than $m_{\min}$ observations; see line 8 of Algorithm 1.

In the remaining part of this section, we present the outcomes of the statistical analysis by Algorithm 1 for two data sets related to the United Kingdom (UK) European Union (EU) membership referendum, which took place on 23 June 2016. For our analysis we chose the significance levels to be $\alpha = 0.1$ and the threshold function w_0, as defined in (2.13). All data used was obtained from https://www.ariva.de on 6 April 2019.

As our first example, we consider the relation of UK's currency, Pound Sterling (GBP), to the Eurozone's currency, the Euro (EUR), and Switzerland's currency, the Swiss franc (CHF). More precisely, we consider daily log returns of the exchange rate of GBP to the United States dollar (USD) as a response Y_t of a linear model as described in (6.1). As predictors we now consider the log returns of EUR to USD ($P_{1,t}$) and CHF to USD ($P_{2,t}$). A graphical representation of the exchange rates and associated log returns for the period from Januar 2016 to March 2019 can be seen in Figure 8. The outcomes of Algorithm 1 are presented visually in the graphs. We chose $m_{\min} := 30$, which corresponds to six weeks of data. No evidence was found for the set of the first 30 observation (6 Jan 2016 to 16 Feb 2016) to contain a change and so it was used as the stable segment for monitoring. The monitoring went on until 30 Jun 2016 and 9 Jun 2016 was obtained as an estimate for the time of change. In the next step a retrospective test is performed to verify whether the observations from 9 Jun 2016 to 20 Jul 2016 contain any changes. No evidence for a change was found. The monitoring procedure is restarted with these 30 observations as the stable segment and monitoring continues until it is stopped on 21 Nov 2016 and the 15 Aug 2016 is identified as a time for a possible change. The period from 15 Aug to 21 Nov 2016 is considered as a candidate for a stable segment and no evidence against this is found. The 71 observation in this segment are used for the third iteration of monitoring,

Data: Univariate responses Y_t and bivariate predictors $(P_{1,t}, P_{2,t})$;

Input: Minimum size $m_{\min}$ for the stable segment, γ to be used for the threshold function w_γ, significance level α;

Output: A sequence of change points $c_1, c_2, \ldots$ and labels $T_1, T_2, \ldots$ that indicate whether the changes were identified by the retrospective test during setup (R) or by the sequential procedure during monitoring (S);

1 begin

2 $m := m_{\min}$, $i := 1$, $s := 1$;

3   **repeat**

       // Find stable segment

4   **while** *retrospective test does reject* **do**

5   Apply test from Section B to $(Y_t, P_{1,t}, P_{2,t})$, $t = s, \ldots, s + m - 1$;

6   **if** *test rejected* **then**

7   Record the change point: $c_i := s + j^* - 1$, $T_i := R$, $i++$;

8   Prepare the next test: $s := j^*$, $m := \max\{j^* - s, m_{\min}\}$;

9   **end**

10   **end**

       // Monitor for changes

11   **repeat**

12   Compute detector $\hat{E}_m(k)$ from $(Y_t, P_{1,t}, P_{2,t})$, $t = s, \ldots, s + m + k$;

13   **until** $\hat{E}_m(k)$ *exceeds the critical curve* $c_\alpha w_\gamma(k/m)$;

       // Estimate time of change

14   Compute $\hat{\ell}$ defined in (6.2);

15   Record the change point: $c_i := s + \hat{\ell}$, $T_i := S$, $i++$;

16   Prepare the restart: $m := \max\{s + m + k - c_i + 1, m_{\min}\}$, $s := c_i$;

17   **until** *terminated by user*;

18 end

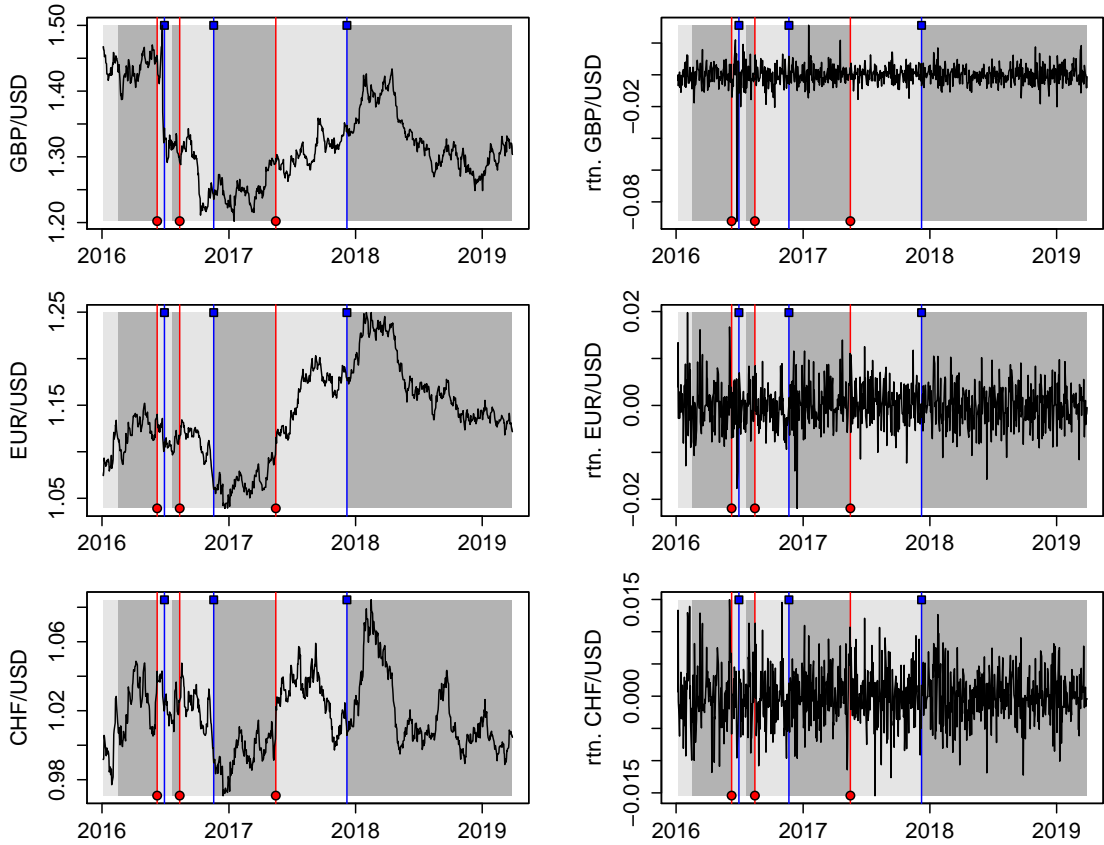**Algorithm 1:** Procedure for repeated monitoring.

Figure 8: *Results of Algorithm 1 applied to log returns of three foreign exchange rates. Response: GBP/USD (top), predictors: EUR/USD (middle) and CHF/USD (bottom). Settings for Algorithm 1 were $m_{\min} := 30$, $\alpha = 0.1$, $\gamma = 0$. Shaded areas indicate observations that were used as the stable segment (light gray) or monitored for changes (dark gray). Vertical lines indicate times the sequential procedure rejected (blue box) or the estimate time of change (red circle). The estimated times of the change points (times the sequential procedure stopped in brackets) were: 2016-06-09 (2016-06-30), 2016-08-15 (2016-11-21), and 2017-05-17 (2017-12-08).*

which continues until 8 Dec 2017, when 17 May 2017 is identified as a possible time of a change. Finally, a retrospective test is applied to the observations from 17 May 2017 to 8 Dec 2017 and as no evidence of a change is found monitoring commences with those 148 observations as a stable set without stopping until the end of the available data.

As our second example, we consider the relation of the UK's market to that of the United States (US) and the EU. More precisely, we consider daily log returns of the FTSE 100, a share index of the 100 companies listed on the London Stock Exchange with the highest market capitalization, as a response Y_t of the linear model described in (6.1). As predictors we consider the log returns of two similarly constructed indices that are related to the US and EU markets, namely the S&P 500 ($P_{1,t}$) and the EuroStoxx 50 ($P_{2,t}$). A graphical representation of the prices and log returns for the period from January 2016 to March
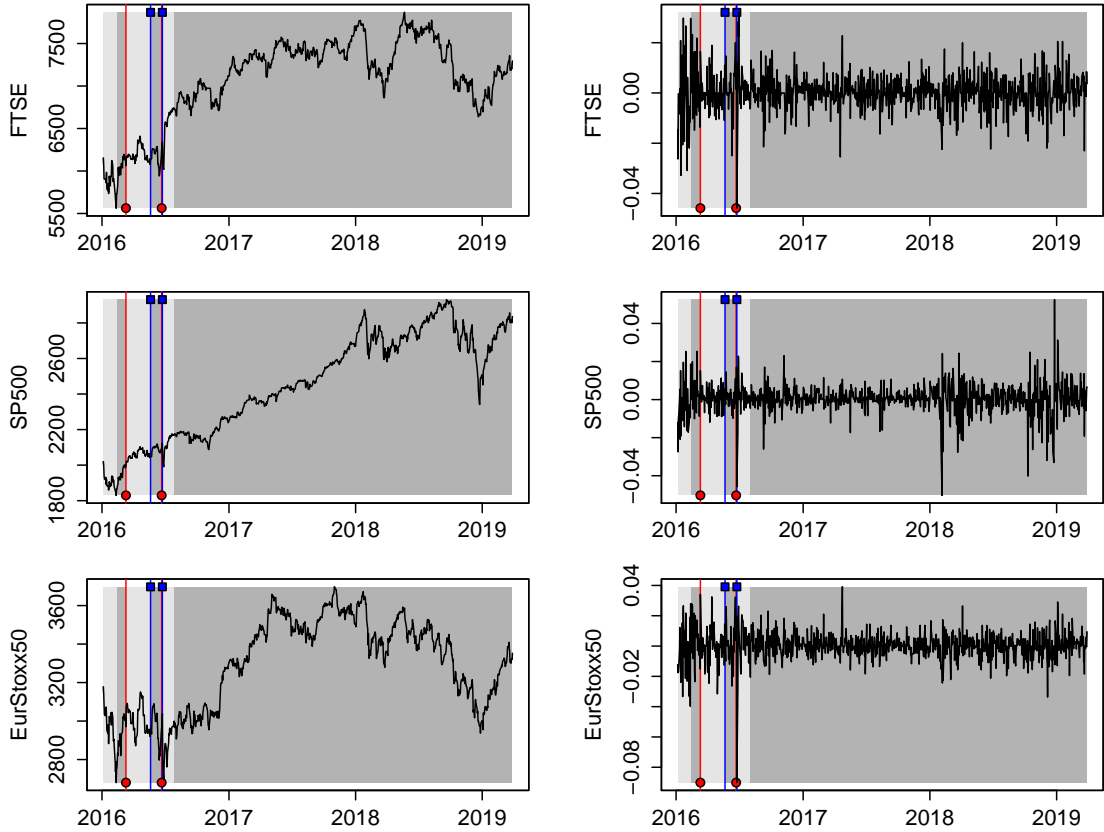
Figure 9: *Results of Algorithm 1 applied to log returns of three market indices. Response: FTSE 100 (top), predictors: S&P 500 (middle) and EuroStoxx 50 (bottom). Settings for Algorithm 1 were $m_{\min} := 30$, $\alpha = 0.1$, $\gamma = 0$. Shaded areas indicate observations that were used as the stable segment (light gray) or monitored for changes (dark gray). Vertical lines indicate times the sequential procedure rejected (blue box) or the estimate time of change (red circle). The estimated times of the change points (times the sequential procedure stopped in brackets) were: 2016-03-10 (2016-05-20), and 2016-06-21 (2016-06-23).*

2019 can be seen in Figure 9. The outcomes of Algorithm 1 are presented visually in the graphs. We chose $m_{\min} := 30$. As the retrospective test affirms that the first 30 observation (5 Jan 2016 to 12 Feb 2016) are stable, they are employed as the stable segment for the first iteration of monitoring. The monitoring went on until 20 May 2016, when the sequential procedure stopped and returned 10 Mar 2016 as the estimated time of the change. The retrospective test identified no changes for the 55 days from 10 Mar 2016 to 20 May 2016 and the second iteration of monitoring then went on until 23 Jun 2016, the day of the UK's EU referendum. The estimated time for a change point was 21 Jun 2016. The retrospective test found no evidence for the period from the 21 Jun 2016 to the 28 Jul 2016, so these 30 days of observations are used as the stable segment with which monitoring continues without stopping until 29 March 2019 (the end of the available data).

7 Summary and outlook

In this paper we developed a new monitoring scheme for change point detection in a parameter of multivariate time series which is applicable in an open-end scenario. Compared to the commonly used methods we replace the estimator of the parameter from the initial sample $X_1, \ldots, X_m$ by an estimator from the sample $X_1, \ldots, X_{m+j}$. We then compare this estimator with the estimator from the sample $X_{m+j+1}, \ldots, X_{m+k}$ for every $j = 0, \ldots, k-1$, For the new statistic the asymptotic distribution under the null hypothesis and the consistency of a corresponding test, which controls the type-1 error, are established. By considering a common class of thresholds w_γ defined in (2.13) the limit reduces to an elementary distribution, for which quantiles can be obtained by straightforward Monte Carlo simulations. Finally, we demonstrate by a comprehensive simulation study that the new monitoring scheme is superior (in terms of testing power) to a benchmark consisting of common methods proposed in the literature. The new statistic can also be used in closed-end scenarios, for which the same superiority in power is observed.

A very challenging subject for future research is the characterization of the asymptotic distribution for the stopping times based on the statistic $\hat{E}$ defined in (2.2). Corresponding results are already known for the methods based on $\hat{Q}$ and $\hat{P}$, see Aue and Horváth (2004) and Fremdt (2014a), respectively. Moreover, as the finite sample properties depend sensitively on the efficient estimation of the long-run variance it is of interest to develop a concept of self-normalization [see Shao and Zhang (2010)], which is applicable in an open-end scenario.

References

Anatolyev, S. and Kosenok, G. (2018). Sequential Testing with Uniformly Distributed Size. *Journal of Time Series Econometrics*, 10(2):1–22.

Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–58.

Aue, A., Dienes, C., Fremdt, S., and Steinebach, J. (2015). Reaction times of monitoring schemes for arma time series. *Bernoulli*, 21(2):1238–1259.

Aue, A., Hörmann, S., Horváth, L., Hušková, M., and Steinebach, J. G. (2012). Sequential testing for the stability of high-frequency portfolio betas. *Econometric Theory*, 28(4):804–837.

Aue, A. and Horváth, L. (2004). Delay time in sequential detection of change. *Statistics & Probability Letters*, 67(3):221–231.

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 Kokoszka, P. (2006). Change-point monitoring in linear models. *The Econometrics Journal*, 9(3):373–403.

Aue, A., Horváth, L., and Reimherr, M. L. (2009). Delay times of sequential procedures for multiple time series regression models. *Journal of Econometrics*, 149(2):174 – 190.

Aue, A., Hörmann, S., Horváth, L., and Hušková, M. (2014). Dependent functional linear models with applications to monitoring structural change. *Statistica Sinica*, 24(3):1043–1073.

Chen, Z. and Tian, Z. (2010). Modified procedures for change point monitoring in linear models. *Mathematics and Computers in Simulation*, 81(1):62 – 75.

Chu, C.-S. J., Stinchcombe, M., and White, H. (1996). Monitoring structural change. *Econometrica*, 64(5):1045–1065.

Csörgö, M. and Horváth, L. (1997). *Limit theorems in change-point analysis*. Wiley series in probability and statistics. Wiley.

Dette, H. and Gösmann, J. (2018). A likelihood ratio approach to sequential change point detection in a genereal class of parameters. *To appear in: Journal of the American Statistical Association; https://arxiv.org/abs/1802.07696*.

Fremdt, S. (2014a). Asymptotic distribution of the delay time in page's sequential procedure. *Journal of Statistical Planning and Inference*, 145:74 – 91.

Fremdt, S. (2014b). Page's sequential procedure for change-point detection in time series regression. *Statistics: A Journal of Theoretical and Applied Statistics*, 48(1):1–28.

Hampel, F., Ronchetti, E., Rousseeuw, P., and Stahel, W. (1986). *Robust statistics: The approach based on influence functions*. Wiley series in probability and mathematical statistics. Probability and mathematical statistics. Wiley.

Hoga, Y. (2017). Monitoring multivariate time series. *Journal of Multivariate Analysis*, 155:105 – 121.

Horváth, L., Hušková, M., Kokoszka, P., and Steinebach, J. (2004). Monitoring changes in linear models. *Journal of Statistical Planning and Inference*, 126(1):225 – 251.

Horváth, L., Kühn, M., and Steinebach, J. (2008). On the performance of the fluctuation test for structural change. *Sequential Analysis*, 27(2):126–140.

Kirch, C. and Weber, S. (2018). Modified sequential change point procedures based on estimating functions. *Electron. J. Statist.*, 12(1):1579–1613.

Lai, T. L. (1995). Sequential changepoint detection in quality control and dynamical systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(4):613–658.

Lai, T. L. (2001). Sequential analysis: Some classical problems and new challenges. 11:303–408.

Leisch, F., Hornik, K., and Kuan, C.-M. (2000). Monitoring structural changes with the generalized fluctuation test. *Econometric Theory*, 16(6):835–854.

Otto, S. and Breitung, J. (2019). A likelihood ratio approach to sequential change point detection in a genereal class of parameters. *http://www.wisostat.uni-koeln.de/sites/statistik/BackwardCUSUM.pdf*.

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.

Pape, K., Wied, D., and Galeano, P. (2016). Monitoring multivariate variance changes. *Journal of Empirical Finance*, 39(Part A):54 – 68.

Schmitz, A. and Steinebach, J. G. (2010). *A note on the monitoring of changes in linear models with dependent errors*, pages 159–174. Springer Berlin Heidelberg, Berlin, Heidelberg.

Shao, X. and Zhang, X. (2010). Testing for change points in time series. *Journal of the American Statistical Association*, 105(491):1228–1240.

Shorack, G. and Wellner, J. (1986). *Empirical processes with applications to statistics*. Wiley series in probability and mathematical statistics: Probability and mathematical statistics. Wiley.

Wied, D. and Galeano, P. (2013). Monitoring correlation change in a sequence of random variables. *Journal of Statistical Planning and Inference*, 143(1):186 – 196.

Zeileis, A. (2004). Econometric computing with hc and hac covariance matrix estimators. *Journal of Statistical Software*, 11(10):1–17.

A Proofs of the results in Section 2

Proof of Theorem 2.6: In the proof we use the following extra notation. Define the statistic

$$E_m(k) = m^{-1/2} \max_{j=0}^{k-1} (k-j) \left\| \hat{\theta}_1^{m+j} - \hat{\theta}_{m+j+1}^{m+k} \right\|_{\Sigma^{-1}}, \tag{A.1}$$

where we have replaced the long-run variance estimator $\hat{\Sigma}$ by the (unknown) true long-run variance Σ in definition (2.2). Further define

$$\tilde{E}_m(k) = m^{-1/2} \max_{j=0}^{k-1} (k-j) \left\| \frac{1}{m+j} \sum_{t=1}^{m+j} \mathcal{IF}_t - \frac{1}{k-j} \sum_{t=m+j+1}^{m+k} \mathcal{IF}_t \right\|_{\Sigma^{-1}}, \tag{A.2}$$

where we have replaced the estimator $\hat{\theta}_i^j$ by corresponding averages of the influence function in (A.1). Finally, the triple $(\Omega, \mathcal{A}, \mathbb{P})$ will denote the underlying probability space.

The proof itself is now split into several Lemmas A.1 - A.5. The first Lemma shows that $\tilde{E}_m(k)$ and $E_m(k)$ are (asymptotically) equivalent. Lemma A.2 will approximate $E_m(k)$ by Brownian motions, while Lemma A.3 then yields a limit for this approximation. Lemma A.4 finishes the proof by plugging in the covariance estimator, meaning that $\hat{E}_m(k)$ and $E_m(k)$ are asymptotically equivalent. Finally, Lemma A.5 will establish the other representation of the limit distribution in (2.11). In each Lemma, we suppose that the assumptions of Theorem 2.6 are valid.

Lemma A.1 (Remove Remainders) *It holds that*

$$\sup_{k=1}^{\infty} \frac{1}{w(k/m)} \left| E_m(k) - \tilde{E}_m(k) \right| = o_{\mathbb{P}}(1)$$

as $m \to \infty$.

Proof. By the (reverse) triangle inequality and the linearization in (2.7) we obtain

$$\left| E_m(k) - \tilde{E}_m(k) \right|$$

$$\leq m^{-1/2} \max_{j=0}^{k-1} (k-j) \left\| \hat{\theta}_1^{m+j} - \hat{\theta}_{m+j+1}^{m+k} - \frac{1}{m+j} \sum_{t=1}^{m+j} \mathcal{IF}_t + \frac{1}{k-j} \sum_{t=m+j+1}^{m+k} \mathcal{IF}_t \right\|_{\Sigma^{-1}}$$

$$= m^{-1/2} \max_{j=0}^{k-1} (k-j) \left\| R_{1,m+j} - R_{m+j+1,m+k} \right\|_{\Sigma^{-1}},$$

where we used that θ_t is constant for the last equality. Next, we obtain that

$$\sup_{k=1}^{\infty} \frac{1}{m^{1/2}w(k/m)} \max_{j=0}^{k-1}(k-j)\big\|R_{m+j+1,m+k}\big\|_{\Sigma^{-1}}$$

$$= \sup_{k=1}^{\infty} \frac{(m+k)^{1/2}}{m^{1/2}w(k/m)} \max_{j=0}^{k-1} \frac{k-j}{(m+k)^{1/2}}\big\|R_{m+j+1,m+k}\big\|_{\Sigma^{-1}} \tag{A.3}$$

$$\leq \sup_{k=1}^{\infty} \frac{(m+k)^{1/2}}{m^{1/2}w(k/m)} \sup_{k=1}^{\infty} \max_{1\leq i<j\leq m+k} \frac{j-i+1}{(m+k)^{1/2}}\big\|R_{i,j}\big\|_{\Sigma^{-1}} .$$

Similar as in (A.3) it holds

$$\sup_{k=1}^{\infty} \frac{1}{m^{1/2}w(k/m)} \max_{j=0}^{k-1}(k-j)\big\|R_{1,m+j}\big\|_{\Sigma^{-1}}$$

$$\leq \sup_{k=1}^{\infty} \frac{m+k}{mw(k/m)} \sqrt{m} \sup_{k=1}^{\infty} \max_{j=0}^{k-1} \big\|R_{1,m+j}\big\|_{\Sigma^{-1}} \tag{A.4}$$

$$= \sup_{k=1}^{\infty} \frac{m+k}{mw(k/m)} \sup_{k=0}^{\infty} \sqrt{m}\big\|R_{1,m+k}\big\|_{\Sigma^{-1}}$$

Using Assumption 2.3 for the threshold w we obtain

$$\sup_{k=1}^{\infty} \frac{m+k}{mw(k/m)} = \sup_{k=1}^{\infty} \frac{1+\dfrac{k}{m}}{w(k/m)} \leq \sup_{t>0} \frac{1+t}{w(t)} < \infty . \tag{A.5}$$

and by similar arguments it holds that

$$\sup_{k=1}^{\infty} \frac{(m+k)^{1/2}}{\sqrt{m}w(k/m)} < \infty . \tag{A.6}$$

Further note that due to Assumption 2.4 with probability one

$$\sup_{k=1}^{\infty} \sqrt{m}\big|R_{1,m+k}\big| \leq \sup_{k=1}^{\infty} \sqrt{m+k}\big|R_{1,m+k}\big| = \sup_{k=m}^{\infty} \sqrt{k}\big|R_{1,k}\big|$$

$$\overset{j=k,i=1}{\leq} \sup_{k=m}^{\infty} \max_{1\leq i<j\leq k} \frac{j-i+1}{\sqrt{k}}\big|R_{i,j}\big| = o(1) . \tag{A.7}$$

Now combining (A.5), (A.6) and (A.7) the bounds derived in (A.3) and (A.4) are of order $o_{\mathbb{P}}(1)$, which finishes the proof of Lemma A.1. $\qquad\square$

For the proof of the next Lemma we can proceed (roughly) similar to the proof of Lemma 5.2 in Fremdt (2014b).

Lemma A.2 (Approximation with Brownian motions) *Define*

$$P_m(k) = \frac{1}{\sqrt{m}} \max_{j=0}^{k-1} \Big|W_{m,1}(k) - \frac{m+k}{m+j}W_{m,1}(j) + \frac{k-j}{m+j}W_{m,2}(m)\Big| ,$$

then

$$\sup_{k=1}^{\infty} \frac{1}{w(k/m)} \left| \tilde{E}_m(k) - P_m(k) \right| = o_{\mathbb{P}}(1)$$

as $m \to \infty$.

Proof. For the remainder of the proof let $W_{m,i}^{\Sigma} := \sqrt{\Sigma} W_{m,i}$ for $i = 1, 2$ and note that this implies

$$P_m(k) = \frac{1}{\sqrt{m}} \max_{j=0}^{k-1} \left\| W_{m,1}^{\Sigma}(k) - \frac{m+k}{m+j} W_{m,1}^{\Sigma}(j) + \frac{k-j}{m+j} W_{m,2}^{\Sigma}(m) \right\|_{\Sigma^{-1}}.$$

The last display and the (reverse) triangle inequality then yield

$$
\begin{aligned}
&\left| \tilde{E}_m(k) - P_m(k) \right| \\
&\leq \frac{1}{\sqrt{m}} \max_{j=0}^{k-1} \left\| \frac{k-j}{m+j} \sum_{t=1}^{m+j} \mathcal{IF}_t - \sum_{t=m+j+1}^{m+k} \mathcal{IF}_t \right. \\
&\qquad + \left. W_{m,1}^{\Sigma}(k) - \frac{m+k}{m+j} W_{m,1}^{\Sigma}(j) + \frac{k-j}{m+j} W_{m,2}^{\Sigma}(m) \right\|_{\Sigma^{-1}} \qquad (A.8) \\
&\leq \frac{1}{\sqrt{m}} \left\| \sum_{t=m+1}^{m+k} \mathcal{IF}_t - W_{m,1}^{\Sigma}(k) \right\|_{\Sigma^{-1}} + \frac{1}{\sqrt{m}} \max_{j=0}^{k-1} \frac{m+k}{m+j} \left\| \sum_{t=m+1}^{m+j} \mathcal{IF}_t - W_{m,1}^{\Sigma}(j) \right\|_{\Sigma^{-1}} \\
&\qquad + \frac{1}{\sqrt{m}} \max_{j=0}^{k-1} \frac{k-j}{m+j} \left\| \sum_{t=1}^{m} \mathcal{IF}_t - W_{m,2}^{\Sigma}(m) \right\|_{\Sigma^{-1}}.
\end{aligned}
$$

We will treat the three summands of the last display separately. Using the definition of the operator norm, we derive the following bound for the first summand

$$
\sup_{k=1}^{\infty} \frac{1}{w(k/m)\sqrt{m}} \left\| \sum_{t=m+1}^{m+k} \mathcal{IF}_t - W_{m,1}^{\Sigma}(k) \right\|_{\Sigma^{-1}}
$$

$$
\leq \sup_{k=1}^{\infty} \frac{k^{\xi}}{w(k/m)\sqrt{m}} \sup_{k=1}^{\infty} \frac{1}{k^{\xi}} \left\| \sum_{t=m+1}^{m+k} \mathcal{IF}_t - W_{m,1}^{\Sigma}(k) \right\|_{\Sigma^{-1}}
$$

By Assumption 2.2 and the estimate $\|x\|_{\Sigma^{-1}} \leq \|\Sigma^{-1}\|_{op}^{1/2} |x|$ for all $x \in \mathbb{R}^p$, the second factor is of order $\mathcal{O}_{\mathbb{P}}(1)$. Since w has a lower bound ℓ_w we obtain for the first factor, that

$$
\sup_{k=1}^{m} \frac{k^{\xi}}{w(k/m)\sqrt{m}} \leq \sup_{k=1}^{m} \frac{k^{\xi}}{\ell_w \sqrt{m}} = \frac{m^{\xi-1/2}}{\ell_w} = o(1) \, ,
$$

$$
\sup_{k=m+1}^{\infty} \frac{k^{\xi}}{w(k/m)\sqrt{m}} \leq \sup_{k=m+1}^{\infty} \frac{1}{k^{1/2-\xi}} \sup_{k=1}^{\infty} \frac{k^{1/2}}{w(k/m)\sqrt{m}} \leq \frac{1}{m^{1/2-\xi}} \sup_{t>0} \frac{\sqrt{t}}{w(t)} = o(1) \, .
$$

Next, we can bound the second summand of the right-hand side in (A.8) by

$$\sup_{k=1}^{\infty} \frac{1}{w(k/m)\sqrt{m}} \max_{j=0}^{k-1} \frac{m+k}{m+j} \Big\| \sum_{t=m+1}^{m+j} \mathcal{IF}_t - W_{m,1}^{\Sigma}(j) \Big\|_{\Sigma^{-1}}$$

$$\leq \sup_{k=1}^{\infty} \frac{m+k}{w(k/m)\sqrt{m}} \max_{j=0}^{k-1} \frac{1}{j^{\xi} m^{1-\xi}} \Big\| \sum_{t=m+1}^{m+j} \mathcal{IF}_t - W_{m,1}^{\Sigma}(j) \Big\|_{\Sigma^{-1}}$$

$$\leq \sup_{k=1}^{\infty} \frac{(m+k)m^{\xi-1}}{w(k/m)\sqrt{m}} \sup_{k=1}^{\infty} \frac{1}{k^{\xi}} \Big\| \sum_{t=m+1}^{m+k} \mathcal{IF}_t - W_{m,1}^{\Sigma}(k) \Big\|_{\Sigma^{-1}},$$

where we used that $(m+j) = (m+j)^{\xi}(m+j)^{1-\xi} \geq j^{\xi} m^{1-\xi}$. Using again Assumption 2.2, the second factor in the last display is of order $\mathcal{O}_{\mathbb{P}}(1)$. Moreover, following the idea of the proof of Lemma 3 in Aue et al. (2006) it holds that

$$
\begin{aligned}
\sup_{k=1}^{m} \frac{(m+k)m^{\xi-1}}{w(k/m)\sqrt{m}} &\leq \sup_{k=1}^{m} \frac{(m+k)m^{\xi-3/2}}{\ell_w} = \frac{2m^{\xi-1/2}}{\ell_w} = o(1), \\
\sup_{k=m+1}^{\infty} \frac{(m+k)m^{\xi-1}}{w(k/m)\sqrt{m}} &= m^{\xi-1/2} \sup_{k=m+1}^{\infty} \frac{1+k/m}{w(k/m)} \leq m^{\xi-1/2} \sup_{t>1} \frac{1+t}{w(t)} = o(1)
\end{aligned}
\tag{A.9}
$$

and it remains to treat the third summand of the right-hand side in (A.8), which can be bounded by

$$\sup_{k=1}^{\infty} \frac{1}{w(k/m)\sqrt{m}} \max_{j=0}^{k-1} \frac{k-j}{m+j} \Big\| \sum_{t=1}^{m} \mathcal{IF}_t - W_{m,2}^{\Sigma}(m) \Big\|_{\Sigma^{-1}}$$

$$\leq \sup_{k=1}^{\infty} \frac{1}{w(k/m)\sqrt{m}} \frac{k}{m} \Big\| \sum_{t=1}^{m} \mathcal{IF}_t - W_{m,2}^{\Sigma}(m) \Big\|_{\Sigma^{-1}}$$

$$\leq \sup_{k=1}^{\infty} \frac{km^{\xi-1}}{w(k/m)\sqrt{m}} \frac{1}{m^{\xi}} \Big\| \sum_{t=1}^{m} \mathcal{IF}_t - W_{m,2}^{\Sigma}(m) \Big\|_{\Sigma^{-1}}.$$

Using Assumption 2.2 and the arguments in (A.9) this term is of order $o_{\mathbb{P}}(1)$, which finishes the proof of Lemma A.2. $\qquad \square$

Lemma A.3 (Obtain limit process) *The following weak convergence holds*

$$\sup_{k=1}^{\infty} \frac{P_m(k)}{w(k/m)} \overset{\mathcal{D}}{\Longrightarrow} \sup_{0 \leq t < \infty} \max_{0 \leq s \leq t} \frac{1}{w(t)} \Big| W_1(t) - \frac{1+t}{1+s} W_1(s) + \frac{t-s}{1+s} W_2(1) \Big|$$

as $m \to \infty$, where W_1 and W_2 denote independent, p-dimensional standard Brownian motions.

Proof. First note that due to the scaling properties of the Brownian motion [see for example

Chapter 2 of Shorack and Wellner (1986)], it holds in distribution that

$$P_m(k) \overset{\mathcal{D}}{=} \max_{j=0}^{k-1} \left| W_1(k/m) - \frac{1+k/m}{1+j/m} W_1(j/m) + \frac{k/m - j/m}{1+j/m} W_2(1) \right| := \widetilde{P}_m(k) \qquad \text{(A.10)}$$

and so within the proof we will - without loss of generality - only consider $\widetilde{P}_m(k)$. Additionally, we define the processes

$$L^{(1)}(s,t) := \frac{1}{w(t)} \left| \frac{t+1}{s+1} W_1(s) - W_1(t) \right| ,$$

$$L^{(2)}(s,t) := \frac{1}{w(t)} \frac{t-s}{s+1} |W_2(1)| .$$

We will show that $L^{(i)}(s,t)$ for $i = 1,2$ are uniformly continuous on $\mathbb{R}_\Delta^+ = \{(s,t) \in \mathbb{R}^2 \mid 0 \leq s \leq t\}$ with probability one.

Case $L^{(1)}$: In the following let $\varepsilon > 0$ be fixed but arbitrary. For almost every $\omega \in \Omega$ the law of iterated logarithm implies that we can choose a $C = C(\varepsilon, \omega) > 0$ sufficiently large such that

$$\sup_{t \geq C} \frac{|W_1(t)|}{t} < \frac{\varepsilon}{8B} , \qquad \text{(A.11)}$$

with B defined by

$$B := \sup_{t > 0} \frac{t+2}{w(t)} < \infty .$$

Depending on C, we can split $\mathbb{R}_\Delta^+$ into the (overlapping) sets

$$\mathbb{R}_\Delta^+ = \mathcal{M}_1(C) \cup \mathcal{M}_2(C) \cup \mathcal{M}_3(C) , \qquad \text{(A.12)}$$

where we use the definitions

$$\mathcal{M}_1(C) = \mathbb{R}_\Delta^+ \cap [0, C+1]^2 ,$$
$$\mathcal{M}_2(C) = \mathbb{R}_\Delta^+ \cap [0, C+1] \times [C, \infty) ,$$
$$\mathcal{M}_3(C) = \mathbb{R}_\Delta^+ \cap [C, \infty)^2 .$$

Further let d denote the maximum distance that is

$$d\big((s_1, t_1), (s_2, t_2)\big) = \max \big\{ |s_1 - s_2|, |t_1 - t_2| \big\} .$$

Note that by construction of the decomposition in (A.12), whenever $d((s_1, t_1), (s_2, t_2)) < \delta$ for sufficiently small $\delta > 0$, then there's $j \in 1, 2, 3$, such that both pairs are in the same

subset $\mathcal{M}_j(C)$. Thus the a.s. uniform continuity of $L^{(1)}$ follows if we can choose $\delta > 0$ sufficiently small, such that

$$\sup_{\substack{d((s_1,t_1),(s_2,t_2))<\delta \\ (s_1,t_1),(s_2,t_2)\in\mathcal{M}_j(C)}} \left| L^{(1)}(s_1,t_1) - L^{(1)}(s_2,t_2) \right| < \varepsilon \tag{A.13}$$

for $j = 1, 2, 3$. In the following, we will treat each subset separately.

Set $\mathcal{M}_1(C)$: As this set is compact, the (ordinary) almost sure continuity of $L^{(1)}$ already implies that (A.13) holds for $j = 1$ and $\delta > 0$ sufficiently small.

Set $\mathcal{M}_2(C)$: We have the following bound

$$\sup_{\substack{d((s_1,t_1),(s_2,t_2))<\delta \\ (s_1,t_1),(s_2,t_2)\in\mathcal{M}_2(C)}} \left| L^{(1)}(s_1,t_1) - L^{(1)}(s_2,t_2) \right|$$

$$\leq \sup_{\substack{|t_1-t_2|<\delta \\ 0\leq s\leq C+1,\ t_1,t_2>C}} \left| L^{(1)}(s,t_1) - L^{(1)}(s,t_2) \right| + \sup_{\substack{|s_1-s_2|<\delta \\ 0\leq s_1,s_2\leq C+1,\ t>C}} \left| L^{(1)}(s_1,t) - L^{(1)}(s_2,t) \right|. \tag{A.14}$$

We will treat both summands of the last display individually. The first one can be bounded again as follows

$$\sup_{\substack{|t_1-t_2|<\delta \\ 0\leq s\leq C+1,\ t_1,t_2>C}} \left| L^{(1)}(s,t_1) - L^{(1)}(s,t_2) \right|$$

$$\leq \sup_{\substack{|t_1-t_2|<\delta \\ 0\leq s\leq C+1,\ t_1,t_2>C}} \left| \frac{1}{w(t_1)} - \frac{1}{w(t_2)} \right| \left| \frac{t_1+1}{s+1}W_1(s) - W_1(t_1) \right| \tag{A.15}$$

$$+ \frac{1}{w(t_2)}\left| \frac{t_1-t_2}{s+1}W_1(s) - W_1(t_1) + W_1(t_2) \right|$$

and again we will treat both terms separately. For the first term note that

$$\sup_{\substack{|t_1-t_2|<\delta \\ 0\leq s\leq C+1,\ t_1,t_2>C}} \frac{|w(t_2)-w(t_1)|}{\ell_w}\frac{t_1+1}{w(t_1)}\left| \frac{W_1(s)}{s+1} - \frac{W_1(t_1)}{t_1+1} \right|$$

$$\leq 2 \sup_{|t_1-t_2|<\delta} \frac{|w(t_2)-w(t_1)|}{\ell_w} \sup_{t\geq C}\frac{t+1}{w(t)} \sup_{t\geq 0}\frac{|W_1(t)|}{t+1},$$

and since w is uniformly continuous this expression is smaller than $\varepsilon/3$ for sufficiently small

δ. For the second term of the right-hand side of (A.15), note that we have the bound

$$\sup_{\substack{|t_1-t_2|<\delta \\ 0\leq s\leq C+1,\ t_1,t_2>C}} \frac{1}{w(t_2)}\left|\frac{t_1-t_2}{s+1}W_1(s)-W_1(t_1)+W_1(t_2)\right|$$

$$\leq \sup_{\substack{|t_1-t_2|<\delta \\ 0\leq s\leq C+1,\ t_1,t_2>C}} \frac{|t_1-t_2|}{w(t_2)}\frac{|W_1(s)|}{s+1}+\frac{t_1+1}{w(t_2)}\frac{|W_1(t_1)|}{t_1+1}+\frac{t_2+1}{w(t_2)}\frac{|W_1(t_2)|}{t_2+1}$$

$$\leq \frac{\delta}{\ell_w}\sup_{s\geq 0}\frac{|W_1(s)|}{s+1}+2\sup_{t>0}\frac{t+2}{w(t)}\sup_{t>C}\frac{|W_1(t)|}{t+1}\ ,$$

and by the choice of C in (A.11) and for sufficiently small δ this is bounded by $\varepsilon/3$. To complete the treatment of $\mathcal{M}_2(C)$ it only remains to examine the second term on the right-hand side of (A.14). We obtain that

$$\sup_{\substack{|s_1-s_2|<\delta \\ 0\leq s_1,s_2\leq C+1,\ t>C}} \left|L^{(1)}(s_1,t)-L^{(1)}(s_2,t)\right|$$

$$\leq \sup_{t>0}\frac{t+1}{w(t)}\sup_{\substack{|s_1-s_2|<\delta \\ 0\leq s_1,s_2\leq C+1}}\left|\frac{W_1(s_1)}{s_1+1}-\frac{W_1(s_2)}{s_2+1}\right|\ ,$$

which can be bounded by $\varepsilon/3$ for sufficently small δ since the first factor is a constant and the function $f(s)=W(s+1)/(s+1)$ is uniformly continuous on the compact set $[0,C+1]$ with probability one.

Set $\mathcal{M}_3(C)$: Note that

$$\sup_{\substack{d((s_1,t_1),(s_2,t_2))<\delta \\ (s_1,t_1),(s_2,t_2)\in\mathcal{M}_3(C)}} \left|L^{(1)}(s_1,t_1)-L^{(1)}(s_2,t_2)\right|$$

$$\leq 2\sup_{s,t>C}|L^{(1)}(s,t)|\leq 2\sup_{t>0}\frac{t+1}{w(t)}\sup_{s,t>C}\left|\frac{W_1(s)}{s+1}-\frac{W_1(t)}{t+1}\right|\leq 4\sup_{t>0}\frac{t+1}{w(t)}\sup_{t\geq C}\frac{|W_1(t)|}{t}<\varepsilon\ ,$$

where we used the choice of C in (A.11) for the last estimate.

This completes the third case and so the almost sure uniform continuity of $L^{(1)}$ on the set $\mathbb{R}_\Delta^+$ is established.

Case $L^{(2)}$: Again let $\varepsilon>0$ and suppose that $d\big((s_1,t_1),(s_2,t_2)\big)<\delta$. It holds that

$$\left|L^{(2)}(s_1,t_1)-L^{(2)}(s_2,t_2)\right|\leq\left|L^{(2)}(s_1,t_1)-L^{(2)}(s_2,t_1)\right|+\left|L^{(2)}(s_2,t_1)-L^{(2)}(s_2,t_2)\right|$$

<div align="right">(A.16)</div>

and note for the first summand of the last display that

$$\left|L^{(2)}(s_1,t_1) - L^{(2)}(s_2,t_1)\right| = \frac{|W_2(1)|}{w(t_1)}\left|\frac{t_1-s_1}{s_1+1} - \frac{t_1-s_2}{s_2+1}\right| = \frac{|W_2(1)|(t_1+1)}{w(t_1)}\left|\frac{s_2-s_1}{(s_1+1)(s_2+1)}\right|$$
$$\leq \frac{|W_2(1)|(t_1+1)}{w(t_1)}|s_2-s_1|$$

and by Assumption 2.3 the last term is smaller than $\varepsilon/2$ uniformly for all $t_1 > 0$ if $\delta > 0$ is chosen sufficiently small. It remains to examine the second summand of the right-hand side of (A.16). It holds that

$$|L^{(2)}(s_2,t_1) - L^{(2)}(s_2,t_2)| \leq \left|\frac{1}{w(t_1)} - \frac{1}{w(t_2)}\right|\frac{|t_1-s_2|}{s_2+1}|W_2(1)| + \frac{|W_2(1)|}{w(t_2)}\frac{|t_2-t_1|}{s_2+1}$$
$$\leq \frac{|w(t_1)-w(t_2)|}{w(t_1)w(t_2)}\frac{t_1+\delta}{s_2+1}|W_2(1)| + \frac{|W_2(1)|}{\ell_w}|t_2-t_1|$$
$$\leq \frac{|w(t_1)-w(t_2)|}{\ell_w}\frac{t_1+\delta}{w(t_1)}|W_2(1)| + \frac{|W_2(1)|}{\ell_w}|t_2-t_1| \, ,$$

where we used that $s_2 \leq t_2 \leq t_1 + \delta$ whenever $d\big((s_1,t_1),(s_2,t_2)\big) < \delta$. By Assumption 2.3 the last display is smaller than $\varepsilon/2$ whenever $\delta > 0$ is sufficiently small and so the almost sure continuity of $L^{(2)}$ is shown.

Finally, we can combine our observations to finish the proof. Note that by the results above, also the process $L(s,t) := L^{(1)}(s,t) + L^{(2)}(s,t)$ is uniformly continuous with probability one and further we have the identity

$$\sup_{1\leq t<\infty}\sup_{0\leq s\leq t} L\big(\lfloor s\rfloor/m, \lfloor t\rfloor/m\big) = \widetilde{\sup}_{k=1}^{\infty}\widetilde{P}_m(k)$$

and so the Lemma's claim now follows if we derive that (with probability one)

$$\sup_{1\leq t<\infty}\sup_{0\leq s\leq t} L\big(\lfloor s\rfloor/m, \lfloor t\rfloor/m\big) \underset{m\to\infty}{\Longrightarrow} \sup_{0\leq t<\infty}\sup_{0\leq s\leq t} L(s,t) \, .$$

First, observe that almost surely

$$\sup_{1\leq t<\infty}\sup_{0\leq s\leq t} L\big(\lfloor s\rfloor/m, \lfloor t\rfloor/m\big) = \sup_{0\leq t<\infty}\sup_{0\leq s\leq t} L\big(\lfloor s\rfloor/m, \lfloor t\rfloor/m\big) + o(1)$$

as $m \to \infty$. Next, we can use the almost sure uniform continuity of L since for arbitrary

$\varepsilon > 0$ and almost every $\omega \in \Omega$ we can choose sufficiently large $m = m(\varepsilon, \omega)$ such that

$$\left| \sup_{0 \leq t < \infty} \sup_{0 \leq s \leq t} L\big(\lfloor s \rfloor / m, \lfloor t \rfloor / m\big) - \sup_{0 \leq t < \infty} \sup_{0 \leq s \leq t} L\big(s, t\big) \right|$$

$$\leq \sup_{0 \leq t < \infty} \sup_{0 \leq s \leq t} \left| L\big(\lfloor s \rfloor / m, \lfloor t \rfloor / m\big) - L\big(s/m, t/m\big) \right|$$

$$\leq \sup_{\substack{d((s_1, t_1),(s_2, t_2)) < 1/m \\ (s_1, t_1),(s_2, t_2) \in \mathbb{R}_\Delta^+}} \left| L\big(s_1, t_1\big) - L\big(s_2, t_2\big) \right| < \varepsilon .$$

$\square$

Combining Lemma A.1, A.2 and A.3 we have already proven that

$$\sup_{k=1}^{\infty} \frac{E_m(k)}{w(k/m)} \overset{\mathcal{D}}{\Longrightarrow} \max_{0 \leq t < \infty} \max_{0 \leq s \leq t} \frac{1}{w(t)} \left| W_1(t) - \frac{1+t}{1+s} W_1(s) + \frac{t-s}{1+s} W_2(1) \right| , \qquad \text{(A.17)}$$

and it only remains to investigate the impact of the covariance estimator. Therefore the following Lemma finishes the proof of Theorem 2.6.

Lemma A.4 (Plug in of covariance estimator) *We have that*

$$\sup_{k=1}^{\infty} \frac{E_m(k)}{w(k/m)} - \sup_{k=1}^{\infty} \frac{\hat{E}_m(k)}{w(k/m)} = o_{\mathbb{P}}(1) .$$

Proof. Observe the bound

$$\left| \sup_{k=1}^{\infty} \frac{E_m(k)}{w(k/m)} - \sup_{k=1}^{\infty} \frac{\hat{E}_m(k)}{w(k/m)} \right|$$

$$\leq \sup_{k=1}^{\infty} \frac{1}{w(k/m)\sqrt{m}} \max_{j=0}^{k-1} (k-j) \left| (\hat{\theta}_1^{m+j} - \hat{\theta}_{m+j+1}^{m+k})^\top (\hat{\Sigma}_m^{-1} - \Sigma^{-1})(\hat{\theta}_1^{m+j} - \hat{\theta}_{m+j+1}^{m+k}) \right|^{1/2} . \tag{A.18}$$

Next note that for a symmetric matrix A and an arbitrary vector v the Cauchy-Schwarz inequality implies

$$|v^\top A v|^2 \leq |Av||v| \leq \|A\|_{op} |v|^2 ,$$

and we can bound (A.18) by

$$\left\| \hat{\Sigma}_m^{-1} - \Sigma^{-1} \right\|_{op}^{1/2} \sup_{k=1}^{\infty} \frac{1}{w(k/m)\sqrt{m}} \max_{j=0}^{k-1} (k-j) \left| \hat{\theta}_1^{m+j} - \hat{\theta}_{m+j+1}^{m+k} \right| . \tag{A.19}$$

Since $\hat{\Sigma}$ is a consistent estimator of Σ, an application of the continuous mapping theorem

45

yields

$$\left\|\hat{\Sigma}_m^{-1} - \Sigma^{-1}\right\|_{op} = o_{\mathbb{P}}(1) . \tag{A.20}$$

Next, the definition of the operator norm yields

$$\sup_{k=1}^{\infty} \frac{1}{w(k/m)\sqrt{m}} \max_{j=0}^{k-1} (k-j) \left|\hat{\theta}_1^{m+j} - \hat{\theta}_{m+j+1}^{m+k}\right|$$

$$\leq \|\Sigma^{1/2}\|_{op} \sup_{k=1}^{\infty} \frac{1}{w(k/m)\sqrt{m}} \max_{j=0}^{k-1} (k-j) \left\|\hat{\theta}_1^{m+j} - \hat{\theta}_{m+j+1}^{m+k}\right\|_{\Sigma^{-1}}$$

$$= \|\Sigma^{1/2}\|_{op} \sup_{k=1}^{\infty} \frac{E_m(k)}{w(k/m)} = \mathcal{O}_{\mathbb{P}}(1) .$$

Now a combination of (A.17) and (A.20) implies that the expression in (A.19) is of order $o_{\mathbb{P}}(1)$, which completes the proof of Lemma A.4 and thus also the proof of Theorem 2.6.

Combining Lemmas A.1, A.2, A.3 and A.4 we have now established that

$$\sup_{k=1}^{\infty} \frac{\hat{E}_m(k)}{w(k/m)} \overset{\mathcal{D}}{\Longrightarrow} \sup_{0 \leq t < \infty} \max_{0 \leq s \leq t} \frac{1}{w(t)} \left| W_1(t) - \frac{1+t}{1+s} W_1(s) + \frac{t-s}{1+s} W_2(1) \right|$$

and it remains to show the equality in distribution stated in (2.11).

Lemma A.5 (Simplify limit distribution) *It holds that*

$$\sup_{0 \leq t < \infty} \max_{0 \leq s \leq t} \frac{1}{w(t)} \left| W_1(t) - \frac{1+t}{1+s} W_1(s) + \frac{t-s}{1+s} W_2(1) \right|$$

$$\overset{\mathcal{D}}{=} \sup_{0 \leq t < \infty} \max_{0 \leq s \leq t} \frac{t+1}{w(t)} \left| W\left(\frac{t}{t+1}\right) - W\left(\frac{s}{s+1}\right) \right| ,$$

where W is a standard p-dimensional Brownian motion.

Proof. In the following let Z denote a standard Gaussian random variable, that is independent of W_1. Observe that

$$\frac{1}{w(t)} \left| W_1(t) - \frac{1+t}{1+s} W_1(s) + \frac{t-s}{1+s} W_2(1) \right|$$

$$= \frac{1}{w(t)(s+1)} \left| (s+1) W_1(t) - (t+1) W_1(s) + (t-s) W_2(1) \right|$$

$$\overset{\mathcal{D}}{=} \frac{1}{w(t)(s+1)} \left| (s+1) W(t) - (t+1) W(s) - (t-s) Z \right|$$

$$= \frac{1}{w(t)(s+1)} \left| (s+1) W(t) - (t+1) W(s) - (t-s) Z + stZ - stZ \right|$$

$$= \frac{1}{w(t)(s+1)} \left| (s+1)\{W(t) - tZ\} - (t+1)\{W(s) - sZ\} \right| . \tag{A.21}$$

Following Horváth et al. (2004), Fremdt (2014b), computing the covariance function implies the following identity (in distribution)

$$\left\{W(t) - tZ\right\}_{t\geq0} \overset{\mathcal{D}}{=} \left\{(1+t)W\left(\frac{t}{t+1}\right)\right\}_{t\geq0}.$$

Applying this to (A.21) yields

$$\frac{1}{w(t)(s+1)}\left|(t+1)\{W(s)-sZ\} - (s+1)\{W(t)-tZ\}\right|$$

$$\overset{\mathcal{D}}{=}\frac{1}{w(t)(s+1)}\left|(t+1)(s+1)W\left(\frac{s}{s+1}\right) - (t+1)(s+1)W\left(\frac{t}{t+1}\right)\right|$$

$$=\frac{1}{w(t)}\left|(t+1)W\left(\frac{s}{s+1}\right) - (t+1)W\left(\frac{t}{t+1}\right)\right|.$$

This completes the proof of Lemma A.5 and also of Theorem 2.6.

$\square$

Proof of Lemma 2.8: Using the definition $w(t) = (1+t)\max\left\{\left(\frac{t}{1+t}\right)^{\gamma}, \varepsilon\right\}$, we obtain that

$$\sup_{0\leq t<\infty}\max_{0\leq s\leq t}\frac{1}{w(t)}\left|(t+1)W\left(\frac{s}{s+1}\right) - (t+1)W\left(\frac{t}{t+1}\right)\right|$$

$$=\sup_{0\leq t<\infty}\max_{0\leq s\leq t}\frac{1}{\max\left\{\left(\frac{t}{1+t}\right)^{\gamma}, \varepsilon\right\}}\left|W\left(\frac{s}{s+1}\right) - W\left(\frac{t}{t+1}\right)\right|$$

$$=\sup_{0\leq t<\infty}\max_{0\leq s\leq t}\frac{1}{\max\left\{\left(\frac{t}{1+t}\right)^{\gamma}, \varepsilon\right\}}\left|W\left(\frac{s}{s+1}\right) - W\left(\frac{t}{t+1}\right)\right|$$

$$=\sup_{0\leq t<1}\max_{0\leq s\leq t}\frac{1}{\max\{t^{\gamma}, \varepsilon\}}\left|W(s) - W(t)\right|,$$

where we used that the mapping $x \mapsto x/(1+x)$ is bijective and increasing on the domain $[0,\infty)$ with co-domain $[0,1)$.

$\square$

Proof of Theorem 2.11: For m sufficiently large we have $k^* < m$. Then observe the lower bound

$$\sup_{k=1}^{\infty}\frac{\hat{E}_m(k)}{w(k/m)} = \sup_{k=1}^{\infty}\frac{1}{w(k/m)m^{1/2}}\max_{j=0}^{k-1}(k-j)\left\|\hat{\theta}_1^{m+j} - \hat{\theta}_{m+j+1}^{m+k}\right\|_{\hat{\Sigma}^{-1}}$$

$$\overset{j=k^*}{\geq}\sup_{k^*\leq k<\infty}\frac{1}{w(k/m)m^{1/2}}(k-k^*)\left\|\hat{\theta}_1^{m+k^*} - \hat{\theta}_{m+k^*+1}^{m+k}\right\|_{\hat{\Sigma}^{-1}}$$

$$\overset{k=m}{\geq}\frac{m-k^*}{m^{1/2}w(1)}\left\|\hat{\theta}_1^{m+k^*} - \hat{\theta}_{m+k^*+1}^{2m}\right\|_{\hat{\Sigma}^{-1}}.$$

The last display is bounded from below by

$$\frac{m-k^*}{m^{1/2}w(1)}\left\|\theta^{(1)}-\theta^{(2)}\right\|_{\hat{\Sigma}^{-1}} - \frac{m-k^*}{m^{1/2}w(1)}\left\|\hat{\theta}_1^{m+k^*}-\theta^{(1)}\right\|_{\hat{\Sigma}^{-1}}$$
$$- \frac{m-k^*}{m^{1/2}w(1)}\left\|\hat{\theta}_{m+k^*+1}^{2m}-\theta^{(2)}\right\|_{\hat{\Sigma}^{-1}} .$$

Using that $\hat{\Sigma}$ is (weakly) convergent with non-singular limit, the first summand above diverges to infinity and so it suffices to proof that the remaining two summands are of order $\mathcal{O}_{\mathbb{P}}(1)$. This follows from the linearization in equation (2.7) since this gives

$$\frac{m-k^*}{m^{1/2}w(1)}\big(\hat{\theta}_1^{m+k^*}-\theta^{(1)}\big)$$
$$= \frac{m-k^*}{m^{1/2}(m+k^*)w(1)}\sum_{t=1}^{m+k^*}\mathcal{IF}_t \ + \ \frac{m-k^*}{m^{1/2}w(1)}R_{1,m+k^*} = \mathcal{O}_{\mathbb{P}}(1)$$

and

$$\frac{m-k^*}{m^{1/2}w(1)}\big(\hat{\theta}_{m+k^*+1}^{2m}-\theta^{(2)}\big)$$
$$= \frac{1}{m^{1/2}w(1)}\sum_{t=m+k^*+1}^{2m}\mathcal{IF}_t \ + \ \frac{(m-k^*)}{m^{1/2}w(1)}R_{m+k^*+1,2m} = \mathcal{O}_{\mathbb{P}}(1) ,$$

where we also applied Assumption 2.9. Now the claim follows. $\qquad\square$

B Retrospective test for changes in linear models

Consider the setting outlined in Section 3.2. In this situation we would like to test for stability of the regression vectors in the initial sample, that is

$$H_0 : \beta_1 = \cdots = \beta_m .$$

Following the ideas stated in chapter 1 of Csörgö and Horváth (1997) it is sensible to employ the statistic

$$T_m = \frac{1}{m^{3/2}}\max_{j=1}^{m} j(m-j)\|\hat{\beta}_1^j - \hat{\beta}_{j+1}^m\|_{\hat{\Sigma}} , \tag{B.1}$$

where $\hat{\beta}$ was defined in (3.3). A sensible estimate for the time of the change is given by

$$j^* := \arg\max_{j=1}^{m} j(m-j)\|\hat{\beta}_1^j - \hat{\beta}_{j+1}^m\|_{\hat{\Sigma}} \tag{B.2}$$

Given the assumptions of Section 3.2 and using the theory of the reference one can verify that under the null hypothesis above

$$T_n \overset{\mathcal{D}}{\Longrightarrow} \max_{t \in [0,1]} |B(t)| \ ,$$

where B denotes a p-dimensional Brownian bridge (with independent components). Thus rejecting the null whenever

$$T_n > b_{1-\alpha}$$

with $b_{1-\alpha}$ denoting the $(1-\alpha)$ quantile of the limit distribution yields an appropriate testing procedure for the retrospective problem.