

Gösmann, Josua; Kley, Tobias; Dette, Holger

Article — Published Version

A new approach for open-end sequential change point monitoring

Journal of Time Series Analysis

Provided in Cooperation with:

John Wiley & Sons

Suggested Citation: Gösmann, Josua; Kley, Tobias; Dette, Holger (2021) : A new approach for open-end sequential change point monitoring, Journal of Time Series Analysis, ISSN 1467-9892, John Wiley & Sons, Ltd, Oxford, UK, Vol. 42, Iss. 1, pp. 63-84, <https://doi.org/10.1111/jtsa.12555>

This Version is available at:

<https://hdl.handle.net/10419/230290>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/4.0/>

ORIGINAL ARTICLE

A NEW APPROACH FOR OPEN-END SEQUENTIAL CHANGE POINT MONITORING

JOSUA GÖSMANN,^{a*} TOBIAS KLEY^a AND HOLGER DETTE^a

^a *Fakultät für Mathematik, Ruhr-Universität Bochum, Bochum, Germany*

^b *School of Mathematics, University of Bristol, Bristol, UK*

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. By means of a simulation study it is demonstrated that the new method outperforms the commonly used procedures with respect to power and the feasibility of our approach is illustrated by analyzing two data examples.

Received 23 September 2019; Accepted 27 July 2020

Keywords: Change point analysis; open-end procedures; sequential monitoring

JEL: C01; C22

MOS subject classification: 62L99; 62F03.

1. INTRODUCTION

Nowadays, nearly all fields of applications require sophisticated statistical modeling and statistical inference to draw scientific conclusions from the observed data. In many cases, data are 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*. Due to its universality, methods for change point analysis have a vast field of possible applications – ranging from natural sciences, like biology and meteorology, to humanities, like economics, finance, and 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 are 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 arrive 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* or *online change point detection*.

* Correspondence to: Josua Gösmann, Ruhr-Universität Bochum, Fakultät für Mathematik, 44780 Bochum, Germany. E-mail: josua.goesmann@ruhr-uni-bochum.de

In the major part of the 20th century the problem of sequential change point detection was tackled using procedures, which are optimized to have a minimal detection delay but do usually not control the probability of a false alarm (type I error). These methods are called control charts and a comprehensive review can be found in Lai (1995, 2001). 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 article 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

$$H_0 : \theta_1 = \dots = \theta_m = \theta_{m+1} = \theta_{m+2} = \dots, \quad (1.1)$$

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

$$H_1 : \exists k^* \in \mathbb{N} : \theta_1 = \dots = \theta_{m+k^*-1} \neq \theta_{m+k^*} = \theta_{m+k^*+1} = \dots \quad (1.2)$$

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. Since the seminal paper of Chu *et al.* (1996) several authors have worked in this area. Berkes *et al.* (2004) designed a detector for changes in the coefficient in the parameters of a GARCH-process. Horváth *et al.* (2004), Aue *et al.* (2006, 2009b, 2014), and Fremdt (2015) 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 (2020) 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

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

(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, 2009b, 2014), Schmitz and Steinebach (2010), or Hoga (2017). Other authors suggest using a function of the differences

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

(in dependence of k) and the corresponding procedures are usually called Page-CUSUM tests (see Fremdt, 2015; Aue, *et al.* 2015; or Kirch and Weber, 2018 among others). As an alternative we propose – following ideas of Dette

and Gösmann (2019) – a monitoring scheme based on a function of the differences

$$\{\hat{\theta}_1^{m+j} - \hat{\theta}_{m+j+1}^{m+k}\}_{j=0, \dots, k-1}. \quad (1.5)$$

A possible advantage of (1.5) over (1.3) is the screening for all potential 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.

This article 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 weight 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 (2015), 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. We present a small simulation study in Section 4, where we compare our approach to those developed by Horváth *et al.* (2004) and Fremdt (2015). 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. In Section 5 we illustrate our approach and compare it to other monitoring schemes by applying it to two examples where the parameter of a linear model of financial data is monitored around the time of the United Kingdom European Union membership referendum 2016. Finally, all proofs are deferred to the Supporting information (Gösmann *et al.*, 2020), in which we additionally provide some extra simulation results and briefly discuss how our statistic can be used in closed-end scenarios.

2. ASYMPTOTIC PROPERTIES

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

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

the empirical distribution function of observations X_i, \dots, 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, \dots, X_j .

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

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

where the statistic $\hat{\Sigma}_m$ denotes an estimator of the long-run variance (LRV) matrix Σ (defined in Assumption 2.3) 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 weight function, which is sometimes also called *threshold function*, say w . If

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

occurs, 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 2.7 to determine the constant $c(\alpha)$ involved in (2.3), such that the test keeps a nominal level of α (asymptotically as $m \rightarrow \infty$).

Remark 2.1 The statistic (2.2) is related to a detection scheme, which was recently proposed by Dette and Gösmann (2019) for the closed-end case, where monitoring ends with observation mT , for some $T \in \mathbb{N}$. These authors considered the statistic

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

and showed

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

where W denotes a p -dimensional Brownian motion and throughout this article the symbol \xRightarrow{D} denotes weak convergence (in the space under consideration). (To avoid confusion, note that in Dette and Gösmann (2019) the weight function was defined as $w = \frac{1}{\bar{w}}$ for some appropriate function \bar{w} .) However, this statistic cannot be considered in an open-end scenario for the typical weight functions considered in the literature satisfying $\limsup_{t \rightarrow \infty} tw(t) < \infty$ (in this case the limit on the right-hand side of (2.5) would be almost surely infinite for $T = \infty$). As weight functions satisfying $\limsup_{t \rightarrow \infty} t^2 w(t) < \infty$ will cause a loss in power as indicated 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). The remaining weight factor $(k-j)$ is retained as it allocates smaller weights to the case when the post-change estimator $\hat{\theta}_{m+j+1}^{m+k}$ contains greater uncertainty as j is close to k .

Remark 2.2 An essential disadvantage of closed-end scenarios as considered in Dette and Gösmann (2019) is the problem of choosing the end-point of monitoring before the procedure is launched. This problem drops out when open-end scenarios are employed, where monitoring can (theoretically) proceed forever if no change has been detected. Even if the statistical problems of closed- and open-end scenarios are naturally related, the reader should note, that the mathematical/technical access to both problem is completely different. In the closed-end case it is usually sufficient to assume the existence of functional central limit theorems (FCLTs) as the underlying time frame is compact (see for instance Aue *et al.*, 2012; Wied and Galeano, 2013; Pape *et al.*, 2016; Dette and Gösmann, 2019). To the best of the authors' knowledge, an FCLT is insufficient in the open-end case and one commonly assumes stronger, uniform stochastic approximations or combines an FCLT with Háyék-Réyni type inequalities (see also Section 2, Horváth *et al.*, 2004; Aue *et al.*, 2009a, 2009b; Fremdt 2014, 2015; Kirch and Weber, 2018).

To discuss the asymptotic properties of our approach, we require the following notation. We denote the non-negative reals by $\mathbb{R}_{\geq 0}$ and define $\mathbb{R}_+ := \mathbb{R}_{\geq 0} \setminus \{0\}$. The symbol $\xRightarrow{\mathbb{P}}$ denotes convergence in probability. The process $\{W(s)\}_{s \in [0, \infty)}$ will represent a standard p -dimensional Brownian motion with independent components. 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. By $\lfloor x \rfloor$ for $x \in \mathbb{R}$ we denote the largest integer smaller or equal to x . For the sake of a clear distinction we will employ $\sup_{i=1}^n a(i)$ 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

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

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

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

with asymptotically negligible remainder terms R_{ij} . 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 2.3 (Approximation) *The time series $\{X_t\}_{t \in \mathbb{Z}}$ is (strictly) stationary, such that $F_t = F$ for all $t \in \mathbb{Z}$. Furthermore, 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*

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

and

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

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

Assumption 2.4 (Weight function) *The weight function $w : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ is of the form*

$$w(t) = \tilde{w}(t) I\{t_w \leq t \leq T_w\} \quad (2.10)$$

for $t_w \geq 0$ and $T_w \in \mathbb{R}_+ \cup \{\infty\}$. Further $\tilde{w} : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_+$ is a positive continuous function and in case of $T_w = \infty$ it satisfies additionally

- (1) $\limsup_{t \rightarrow \infty} t \tilde{w}(t) < \infty$,
- (2) $1/\tilde{w}$ is uniformly continuous on $\mathbb{R}_{\geq 0}$.

Assumption 2.5 (Linearization) *The remainder terms in the linearization (2.7) satisfy*

$$\max_{\substack{i,j=1 \\ i < j}}^k \frac{(j-i+1)}{\sqrt{k}} |R_{i,j}| = o(1) \quad (2.11)$$

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

Remark 2.6 Let us give a brief explanation on the assumptions stated above.

- (i) Assumption 2.3 is a uniform invariance principle and frequently used in the (sequential) change point literature (see eg, Aue *et al.*, 2006 or Fremdt, 2015 among others). Following the lines of Aue *et al.* (2006) Assumption 2.3 can be verified by employing the multivariate strong approximation results derived by Eberlein (1986). This is already spelled out for augmented GARCH-processes in Lemma A.1 of Aue *et al.* (2006) for the one-dimensional case. Assumption 2.3 is stronger than a functional central limit theorem (FCLT), which is usually sufficient to work in a closed-end setup (see eg, Wied and Galeano, 2013; Pape *et al.*, 2016; or Dette and Gösmann, 2019). Another possible starting point to cope with open-end scenarios is an FCLT for any fixed time horizon together combined with Háyék-Rényi-Inequalities (see eg, Kirch and Weber, 2018 or Kirch and Stoeck, 2019). As this is less frequently used in the literature, we will remain with the other approach.
- (ii) Assumption 2.4 gives restrictions on the feasible set of weight functions, which are required for the existence of a (weak) limit derived in Theorem 2.7. The cutoffs defined in (2.10) serve only for technical purposes. By choosing $t_w > 0$ a delay at monitoring start is introduced, which can avoid problems with false alarms due to instability (see also Kirch and Weber, 2018). Selecting $T_w < \infty$ allows to additionally cover closed-end scenarios by our theory, which we briefly discuss in Section C of the Supporting information (Gösmann *et al.*, 2020). Note that in case of $t_w = 0$ and $T_w = \infty$ the cutoffs disappear, such that w and \tilde{w} coincide.
- (iii) It is worth mentioning that it is also possible to define the functions w, \tilde{w} on the smaller domain \mathbb{R}_+ , while additionally demanding that $\lim_{t \rightarrow 0} t^\gamma \tilde{w}(t) = 0$ for a constant $0 \leq \gamma < 1/2$. In this case, the assumption for the remainders in (2.11) has to be replaced by

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

which would have the upside to allow for an unbounded weighting at zero. However, for the sake of a transparent presentation, we use Assumption 2.4 here, as this also simplifies the technical arguments in the proofs later on.

- (iv) Assumption 2.5 is crucial for the proof of our main theorem and directly implies

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

Note that in the location model $\theta(F) = \mathbb{E}_F[X]$ we have $R_{i,j} = 0$ and (2.11) obviously holds. In general however, Assumption 2.5 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 (2019), where the estimate (2.11) has been verified for convergence in probability for different functionals including quantiles and variance.

The following result is the main theorem of this section.

Theorem 2.7 Assume that the null hypothesis (1.1) and Assumptions 2.3–2.5 hold. If further $\hat{\Sigma}_m$ is a consistent and non-singular estimator of the LRV matrix Σ , it holds that

$$\sup_{k=1}^{\infty} w(k/m) \hat{E}_m(k) \xRightarrow{D} \sup_{0 \leq t < \infty} \max_{0 \leq s \leq t} (t+1)w(t) \left| W\left(\frac{s}{s+1}\right) - W\left(\frac{t}{t+1}\right) \right|, \quad (2.12)$$

as $m \rightarrow \infty$, where W is a p -dimensional Brownian motion with independent components and $|\cdot|$ denotes the Euclidean norm.

For the sake of completeness, the reader should note that due to Assumption 2.4 the asymptotic behavior of the weight function guarantees that the random variable on the right-hand side of (2.12) is finite (with probability one). In light of Theorem 2.7 one can choose a constant $c(\alpha)$, such that

$$\mathbb{P}\left(\sup_{0 \leq t < \infty} \max_{0 \leq s \leq t} (t+1)w(t) \left| W\left(\frac{s}{s+1}\right) - W\left(\frac{t}{t+1}\right) \right| > c(\alpha)\right) \leq \alpha. \quad (2.13)$$

Note that for Theorem 2.7 we only require that $\hat{\Sigma}_m$ is a consistent estimator for the LRV as $m \rightarrow \infty$. Under both, H_0 and H_1 , such an estimator should be computed from the initial stable set, which prevents the estimate from being corrupted by possible changes/breaks (see also the discussion in Section 4). In practice, the actual choice of LRV-estimator depends on the concrete application and is crucial for the performance of the procedure. A more extensive discussion on LRV-estimation (not only for change point problems) can be found in Andrews (1991) or Shao and Zhang (2010).

Corollary 2.8 then states that our approach leads to a level α detection scheme.

Corollary 2.8 Grant the assumptions of Theorem 2.7 and further let $c(\alpha)$ satisfy inequality (2.13), then

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

The limit distribution obtained in Theorem 2.7 strongly depends on the considered weight function. A special family of functions that has received considerable attention (see Horváth *et al.* (2004), Fremdt, 2015; Kirch and Weber, 2018 among many others) is given by

$$w_{\gamma}(t) = (1+t)^{-1} \max\left\{\left(\frac{t}{1+t}\right)^{\gamma}, \varepsilon\right\}^{-1} \quad \text{with} \quad 0 \leq \gamma < 1/2, \quad (2.14)$$

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 (2.12) can be simplified to an expression that is more easily tractable via simulations. Straightforward calculations yield that Assumption 2.4 is satisfied by the function w_{γ} and the limit distribution in Theorem 2.7 simplifies as follows.

Corollary 2.9 For a p -dimensional Brownian motion W with independent components it holds that

$$\begin{aligned} & \sup_{0 \leq t < \infty} \max_{0 \leq s \leq t} (t+1)w_{\gamma}(t) \left| W\left(\frac{s}{s+1}\right) - W\left(\frac{t}{t+1}\right) \right| \\ & \stackrel{D}{=} \sup_{0 \leq t < 1} \max_{0 \leq s \leq t} \frac{1}{\max\{t', \varepsilon\}} \left| W(t) - W(s) \right| := L_{1,\gamma}. \end{aligned}$$

Remark 2.10 The cumulative distribution function of the random variable on the right-hand side in Corollary 2.9 is hard to derive in general. However in the case of $\gamma = 0$ and dimension $p = 1$, an explicit formula can be obtained. Therefor note that (if we ignore the cutoff constant ε) the following identity holds with probability one

$$\sup_{0 \leq t < 1} \max_{0 \leq s \leq t} |W(t) - W(s)| = \max_{0 \leq t \leq 1} W(t) - \min_{0 \leq t \leq 1} W(t),$$

where the distribution on the right-hand side is known as the *Range of a Brownian motion* (see for instance Feller, 1951). Its distribution function can be found in Borodin and Salminen (1996, p. 146) and is given by

$$F_{L_1, \gamma=0}(x) = 1 + 8 \sum_{k=1}^{\infty} (-1)^k \cdot k \cdot (1 - \Phi(kx)), \quad (2.15)$$

where Φ denotes the c.d.f. of a standard Gaussian random variable. A corresponding result holds for the limit distribution in a closed-end scenario, see Section C of the Supporting information (Gösmann *et al.*, 2020), where an additional parameter in the distribution function is associated with the monitoring length.

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

Assumption 2.11 Under the alternative H_1 defined in (1.2) let

$$\theta_m^{(1)} := \theta(F_1) = \theta(F_2) = \dots = \theta(F_{m+k_m^*-1}) \neq \theta_m^{(2)} := \theta(F_{m+k_m^*}) = \theta(F_{m+k_m^*+1}) = \dots,$$

where the position of the change within the monitoring data $k_m^* \in \mathbb{N}$ may depend on m . For the size of change suppose that

$$\sqrt{m} |\theta_m^{(1)} - \theta_m^{(2)}| \xrightarrow{m \rightarrow \infty} \infty.$$

Further assume that the process $\{\mathcal{IF}_t\}_{t \in \mathbb{Z}}$ and the remainders defined in Assumption 2.5 are of the following order before the change

$$\frac{1}{\sqrt{m+k_m^*}} \left| \sum_{t=1}^{m+k_m^*-1} \mathcal{IF}_t \right| = \mathcal{O}_{\mathbb{P}}(1) \quad \text{and} \quad \sqrt{m+k_m^*} |R_{1, m+k_m^*-1}| = \mathcal{O}_{\mathbb{P}}(1). \quad (2.16)$$

For the period following the change, assume that there exists a constant $c_a > 0$ and distinct two cases:

(1) If $k_m^*/m = \mathcal{O}(1)$, suppose that

$$\frac{1}{\sqrt{m}} \left| \sum_{t=m+k_m^*}^{m+k_m^*+\lfloor c_a m \rfloor} \mathcal{IF}_t \right| = \mathcal{O}_{\mathbb{P}}(1) \quad \text{and} \quad \sqrt{m} |R_{m+k_m^*, m+k_m^*+\lfloor c_a m \rfloor}| = \mathcal{O}_{\mathbb{P}}(1) \quad (2.17)$$

and for the cutoff constants in (2.10), assume that $t_w < k_m^*/m + c_a \leq T_w$.

(2) If $k_m^*/m \rightarrow \infty$, suppose that

$$\frac{1}{\sqrt{k_m^*}} \left| \sum_{t=m+k_m^*}^{m+k_m^*+\lfloor c_a k_m^* \rfloor} \mathcal{IF}_t \right| = \mathcal{O}_{\mathbb{P}}(1) \quad \text{and} \quad \sqrt{k_m^*} |R_{m+k_m^*, m+k_m^*+\lfloor c_a k_m^* \rfloor}| = \mathcal{O}_{\mathbb{P}}(1). \quad (2.18)$$

Assume additionally that the weight function satisfies $T_w = \infty$ and

$$\liminf_{t \rightarrow \infty} t\tilde{w}(t) > 0. \quad (2.19)$$

Remark 2.12 The assumptions stated above are substantially weaker than those used to investigate the asymptotic properties of $\sup_{k=1}^{\infty} w(\frac{k}{m})\hat{E}_m(k)$ 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 2.3 and the uniform negligibility of the remainders in Assumption 2.5. It is easy to see, that the conditions on the sequence \mathcal{IF}_t are already satisfied if both, its phases before and after the change fulfill a central limit theorem. Finally, it is worth mentioning that the assumptions for the change position k_m^* and size $|\theta_m^{(1)} - \theta_m^{(2)}|$ are very flexible as we allow both quantities to depend on m , where the latter can also tend to zero (sufficiently slow as $m \rightarrow \infty$).

For *early* changes, that is $k_m^*/m = \mathcal{O}(1)$, it is obvious that the change has to occur before monitoring is stopped, where the inequality $k_m^*/m \leq T_w - c_a$ ensures that there is enough data, such that it can actually be detected. On the other hand, the motivation for the inequality $t_w < k_m^*/m + c_a$ is slightly more technical. Roughly spoken, it guarantees, that the time frame $m + k_m^*, \dots, m + k_m^* + c_a m$, which follows the change, is not completely covered by the weight function's cutoff at monitoring start. For exactly this time frame we know by assertion (2.17), that the time series still behaves reasonable.

For *late* changes, that is $k_m^*/m \rightarrow \infty$, it is by Assumption 2.11 not allowed to use a cutoff ($T_w < \infty$) in the weight function. Here we rely on the extra assumption in (2.19), which defines a lower bound for the growth rate of the weight function. Heuristically, this is necessary as it guarantees, that a sufficient amount of weight is assigned even to late time points. The reader should note that this assumption is obviously fulfilled by the standard weighting defined in (2.14).

The next theorem yields consistency under the alternative hypothesis.

Theorem 2.13 Assume that the alternative hypothesis (1.2) and Assumptions 2.4 and 2.11 hold. If further $\hat{\Sigma}_m$ is non-singular and weakly convergent to a non-singular, deterministic matrix, it holds that

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

Consequently, $\lim_{m \rightarrow \infty} \mathbb{P}\left(\sup_{k=1}^{\infty} w(k/m)\hat{E}_m(k) > c\right) = 1$ holds for any constant $c \in \mathbb{R}$.

3. SOME SPECIFIC CHANGE POINT PROBLEMS

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 (2019).

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, 2014; or Hoga, 2017 among many others).

Is it 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 2.5 and the corresponding parts of Assumption 2.11 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 2.3, which also implies the remaining part of Assumption 2.11 (see also the discussion in Remark 2.6). In this situation both, Theorems 2.7 and 2.13 are valid provided that the chosen weighting fulfills Assumption 2.4.

3.2. Changes in Linear Models

Consider the time-dependent linear model

$$Y_t = P_t^\top \beta_t + \varepsilon_t, \quad (3.1)$$

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.*, 2009b; Fremdt, 2015 among many others).

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

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

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

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

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

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

which leads to the estimators

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

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

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

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

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

which directly gives $\mathbb{E}[\mathcal{IF}_t] = 0$. Assuming additionally stationarity of $\{\varepsilon_t\}_{t \in \mathbb{N}}$, it follows that the random sequence $\{X_t\}_{t \in \mathbb{N}}$ is stationary under the null hypothesis. In this case, the linearization defined in (2.7) simplifies to

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

Consequently, the remainders in (2.7) vanish and Assumption 2.5 is obviously satisfied. Next, note that the LRV matrix is given by

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

with $\Gamma = \sum_{t \in \mathbb{Z}} \text{Cov}(Y_0 P_0, Y_t P_t)$, which can be estimated by $\hat{\Sigma}_m = M^{-1} \hat{\Gamma} M^{-1}$ 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

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

and for this reason it 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 Theorems 2.7 and 2.13.

Corollary 3.1 Assume that the predictor sequence $\{P_t\}_{t \in \mathbb{N}}$ and the centered sequence $\{\varepsilon_t\}_{t \in \mathbb{N}}$ are strictly stationary and the second moment matrix $M = \mathbb{E}[P_1 P_1^\top]$ is non-singular. Further suppose that the sequences $\{P_t\}_{t \in \mathbb{N}}$ and $\{\varepsilon_t\}_{t \in \mathbb{N}}$ are independent and let the weight function under consideration fulfill Assumption 2.4.

- (i) Under the null hypothesis H_0 of no change, it follows that the sequence $\{\mathcal{IF}_t\}_{t \in \mathbb{N}}$ defined in (3.4) is strictly stationary. Assume further that this sequence admits the approximation in Assumption 2.3 and that $\hat{\Gamma}_m$ is a non-singular, consistent estimator of the non-singular LRV matrix Γ defined in (3.6). Then monitoring based on the statistic \hat{E} in (3.7) is an asymptotic level α procedure.

- (ii) Under the alternative hypothesis H_1 suppose that the remaining parts of Assumption 2.11 are fulfilled. If further $\hat{\Gamma}$ is non-singular and weakly convergent to a non-singular, deterministic matrix, the monitoring based on the statistic \hat{E} in (3.7) is consistent.

Remark 3.2 If one replaces (the unknown) moment matrix M on the right-hand side of (3.3) by an appropriate estimate, that is

$$\hat{\beta}_i^j = \frac{(\hat{M}_i^j)^{-1}}{j-i+1} \sum_{t=i}^j P_t Y_t \quad \text{with} \quad \hat{M}_i^j = \frac{1}{j-i+1} \sum_{t=i}^j P_t P_t^\top,$$

one obtains a modified statistic given by

$$\begin{aligned} \hat{E}_m(k) &= m^{-1/2} \max_{j=0}^{k-1} (k-j) \left\| \hat{\beta}_{m+j+1}^{m+k} - \hat{\beta}_1^{m+j} \right\|_{\hat{\Sigma}_m^{-1}} \\ &= m^{-1/2} \max_{j=0}^{k-1} \left\| (\hat{M}_{m+j+1}^{m+k})^{-1} \sum_{t=m+j+1}^{m+k} P_t (Y_t - P_t^\top \hat{\beta}_1^{m+j}) \right\|_{\hat{\Sigma}_m^{-1}}, \end{aligned} \quad (3.8)$$

where $\hat{\Sigma}_m$ denotes an appropriate LRV estimator. Note that in this situation the dependence of the unknown moment matrix M (or its estimators) cannot cancel out as observed in (3.7). The modified statistic can be reasonable to employ if – for example – possible changes in the distribution of P_t have to be taken into account. However as (3.8) illustrates the modified statistic \hat{E} can equivalently be written as weighted residual-based approach. This kind of phenomena is already known in the literature, as Hušková and Koubková (2005) describe this for a similar statistic in linear models.

4. FINITE SAMPLE PROPERTIES

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

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

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 (2015) and has been already reconsidered by Kirch and Weber (2018). In the simulation study we will restrict ourselves to the commonly used class of weight functions w_γ defined in (2.14), where we set the involved, technical constant $\varepsilon = 10^{-10}$ when computing the statistics. Under the assumptions made in Section 2, it can be shown by similar arguments as given in Section A of the Supporting information (Gösmann *et al.*, 2020) that

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

Table I. $(1-\alpha)$ -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 Horváth *et al.* (2004) and Fremdt (2015) respectively. The quantiles for $L_{1,0}$ for $p = 1$ were computed with respect to formula (2.15).

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
1	0	3.0233	2.4977	2.2412	2.7912	2.2365	1.9497	2.8262	2.2599	1.9914
	0.25	3.1050	2.5975	2.3542	2.9445	2.3860	2.1060	2.9638	2.4296	2.1758
	0.45	3.4269	2.9701	2.7398	3.3015	2.7992	2.5437	3.3817	2.9241	2.7002
	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

and

$$\sup_{k=1}^{\infty} w_{\gamma} \left(\frac{k}{m} \right) \hat{P}_m(k) \xrightarrow{D} \sup_{0 \leq t < 1} \max_{0 \leq s \leq t} \frac{1}{\max\{t', \varepsilon\}} \left| W(t) - \frac{1-t}{1-s} W(s) \right| =: L_{3,\gamma}, \quad (4.3)$$

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 (2015), 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 2.9. By (4.2), (4.3), and Corollary 2.8 the necessary critical values for the procedures \hat{E} , \hat{Q} , and \hat{P} combined with weighting 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 I for dimensions $p = 1$ and $p = 2$ and have been calculated by 10,000 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, we will examine the finite sample properties of the three statistics for the detection of changes in the mean and in the regression 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 $\alpha = 0.05$.

4.1. Changes in the Mean

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 outlined in Section 3.1. Here we test the null hypothesis of no change which is given by

$$H_0 : \mu_1 = \cdots = \mu_m = \mu_{m+1} = \mu_{m+2} = \cdots, \quad (4.4)$$

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

$$H_1 : \exists k^* \in \mathbb{N} : \mu_1 = \cdots = \mu_{m+k^*-1} \neq \mu_{m+k^*} = \mu_{m+k^*+1} = \cdots. \quad (4.5)$$

We will consider four different data generating models, one white noise process and three autoregressive processes with different levels of temporal dependence controlled by the AR-parameter. To be precise we consider the models

- (M1) $X_t = \varepsilon_t$,
- (M2) $X_t = 0.1X_{t-1} + \varepsilon_t$,
- (M3) $X_t = 0.5X_{t-1} + \varepsilon_t$,
- (M4) $X_t = 0.7X_{t-1} + \varepsilon_t$,

where $\{\varepsilon_t\}$ is an i.i.d. sequence of standard Gaussian random variables. For the AR(1)-processes defined in models (M2)–(M4), we create a burn-in sample of 100 observations in the first place. To simulate the alternative

Table II. Type I error for the open-end procedures based on \hat{E} , \hat{Q} , and \hat{P} at 5% nominal size. The size of the known stable data was set to $m = 50$ (upper part), $m = 100$ (lower part)

m	γ	(M1)			(M2)		
		\hat{E}	\hat{Q}	\hat{P}	\hat{E}	\hat{Q}	\hat{P}
50	0	4.8%	5.3%	5.3%	8.4%	8.8%	9.0%
	0.25	5.0%	5.0%	5.3%	8.9%	8.4%	8.3%
	0.45	4.5%	4.4%	3.9%	7.5%	7.4%	6.4%
100	0	4.1%	4.4%	4.6%	6.8%	6.3%	6.6%
	0.25	5.0%	5.4%	5.6%	7.3%	6.7%	6.9%
	0.45	6.0%	6.2%	5.2%	7.0%	6.4%	6.0%

Table III. Type I error for the open-end procedures based on \hat{E} , \hat{Q} , and \hat{P} at 5% nominal size. The size of the known stable data was set to $m = 100$ (upper part), $m = 200$ (middle part) and $m = 400$ (lower part). In brackets we report the result of simulations, in which the long-run variance estimator has been replaced by the true long-run variance

m	γ	(M3)			(M4)		
		\hat{E}	\hat{Q}	\hat{P}	\hat{E}	\hat{Q}	\hat{P}
100	0	12.1% (3.2)	11.4% (4.5)	11.8% (4.4)	16.3% (2.9)	15.0% (4.1)	15.9% (4.0)
	0.25	13.7% (3.2)	11.8% (3.8)	12.5% (3.8)	18.1% (3.1)	16.3% (3.5)	17.3% (3.5)
	0.45	13.2% (2.6)	12.2% (2.8)	11.6% (2.3)	16.6% (2.1)	14.3% (2.3)	13.9% (1.8)
200	0	7.6% (3.0)	8.1% (4.0)	8.4% (4.1)	9.4% (2.6)	10.5% (3.5)	10.6% (3.5)
	0.25	8.4% (3.5)	7.6% (4.0)	8.3% (4.1)	11.4% (3.0)	11.4% (3.8)	11.6% (3.8)
	0.45	8.7% (3.2)	8.1% (3.2)	7.4% (2.8)	11.3% (2.6)	10.6% (2.7)	10.2% (2.4)
400	0	5.0% (2.8)	6.0% (3.5)	6.2% (3.4)	7.3% (2.6)	7.8% (3.6)	8.2% (3.4)
	0.25	6.2% (3.5)	6.2% (4.0)	6.3% (4.1)	8.2% (3.1)	8.1% (3.8)	8.7% (3.6)
	0.45	6.9% (3.1)	6.2% (3.3)	5.7% (2.9)	7.8% (2.9)	7.9% (2.9)	7.0% (2.6)

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 long run variance estimation we employ the well-known quadratic spectral estimator (Andrews, 1991) with its implementation in the R-package ‘sandwich’ (Zeileis, 2004). To take into account the possible appearance of changes, only the initial stable segment X_1, \dots, X_m is used for this estimate. This restriction is standard in the literature (see eg, Horváth *et al.*, 2004; Wied and Galeano, 2013; or Dette and Gösmann, 2019 among many others), and we will briefly discuss ideas to improve this in our outlook in Section 6. The bandwidth involved in the estimator is chosen as $\log_{10}(m)$ for models (M1) and (M2). To take into account the stronger temporal dependence we take a bandwidth of $\log_{10}(m^4)$ for the models (M3) and (M4).

In Tables II and III we display the type I error for the four time series models (M1)–(M4) and different choices of γ in the weight functions. The principal observation for Table II is, that all three statistical procedures offer a reasonable approximation of the desired nominal level of $\alpha = 0.05$ for the models (M1) and (M2). The results for the weak dependent model (M2) are slightly worse than those for the white noise model (M1).

In Table III it can be seen that the nominal approximation is quite imprecise for the stronger dependent models (M3) and (M4) especially for an initial sample size of $m = 100$. This effect seems to be primary caused by a less precise estimation of the LRV and the approximation improves with larger initial sample size m , such that the type I error is considerably closer to 5% for $m = 400$. To support this conjecture, we also report the type I error for

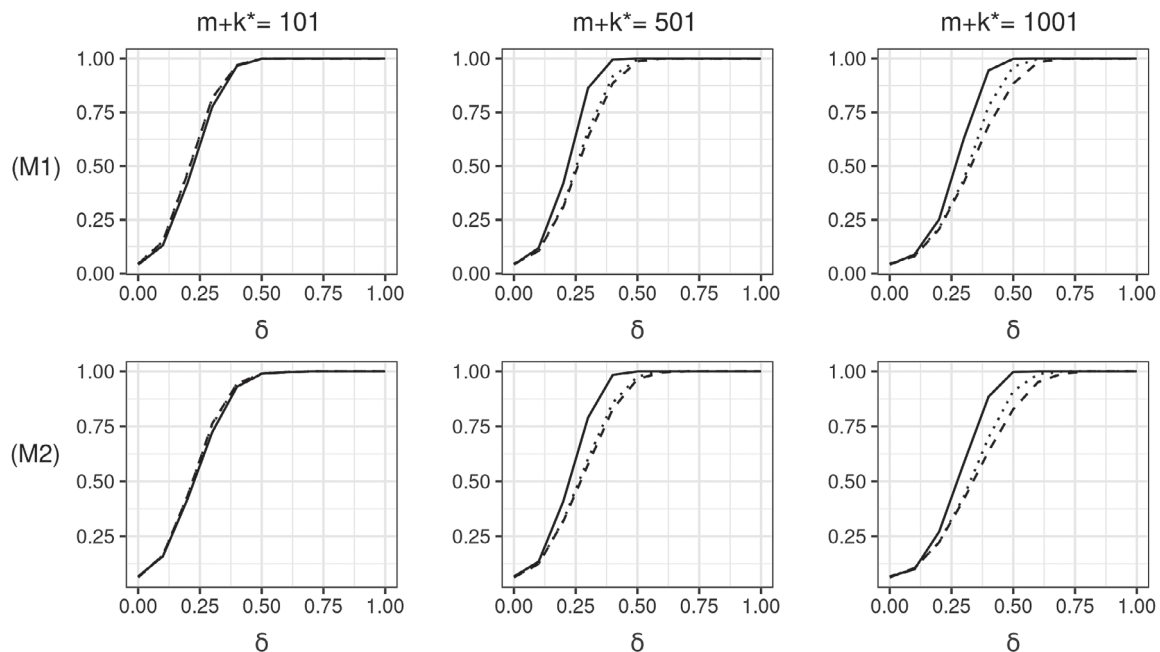


Figure 1. Power of the monitoring procedures for a change in the mean based on the statistics \hat{E} (solid line), \hat{Q} (dashed line) and \hat{P} (dotted line) with $\gamma = 0$ and $m = 100$ at 5% nominal size.

simulations, in which we used the true LRV instead of an estimate, in Table III. This demonstrates a much more sound approximation of the desired test level of 5% for $m = 100$ and $m = 200$.

To discuss the performance under the alternative we illustrate the power of the procedures for increasing values of the change and different change positions for models (M1) and (M2) with $m = 100$ in Figure 1 and for models (M3) and (M4) with $m = 200$ in Figure 2. As the results are very similar we only report the choice $\gamma = 0$ here and provide results for $\gamma = 0.45$ in the Supporting information. The basic tendency observable in Figures 1 and 2 is concordant: 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. The advantage in power is not visible for changes occurring close to the initial training set, where the other procedures perform slightly better.

To give an example, consider the right plot of the first row in Figure 1. Here the test based on the statistic \hat{E} already has a power of 62.8% for a change of $\delta = 0.3$, whereas the tests based on the statistics \hat{P} and \hat{Q} have power of 43.7% and 42.4% 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 other statistics only involve the estimator $\hat{\theta}_1^m$ (see formulae (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 ($m = 50$), 3000 ($m = 100$, $m = 200$), or 4000 ($m = 400$) 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 II error until the specified stopping point. The theory developed in Section 2 also covers the case with a preselected end of the monitoring period. While the statistic for monitoring is the same, the quantile is chosen differently leading to a detector that has higher power if the change is included in the monitoring window and no power if the true change occurs after monitoring ends. We discuss this in the Supporting information.

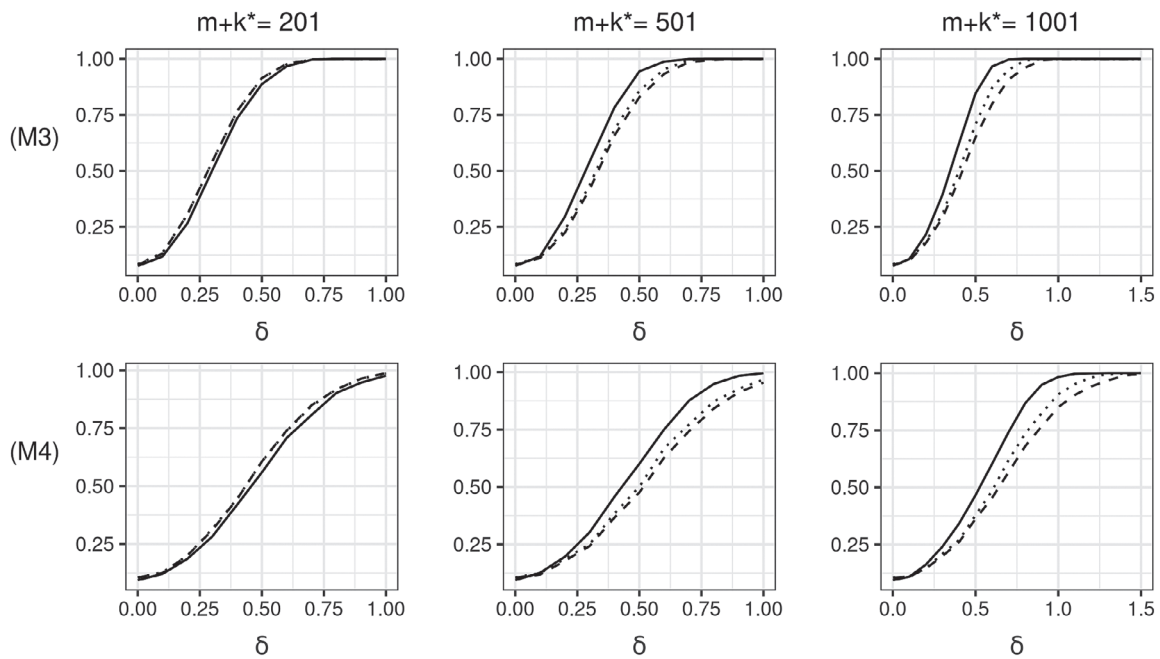


Figure 2. Power of the monitoring procedures for a change in the mean based on the statistics \hat{E} (solid line), \hat{Q} (dashed line) and \hat{P} (dotted line) with $\gamma = 0$ and $m = 200$ at 5% nominal size.

4.2. Changes in Linear Models

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

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

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

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

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

$$(LM1) \ P_t = (1, \sqrt{0.5}Z_t)^\top,$$

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

where Z_t denotes an i.i.d. sequence of $\mathcal{N}(0, 1)$ 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 GARCH(1,1) model (LM2) has been already considered by Fremdt (2015). As pointed out in Section 3.2 the

Table IV. Type I error for the open-end procedures based on \hat{E} , \hat{Q} , and \hat{P} at 5% nominal size. The size of the known stable data was set to $m = 100$

γ	(LM1)			(LM2)		
	\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%

asymptotic variance that needs to be estimated within our procedures is given by

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

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

The problem of detecting changes in the parameter of the linear model has also been addressed using partial sums of the residuals $\hat{\varepsilon}_t = Y_t - P_t^\top \hat{\beta}_t$ in statistics similar to (4.1), where $\hat{\beta}_t$ is an initial estimate of β computed from the initial stable segment. We refer for instance to Chu *et al.* (1996), Horváth *et al.* (2004), who – among many others – use statistics similar to \hat{Q} , or Fremdt (2015), who uses a statistic similar to \hat{P} . Our approach directly compares estimators for the vector β_t , which are derived using the general methodology introduced in Sections 2 and 3. The resulting statistics are obtained replacing $\hat{\theta}$ by $\hat{\beta}$ in (4.1). As pointed out in Remark 3.2, there is a strong connection between methods comparing direct estimates and methods based on weighted residuals, which was already described by Hušková and Koubková (2005). These authors, in particular, demonstrate that these approaches exhibit power against alternatives, that the *plain* residual-based statistics fail to distinguish from the null hypothesis. We also refer to Leisch *et al.* (2000), Hušková and Koubková (2005), and Hušková *et al.* (2007) for a comparison of (plain) residual-based methods with methods using the estimators directly.

In Table IV we display the approximation of the nominal level for the three statistics with different values of the parameter γ in the weight function, where monitoring was stopped after 1500 observations. We observe an acceptable approximation of the nominal level 5% in the case $\gamma = 0$, while the rejection probabilities 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 I error has also been observed by other authors (see eg, Wied and Galeano, 2013) and can be explained by a more sensitive weight 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 3 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} .

5. TWO APPLICATIONS

We apply our methodology alongside two competitors to monitor for changes in linear models. We discuss two examples, related to the United Kingdom European Union membership referendum 2016. We consider the linear model

$$Y_t = \beta_1 P_{1,t} + \beta_2 P_{2,t} + \varepsilon_t, \quad (5.1)$$

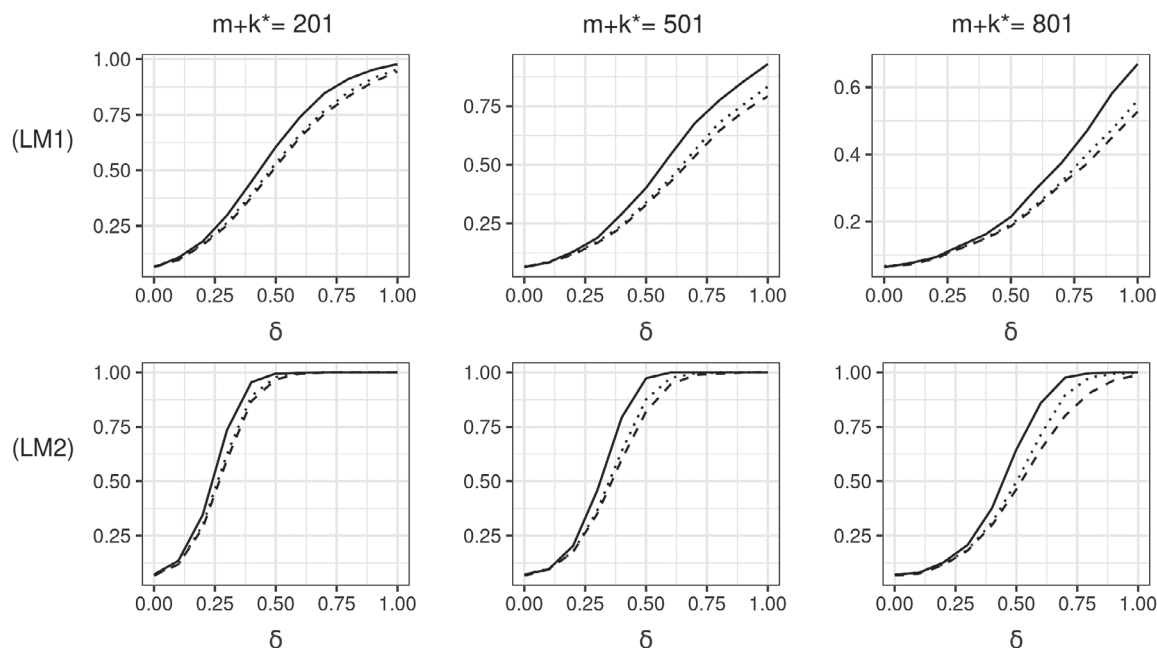


Figure 3. Power of the monitoring procedures for a change in the regression parameters for the open-end procedures based on the statistics \hat{E} (solid line), \hat{Q} (dashed line) and \hat{P} (dotted line) with $\gamma = 0$ and $m = 100$ at 5% nominal size.

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.

Recall that our approach requires a stable segment of m observations in which no changes have yet happened. We choose a stable segment of size $m = 20$ for our analysis of the data and monitor with the three detectors \hat{E} , \hat{P} , and \hat{Q} defined in (2.2) and (4.1) respectively. More precisely, the detectors are updated for every incoming observation, namely $(Y_t, P_{1,t}, P_{2,t})$, and a decision is made, by comparing the detectors with the corresponding thresholds, whether to reject the null hypothesis and stop the procedure or to continue monitoring with the subsequent observation. Monitoring then continues until a change has been detected by each of the three approaches.

For the next monitoring phase another m observations from the time where the last of the three detectors has rejected are used as the next stable segment. Monitoring ends once the end of the available data is reached.

In the remaining part of this section, we present the outcomes of the previously described statistical analysis 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 $\alpha = 0.05$ and the weight function w_0 , as defined in (2.14). All data used was obtained from <https://www.ariva.de> on 26 March 2020.

As our first example, we consider the relation of the 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 (5.1). 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 January 2016 to December 2019 can be seen in Figure 4. The outcomes of the previously described analysis are presented visually in the graphs. The first 20 observation (4 Jan 2016 to 29 Jan 2016, note that we only considered trading days FXCM) were used as the stable segment for monitoring. The monitoring starts on 1 Feb 2016 and went on with all three detectors until 17 Mar 2016 when \hat{P} and \hat{Q} reject, but \hat{E} does not yet reject. Monitoring continues with \hat{E} only until 29 Mar 2019 when the first phase of monitoring ends as all three monitoring procedures have rejected the null hypothesis. The monitoring procedure is then restarted with the 20 observations from the time of rejection

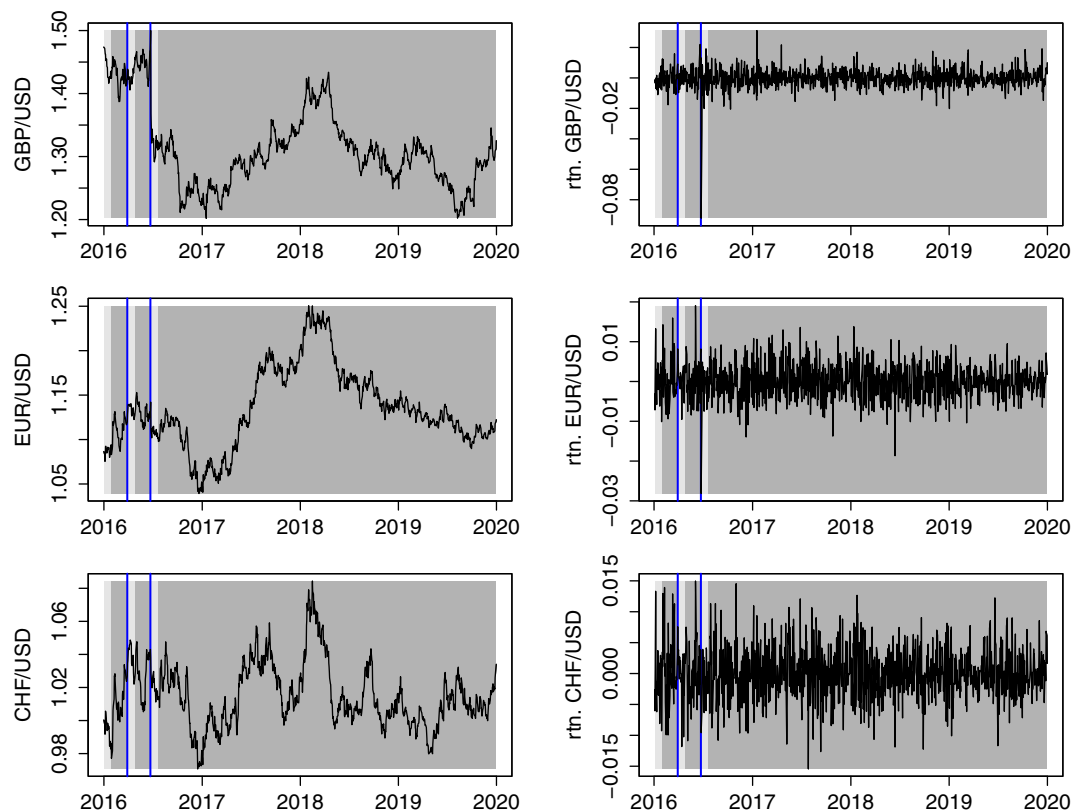


Figure 4. Results of the analysis detailed in Section 5 applied to log returns of three foreign exchange rates. Response: GBP/USD (top), predictors: EUR/USD (middle), and CHF/USD (bottom). Parameters used were $m := 20$, $\alpha = 0.05$ and $\gamma = 0$. Shaded areas indicate observations that were used as the stable segment (light gray) or monitored for changes with \hat{E} (dark gray). Vertical blue lines indicate times the sequential procedure \hat{E} rejected. The times the sequential procedures $\hat{E}/\hat{P}/\hat{Q}$ stopped were: 29 Mar 2016/17 Mar 2016/17 Mar 2016 and 23 Jun 2016/23 Jun 2016/24 Jun 2016. [Color figure can be viewed at wileyonlinelibrary.com]

(29 Mar 2016 to 25 Apr 2016) as the stable segment and monitoring continues from 26 Apr 2016 until 23 Jun 2016 (day of the UK EU referendum), when \hat{E} and \hat{P} reject. After these rejections, monitoring continues for one more day, until 24 Jun 2016, when also \hat{Q} rejects. Finally, the monitoring procedure is restarted with the next 20 observations (24 Jun 2016 to 21 Jul 2016) and monitoring continues until 31 Dec 2019 without rejections by any of the three detectors. In this example, we see that the three detectors behave quite similar, as each of them rejects twice around the time of the UK EU referendum and no further changes afterwards.

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 (5.1). 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 December 2019 can be seen in Figure 5. The outcomes of the previously described analysis are presented visually in the graphs. The first 20 observations (6 Jan 2016 to 1 Feb 2016) were used as the stable segment for the first phase of monitoring. The monitoring starts on 2 Feb 2016 and on with all three detectors until 6 Feb 2017 when \hat{E} rejects, but \hat{P} and \hat{Q} do not yet reject. Monitoring continues with only \hat{P} and \hat{Q} until 16 Mar 2017 when the first phase of monitoring ends with \hat{P} and \hat{Q} also rejecting. In Figure 5, the time that was only monitored by \hat{P} and \hat{Q} is not shaded in gray, because \hat{E} has already rejected.

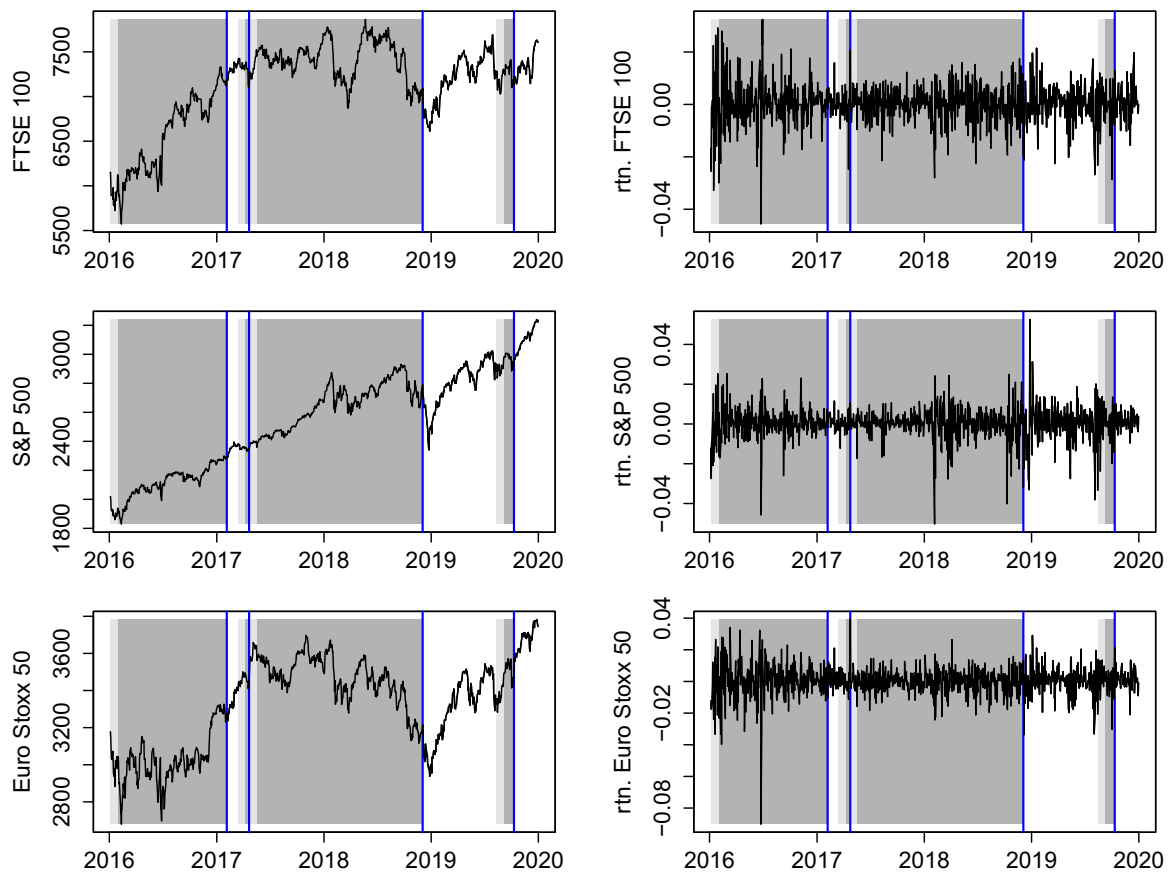


Figure 5. Results of the analysis detailed in Section 5 applied to log returns of three market indices. Response: FTSE 100 (top), predictors: S&P 500 (middle) and EuroStoxx 50 (bottom). Parameters used were $m := 30$, $\alpha = 0.05$ and $\gamma = 0$. Shaded areas indicate observations that were used as the stable segment (light gray) or monitored for changes with \hat{E} (dark gray). Vertical blue lines indicate times the sequential procedure \hat{E} rejected. The times the sequential procedures $\hat{E}/\hat{P}/\hat{Q}$ stopped were: 06 Feb 2016/16 Mar 2016/16 Mar 2016, 24 Apr 2017/18 Apr 2017/24 Apr 2017, 04 Dec 2018/14 Aug 2019/14 Aug 2019, and 11 Oct 2019/11 Oct 2019/did not stop. [Color figure can be viewed at wileyonlinelibrary.com]

For the second phase of monitoring the procedures are then restarted with the 20 observations from the time of rejection (16 Mar 2016 to 7 Apr 2016) as the stable segment and monitoring continues from 10 Apr 2016 until 18 Apr 2017 when \hat{P} rejects. Monitoring continues with only \hat{E} and \hat{Q} until 24 Apr 2017 when the second phase of monitoring ends with \hat{P} and \hat{Q} both also rejecting.

For the third phase of monitoring the procedures are then restarted again with the 20 observations from the time of rejection (24 Apr 2017 to 18 May 2017) as the stable segment and monitoring continues from 19 May 2017 until 4 Dec 2018 when \hat{E} rejects. Monitoring continues with \hat{P} and \hat{Q} only until 14 Aug 2019 when the third phase of monitoring ends with \hat{P} and \hat{Q} both also rejecting.

For the fourth and final phase of monitoring the procedures are then restarted again with the 20 observations from the time of rejection (14 Aug 2019 to 6 Sep 2019) as the stable segment and monitoring continues from 9 Sep 2019 until 11 Oct 2019 when \hat{E} and \hat{P} both reject. Monitoring continues with \hat{Q} only until 31 Dec 2019, the end of the available data, without a rejection of \hat{Q} . In this example, we see that \hat{E} , as expected from the simulations, is capable of detecting changes earlier after a longer period of monitoring. Only in the second period, where the rejection happens early, this is not the case.

6. CONCLUSION AND OUTLOOK

In this article 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, \dots, X_m by an estimator from the sample X_1, \dots, X_{m+j} . We then compare this estimator with the estimator from the sample $X_{m+j+1}, \dots, X_{m+k}$ for every $j = 0, \dots, k-1$. For the new statistic the asymptotic distribution under the null hypothesis and the consistency of a corresponding test, which controls the type I error, are established. By considering a common class of weight functions w_γ defined in (2.14) 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.

For a future research project it is of interest to replace Assumption 2.3 by an FCLT for any fixed time horizon and Hájek-Rényi-inequalities, as done for instance in Kirch and Weber (2018) and Kirch and Stoehr (2019). Since these conditions are slightly weaker, it would be a benefit to establish the results at hand under those conditions.

Another issue – particularly with regard to our simulation study in Section 4 – is that the test level approximation depends sensitively on the efficient estimation of the LRV. The standard approach in our field, which we also followed, is to employ only the initial set for this estimate. As the performance of this is poor for stronger dependent models, it is logical to take a permanently updated estimate into consideration, which fits to the basic message of this work to enhance initial estimates during monitoring. Moreover, one could tackle this problem developing a concept of self-normalization (see Shao and Zhang, 2010), which is applicable in an open-end scenario. However, as the discussion of both ideas is technically involved, it is beyond the scope of this article and left as a promising subject for future research.

Finally, it is a logical next step to also characterize 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 (2014) respectively.

ACKNOWLEDGEMENTS

This work has been supported in part by the Collaborative Research Center ‘Statistical modeling of nonlinear dynamic processes’ (SFB 823, Teilprojekt A1, C1) and the Research Training Group ‘High-dimensional phenomena in probability – fluctuations and discontinuity’ (RTG 2131) of the German Research Foundation (DFG). Moreover, the authors thank Christina Stöhr and Herold Dehling for extremely helpful discussions. We are also grateful to the two unknown referees for their constructive comments on an earlier version of the article and to Ivan Kojadinovic for pointing out some errors. Open access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are provided as online supporting information and were obtained from <https://www.ariva.de> on 26 March 2020.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher’s web site.

REFERENCES

- Anatolyev S, Kosenok G. 2018. Sequential testing with uniformly distributed size. *Journal of Time Series Econometrics* **10**.
- Andrews D. 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**: 817–858.
- Aue A, Dienes C, Fremdt S, Steinebach J. 2015. Reaction times of monitoring schemes for ARMA time series. *Bernoulli* **21**: 1238–1259.
- Aue A, Hörmann S, Horváth L, Hušková M, Steinebach JG. 2012. Sequential testing for the stability of high-frequency portfolio betas. *Econometric Theory* **28**: 804–837.

- Aue A, Hörmann S, Horváth L, Reimherr M. 2009a. Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics* **37**: 4046–4087.
- Aue A, Horváth L. 2004. Delay time in sequential detection of change. *Statistics & Probability Letters* **67**: 221–231.
- Aue A, Horváth L. 2013. Structural breaks in time series. *Journal of Time Series Analysis* **34**: 1–16.
- Aue A, Horváth L, Hušková M, Kokoszka P. 2006. Change-point monitoring in linear models. *The Econometrics Journal* **9**: 373–403.
- Aue A, Horváth L, Reimherr ML. 2009b. Delay times of sequential procedures for multiple time series regression models. *Journal of Econometrics* **149**: 174–190.
- Aue A, Hörmann S, Horváth L, Hušková M. 2014. Dependent functional linear models with applications to monitoring structural change. *Statistica Sinica* **24**: 1043–1073.
- Berkes I, Gombay E, Horváth L, Kokoszka P. 2004. Sequential change-point detection in GARCH(p, q) models. *Econometric Theory* **20**: 1140–1167.
- Borodin A, Salminen P. 1996. Handbook of Brownian motion – facts and formulae. In *Probability and Its Applications* Basel: Birkhäuser.
- Chen Z, Tian Z. 2010. Modified procedures for change point monitoring in linear models. *Mathematics and Computers in Simulation* **81**: 62–75.
- Chu C-SJ, Stinchcombe M, White H. 1996. Monitoring structural change. *Econometrica* **64**: 1045–1065.
- Dette H, Gösmann J. 2019. A likelihood ratio approach to sequential change point detection for a general class of parameters. *Journal of the American Statistical Association*, 10.1080/01621459.2019.1630562, (to appear in print).
- Eberlein E. 1986. On strong invariance principles under dependence assumptions. *The Annals of Probability* **14**: 260–270.
- Feller W. 1951. The asymptotic distribution of the range of sums of independent random variables. *The Annals of Mathematical Statistics* **22**: 427–432.
- Fremdt S. 2014. Asymptotic distribution of the delay time in Page's sequential procedure. *Journal of Statistical Planning and Inference* **145**: 74–91.
- Fremdt S. 2015. Page's sequential procedure for change-point detection in time series regression. *Statistics* **49**: 128–155.
- Gösmann J., Kley T., Dette H. 2020. *Online appendix to: a new approach for open-end sequential change point monitoring*.
- Hampel F, Ronchetti E, Rousseeuw P, Stahel W. 1986. Robust statistics: the approach based on influence functions. In *Wiley Series in Probability and Mathematical Statistics* New York: John Wiley & Sons, Inc.
- Hoga Y. 2017. Monitoring multivariate time series. *Journal of Multivariate Analysis* **155**: 105–121.
- Horváth L, Hušková M, Kokoszka P, Steinebach J. 2004. Monitoring changes in linear models. *Journal of Statistical Planning and Inference* **126**: 225–251.
- Horváth L, Kühn M, Steinebach J. 2008. On the performance of the fluctuation test for structural change. *Sequential Analysis* **27**: 126–140.
- Hušková M, Koubková A. 2005. Monitoring jump changes in linear models. *Journal of Statistical Research* **39**: 51–70.
- Hušková M, Prášková Z, Steinebach J. 2007. On the detection of changes in autoregressive time series I. *Asymptotics. Journal of Statistical Planning and Inference* **137**: 1243–1259.
- Kirch C., Stoeck C. 2019. *Sequential change point tests based on U-statistics*. arXiv e-print arXiv:1912.08580.
- Kirch C, Weber S. 2018. Modified sequential change point procedures based on estimating functions. *Electronic Journal of Statistics* **12**: 1579–1613.
- Lai TL. 1995. Sequential changepoint detection in quality control and dynamical systems. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**: 613–658.
- Lai TL. 2001. Sequential analysis: some classical problems and new challenges. *Statistica Sinica* **11**: 303–408.
- Leisch F, Hornik K, Kuan C-M. 2000. Monitoring structural changes with the generalized fluctuation test. *Econometric Theory* **16**: 835–854.
- Otto S., Breitung J. 2020. *Backward CUSUM for testing and monitoring structural change*. arXiv e-print arXiv:2003.02682.
- Page ES. 1954. Continuous inspection schemes. *Biometrika* **41**: 100–115.
- Page ES. 1955. Control charts with warning lines. *Biometrika* **42**: 243–257.
- Pape K., Wied D., Galeano P. 2016. Monitoring multivariate variance changes. *Journal of Empirical Finance* **39**: 54–68.
- Schmitz A., Steinebach J. G. 2010. A note on the monitoring of changes in linear models with dependent errors. In *Dependence in Probability and Statistics. Lecture Notes in Statistics*, Vol. 200, Doukhan P., Lang G., Surgailis D., Teyssière G. (eds.). Berlin, Heidelberg: Springer; 159–174.
- Shao X, Zhang X. 2010. Testing for change points in time series. *Journal of the American Statistical Association* **105**: 1228–1240.
- Wied D, Galeano P. 2013. Monitoring correlation change in a sequence of random variables. *Journal of Statistical Planning and Inference* **143**: 186–196.
- Zeileis A. 2004. Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software* **11**: 1–17.