A likelihood ratio approach to sequential change point detection

Holger Dette, Josua Gösmann
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany

February 21, 2018

Abstract

In this paper we propose a new approach for sequential monitoring of a parameter of a d-dimensional time series. We consider a closed-end-method, which is motivated by the likelihood ratio test principle and compare the new method with two alternative procedures. We also incorporate self-normalization such that estimation of the long-run variance is not necessary. We prove that for a large class of testing problems the new detection scheme has asymptotic level α and is consistent. The asymptotic theory is illustrated for the important cases of monitoring a change in the mean, variance and correlation. By means of a simulation study it is demonstrated that the new test performs better than the currently available procedures for these problems.

AMS subject classification: 62M10, 62G10, 62G20, 62L99

Keywords and phrases: change point analysis, self-normalization, sequential monitoring, likelihood ratio principle

1 Introduction

An important problem in statistical modeling of time series is the problem of testing for structural stability as changes in the data generating process may have a substantial impact on statistical inference developed under the assumption of stationarity. Because of its importance there exists a large amount of literature, which develops tests for structural breaks in various models and we refer to Aue and Horváth (2013) and Jandhyala et al. (2013) for more recent reviews of the literature. There are essentially to ways how the problem of change point analysis is addressed. A large portion of the literature discusses a-posteriori change point analysis, where the focus is on the detection of structural breaks given a historical data set [see Davis et al. (1995), Csörgö and Horváth (1997), Aue et al. (2009b), Jirak (2015) among many others].

On the other hand, in many applications, such as engineering, medicine or risk management data arrives steadily and therefore several authors have addressed the problem of sequentially monitoring changes in a parameter. Page (1954), Moustakides (1986) and Lai (1995) among others developed detection schemes for models with an infinite time horizon. These always stop and different methods are compared in terms of the average run length. An alternative monitoring ansatz, which on the one hand allows to control (asymptotically) the type I error (if no changes occur) and on the other hand provides the possibility of power analysis was introduced by Chu et al. (1996) in the context of testing the structural stability of the parameters in a linear model. Horváth et al. (2004), Fremdt (2014) extended this approach for linear models with infinite time horizon, while Aue et al. (2012), Wied and Galeano (2013), Pape et al. (2016) developed monitoring procedures for changes in a capital asset pricing model, correlation and variance under the assumption of a finite time horizon.

In this paper we propose a general alternative sequential test in this context, which is applicable for change point analysis of a p-dimensional parameter of a d-dimensional time series if a historical data set from a stable phase is available and then data arrives consecutively. Our approach differs from the current methodology as it is motivated by a likelihood ratio test for a structural change. To be precise, Wied and Galeano (2013) proposed to compare estimates from the historical data set, say $X_1, \ldots, X_m$, with estimators from the sequentially observed data $X_{m+1}, \ldots, X_{m+k}$ [see also Chu et al. (1996); Horváth et al. (2004); Aue et al. (2012); Pape et al. (2016)], while Fremdt (2014) and Kirch and Weber (2017) suggested to compare the estimate from the historical data set with estimates from the sequentially observed data $X_{m+j+1}, \ldots, X_{m+k}$ (for all $j = 0, \ldots, k-1$). In contrast, motivated by the likelihood ratio principle, our approach sequentially compares the estimates from the samples $X_1, \ldots, X_{m+j}$ and $X_{m+j+1}, \ldots, X_{m+k}$ (for all $j = 0, \ldots, k-1$).

Moreover, we also propose a self-normalized test, which avoids the problem of estimating the long-run variance. While the concept of self-normalization has been studied intensively for a-posteriori change point analysis [see Shao and Zhang (2010) and Shao (2015) among many others], to our best knowledge self-normalization has not been studied in the context of sequential monitoring.

The statistical model and the change point problem for a general parameter of the marginal distribution are introduced in Section 2, where we also provide the motivation for the statistic used in the sequential scheme (see Example 2.1) and a discussion of the alternative methods. The asymptotic properties of the new monitoring scheme are investigated in Section 3. In particular we prove a result, which allows to control (asymptotically) the probability of indicating a change in the parameter although there is in fact structural stability (type I error). Moreover, we also show that the new test is consistent and investigate the concept of self-normalization in this context. These asymptotic considerations require several assumptions, which are stated in the general context and verified in Section 4 fro the case of monitoring changes in the mean and variance matrix. In Section 5 the finite sample properties of the new procedure are investigated by means of a simulation study, and we also demonstrate the (empirical) superiority of our approach. Finally, all proofs are deferred to a technical appendix (see Section A).

2 Sequential change point testing

Consider a d-dimensional time series $\{X_j\}_{j\in\mathbb{Z}}$, where X_j has distribution function F_j, and denote by $\theta_j = \theta(F_j)$ a p-dimensional parameter of interest of the distribution of F_j. We are taking the sequential point of view and assume that there exists a historical period of length, say $m \in \mathbb{N}$, such that the process is stable in the sense

$$\theta_1 = \theta_2 = \cdots = \theta_m \ . \tag{2.1}$$

We are interested to monitor if the parameter θ_t changes in the future $m + k \geq m + 1$. The sequence $X_1, \ldots, X_m$ is usually referred to a historical or initial training data set, see for example Chu et al. (1996), Horváth et al. (2004), Wied and Galeano (2013) or Kirch and Weber (2017), among many others. Based on this stretch of "stable" observations a sequential procedure should be conducted to test the hypotheses

$$H_0 : \ \theta_1 = \cdots = \theta_m = \theta_{m+1} = \theta_{m+2} = \ldots , \tag{2.2}$$

against the alternative that the parameter $\theta_{m+k^\star}$ changes for some $k^\star \geq 1$, that is

$$H_1 : \ \exists k^\star \in \mathbb{N} : \ \ \theta_1 = \cdots = \theta_{m+k^\star-1} \neq \theta_{m+k^\star} = \theta_{m+k^\star+1} = \ldots \ , \tag{2.3}$$

In order to motivate our approach in particular the test statistic used in the detection scheme, which will be used in the proposed sequential test, we begin with a very simple example of a change in the mean.

Example 2.1 Consider a sequence $\{X_t\}_{t\in\mathbb{Z}}$ of independent, d-dimensional normal distributed random variables with (positive definite) variance matrix Σ and mean vectors

$$\mu_t = \mathbb{E}[X_t] = \theta(F_t) = \int_{\mathbb{R}^d} x dF_t(x) \in \mathbb{R}^d \ , \quad j = 1, 2, \ldots . \tag{2.4}$$

During the monitoring procedure, we propose to successively test the hypotheses

$$\begin{aligned} & H_0 : \ \mu_1 = \cdots = \mu_m = \mu_{m+1} = \cdots = \mu_{m+k} \ , \\ \text{versus} \quad & H_A^{(k)} : \exists\, j \in \{0, \ldots, k-1\} : \ \ \mu_1 = \cdots = \mu_{m+j} \neq \mu_{m+j+1} = \cdots = \mu_{m+k} \end{aligned} \tag{2.5}$$

based on the sample $X_1, \ldots, X_{m+k}$. Under the assumptions made in this example we can easily derive the likelihood ratio

$$\Lambda_m(k) = \frac{\sup\limits_{\mu\in\mathbb{R}^d} \prod\limits_{t=1}^{m+k} f(X_t, \mu)}{\sup\limits_{\substack{j\in\{0,\ldots,k-1\} \\ \mu^{(1)},\mu^{(2)}\in\mathbb{R}^d}} \prod\limits_{t=1}^{m+j} f(X_t, \mu^{(1)}) \cdot \prod\limits_{t=m+j+1}^{m+k} f(X_t, \mu^{(2)})} \ ,$$

where $f(\cdot,\mu)$ denotes the density of a normal distribution with mean μ and variance matrix Σ (note that the first m observations are assumed to be mean-stable). A careful calculation now proves the identity

$$\begin{aligned} -2\log\big(\Lambda_m(k)\big) &= \max_{j=0}^{k-1} \frac{(m+j)(k-j)}{m+k}\big(\hat\mu_1^{m+j}-\hat\mu_{m+j+1}^{m+k}\big)^\top \Sigma^{-1}\big(\hat\mu_1^{m+j}-\hat\mu_{m+j+1}^{m+k}\big) \\ &= \max_{j=0}^{k-1} \frac{(m+k)(m+j)}{(k-j)}\big(\hat\mu_1^{m+j}-\hat\mu_1^{m+k}\big)^\top \Sigma^{-1}\big(\hat\mu_1^{m+j}-\hat\mu_1^{m+k}\big) \ , \end{aligned} \tag{2.6}$$

where $v^\top$ denotes the transposed of the vector v (usually considered as a column vector) and

$$\hat\mu_i^j = \frac{1}{j-i+1}\sum_{t=i}^{j} X_t$$

is the mean of the observations $X_i,\ldots,X_j$. Consequently the null hypothesis $H_0^{(k)}$ should be rejected in favor of the alternative $H_A^{(k)}$ for large values of the statistic

$$\max_{j=0}^{k-1}(m+j)(k-j)\big(\hat\mu_1^{m+j}-\hat\mu_{m+j+1}^{m+k}\big)^\top \Sigma^{-1}\big(\hat\mu_1^{m+j}-\hat\mu_{m+j+1}^{m+k}\big) \ . \tag{2.7}$$

However, as pointed out in Csörgö and Horváth (1997), the asymptotic properties of a likelihood ratio type statistic of the type (2.7) are difficult to study. For this reason we propose to use a weighted version of (2.7) and consider the statistic

$$\begin{aligned} &\max_{j=0}^{k-1}(m+j)^2(k-j)^2\big(\hat\mu_1^{m+j}-\hat\mu_{m+j+1}^{m+k}\big)^\top \Sigma^{-1}\big(\hat\mu_1^{m+j}-\hat\mu_{m+j+1}^{m+k}\big) \\ =&\max_{j=0}^{k-1}\Big((k-j)\sum_{t=1}^{m+j}X_t-(m+j)\sum_{t=m+j+1}^{m+k}X_t\Big)\Sigma^{-1}\Big((k-j)\sum_{t=1}^{m+j}X_t-(m+j)\sum_{t=m+j+1}^{m+k}X_t\Big) \ , \end{aligned} \tag{2.8}$$

for which (after appropriate normalization) weak convergence of a corresponding sequential empirical process can be established. Note that the right-hand side in (2.8) corresponds to the well known CUSUM statistic, which has become a standard tool for change point detection in a retrospective setting.

Motivated by the previous example we propose to use the statistic

$$\hat D_m(k) = m^{-3}\max_{j=0}^{k-1}(m+j)^2(k-j)^2\big(\hat\theta_1^{m+j}-\hat\theta_{m+j+1}^{m+k}\big)^\top \hat\Sigma_m^{-1}\big(\hat\theta_1^{m+j}-\hat\theta_{m+j+1}^{m+k}\big) \ , \tag{2.9}$$

for monitoring changes in the parameter θ_j, where $\hat\theta_i^j = \theta(\hat F_i^j)$ denotes the estimator obtained from the empirical distribution function

$$\hat F_i^j(z) = \frac{1}{j-i+1}\sum_{t=i}^{j} I\{X_t \le z\}$$

of the observations $X_i, \dots, X_j$ and the matrix $\hat{\Sigma}_m$ corresponds to an estimator of a long-run variance based on $X_1, \dots, X_m$. The scaling by m^{-3} will be necessary to obtain weak convergence in the sequel.

We use the sequence $\{\hat{D}_m(k)\}_{k\in\mathbb{N}}$ in combination with an increasing threshold function $w(\cdot)$ as a monitoring scheme. More precisely, let $T \in \mathbb{N}$ denote a constant factor defining the window of monitoring, then we reject the null hypothesis in (2.2) at the first time $k \in \{1, 2, \dots, Tm\}$ for which the detector $\hat{D}_m(k)$ exceeds the threshold function $w : [0, T] \to \mathbb{R}_+$, that is

$$\hat{D}_m(k) > w(k/m) \ . \tag{2.10}$$

This definition yields a stopping rule defined by

$$\tau_m = \inf \Big\{ 1 \leq k \leq Tm \mid \hat{D}_m(k) > w(k/m) \Big\} \ ,$$

(if the set $\{1 \leq k \leq Tm \mid \hat{D}_m(k) > w(k/m)\}$ is empty we define $\tau_m = \infty$). The threshold function has to be chosen such that the test has asymptotic level α, that is

$$\limsup_{m\to\infty} \mathbb{P}_{H_0}\big(\tau_m < \infty\big) = \limsup_{m\to\infty} \mathbb{P}_{H_0}\bigg(\max_{k=1}^{mT} \frac{\hat{D}_m(k)}{w(k/m)} > 1\bigg) \leq \alpha \ , \tag{2.11}$$

and is consistent, i.e.

$$\lim_{m\to\infty} \mathbb{P}_{H_1}\big(\tau_m < \infty\big) = 1 \ . \tag{2.12}$$

Following Aue et al. (2012) we call this procedure a closed-end method, because monitoring is only performed in the interval $[m + 1, mT]$.

To our best knowledge, the detection scheme defined by (2.10) has not been considered in the literature so far. However, our approach is related to the work of Chu et al. (1996); Aue et al. (2012); Wied and Galeano (2013) and Pape et al. (2016), who investigated sequential monitoring schemes for various parameters (such as the correlation, the variance or the parameters of the capital asset pricing model). In the general situation considered in this section their approach uses

$$\hat{Q}_m(k) = \frac{k^2}{m}\big(\hat{\theta}_1^m - \hat{\theta}_{m+1}^{m+k}\big)^\top \hat{\Sigma}_m^{-1}\big(\hat{\theta}_1^m - \hat{\theta}_{m+1}^{m+k}\big) \tag{2.13}$$

as a basic statistic in the sequential procedure. Note that a sequential scheme based on this statistic measures the differences between the estimator $\hat{\theta}_1^m$ from the initial data and the estimator $\hat{\theta}_{m+1}^{m+k}$ from all observations excluding the training sample. As a consequence - in particular in the case of a rather late change - the estimator $\hat{\theta}_{m+1}^{m+k}$ may be corrupted by observations before the change point, which might lead to a loss of power. Another related procedure uses the statistic

$$\hat{P}_m(k) = \max_{j=0}^{k-1} \frac{(k-j)^2}{m}\big(\hat{\theta}_1^m - \hat{\theta}_{m+j+1}^{m+k}\big)^\top \hat{\Sigma}_m^{-1}\big(\hat{\theta}_1^m - \hat{\theta}_{m+j+1}^{m+k}\big) \tag{2.14}$$

and was recently suggested by Fremdt (2014) and reconsidered by Kirch and Weber (2017). These authors compare the estimate from the data $X_1, \ldots, X_m$ with estimates from the data $X_{m+j+1}, \ldots, X_{m+k}$ (for different values of j). This method may lead to a loss in power in problems with a small sample of historical data and a rather late change point. In contrast our approach compares the estimates of the parameters before and after all potential positions of a change point $j \in \{m+1, \ldots, m+k\}$.
In this paper, we argue that the performance of the change point tests can be improved by replacing $\hat{\theta}_1^m$ by $\hat{\theta}_1^{m+j}$ inside the maximum, which would directly lead to a scheme of the form (2.9). Here, we would like to point to our simulation study in Section 5, which contains many cases where a sequential detection scheme based on the statistic $\hat{D}_m$ outperforms schemes based on $\hat{Q}_m$ or $\hat{P}_m$.

3 Asymptotic properties

In the subsequent discussion we use the following notation. We denote by $\ell^\infty(V_1, V_2)$ the space of all bounded functions $f : V_1 \to V_2$ equipped with sup-norm, where V_1, V_2 are normed linear spaces. The symbols $\xrightarrow{\mathbb{P}}$ and $\xRightarrow{\mathcal{D}}$ mean convergence in probability and weak convergence (in the space under consideration), respectively. The process $\{W(s)\}_{s\in[0,T+1]}$ will usually represent a standard p-dimensional Brownian motion. For a vector $v \in \mathbb{R}^d$, we denote by $|v| = \big(\sum_{i=1}^d v_i^2\big)^{1/2}$ its euclidean norm.

3.1 Weak convergence

Throughout this paper we denote by θ a p-dimensional functional of the d-dimensional distribution function F and define its influence function (assuming its existence) by

$$\mathcal{IF}(x, F, \theta) = \lim_{\varepsilon \searrow 0} \frac{\theta((1-\varepsilon)F + \varepsilon\delta_x) - \theta(F)}{\varepsilon} , \tag{3.1}$$

where $\delta_x(z) = I\{x \le z\}$ is the distribution function of the Dirac measure at the point $x \in \mathbb{R}^d$ and the inequality in the indicator is understood component-wise. Throughout this section we make the following assumptions, which will be verified for several important examples in Section 4 [see also Shao and Zhang (2010) for similar regularity conditions].

Assumption 3.1 *Under the null hypothesis (2.2) we assume that the times series $\{X_t\}_{t\in\mathbb{N}}$ is strictly stationary with $\mathbb{E}[\mathcal{IF}(X_1, F, \theta)] = 0$ and that the weak convergence*

$$\frac{1}{\sqrt{m}} \sum_{t=1}^{\lfloor ms \rfloor} \mathcal{IF}(X_t, F, \theta) \xRightarrow{\mathcal{D}} \sqrt{\Sigma_F} W(s) , \tag{3.2}$$

holds in the space $\ell^\infty([0,T+1],\mathbb{R}^p)$ *as* $m\to\infty$*, where the long-run variance matrix is defined by (assuming convergence of the series)*

$$\Sigma_F = \sum_{t\in\mathbb{Z}} \operatorname{Cov}\big(\mathcal{IF}(X_0,F,\theta),\ \mathcal{IF}(X_t,F,\theta)\big) \in \mathbb{R}^{p\times p} \tag{3.3}$$

and $\{W(s)\}_{s\in[0,T+1]}$ *is a* p*-dimensional (standard) Brownian motion.*

Assumption 3.2 *The remainder terms*

$$R_{i,j} = \hat\theta_i^j - \theta(F) - \frac{1}{j-i+1}\sum_{t=i}^{j}\mathcal{IF}(X_t,F,\theta) \tag{3.4}$$

in the linearization of $\hat\theta_i^j - \theta(F)$ *satisfy*

$$\sup_{1\le i<j\le n}(j-i+1)|R_{i,j}| = o_{\mathbb{P}}(n^{1/2})\ . \tag{3.5}$$

For the statement of our first result we introduce the notation (throughout this paper we use the convention $\hat\theta_z^u = 0$, whenever $z>u$)

$$\tilde{\mathbb{U}}(\ell,z,u) := (u-z)(z-\ell)\big(\hat\theta_{\ell+1}^z - \hat\theta_{z+1}^u\big)\ , \tag{3.6}$$

$$\mathbb{U}(z,u) := \tilde{\mathbb{U}}(0,z,u) = (u-z)z\big(\hat\theta_1^z - \hat\theta_{z+1}^u\big)\ , \tag{3.7}$$

and denote by

$$\Delta_2 = \{(s,t)\in[0,T+1]^2 \mid s\le t\}\ , \tag{3.8}$$

$$\Delta_3 = \big\{(r,s,t)\in[0,T+1]^3 \mid r\le s\le t\big\}. \tag{3.9}$$

the 2- and 3-dimensional triangle in $[0,T+1]^2$ and $[0,T+1]^3$, respectively.

Theorem 3.3 *Let Assumptions 3.1 and 3.2 be satisfied. If the null hypothesis in (2.2) holds, then as* $m\to\infty$

$$\big\{m^{-3/2}\tilde{\mathbb{U}}(\lfloor mr\rfloor,\lfloor ms\rfloor,\lfloor mt\rfloor)\big\}_{(r,s,t)\in\Delta_3} \overset{\mathcal{D}}{\Longrightarrow} \Sigma_F^{1/2}\big\{B(s,t)+B(r,s)-B(r,t)\big\}_{(r,s,t)\in\Delta_3} \tag{3.10}$$

in the space $\ell^\infty(\Delta_3,\mathbb{R}^p)$*, where the process* $\{B(s,t)\}_{(s,t)\in\Delta_2}$ *is defined by*

$$B(s,t) = tW(s) - sW(t)\ ,\quad (s,t)\in\Delta_2, \tag{3.11}$$

and $\{W(s)\}_{s\in[0,T+1]}$ *denotes a* p*-dimensional Brownian motion on the interval* $[0,T+1]$.

Remark 3.4 As a by-product of Theorem 3.3 and the representation (3.7) we obtain the weak convergence of the double-indexed CUSUM-process (3.7), that is ($m \to \infty$)

$$\big\{m^{-3/2} \cdot \mathbb{U}(\lfloor ms \rfloor, \lfloor mt \rfloor)\big\}_{(s,t)\in\Delta_2} \overset{\mathcal{D}}{\Longrightarrow} \{\Sigma_F^{1/2} B(s,t)\}_{(s,t)\in\Delta_2} , \tag{3.12}$$

where Δ_2 denotes the 2-dimensional triangle in $[0, T+1]^2$ and the process B is defined in (3.11). In particular the covariance structure of the process B is given by

$$\mathrm{Cov}\,\big(B(s_1,t_1), B(s_2,t_2)\big) = t_1t_2(s_1 \wedge s_2) - t_1s_2(s_1 \wedge t_2) - s_1t_2(t_1 \wedge s_2) + s_1s_2(t_1 \wedge t_2) \ .$$

Consequently, the process $\{B(s,t)\}_{(s,t)\in\Delta_2}$ can be considered as a natural extension of the standard Brownian bridge as for fixed t the process $\{B(s,t)\}_{s\in[0,t]}$ is a Brownian bridge on the interval $[0,t]$.

Observing the definition (3.7) the statistic (2.9) allows the representation

$$\hat{D}_m(k) = m^{-3} \max_{j=0}^{k-1} |\mathbb{U}^\top(m+j, m+k)\hat{\Sigma}_m^{-1}\mathbb{U}(m+j, m+k)| \ , \tag{3.13}$$

and we obtain the following Corollary as a consequence of Theorem 3.3.

Corollary 3.5 *Let the assumptions of Theorem 3.3 be satisfied. If the null hypothesis in* (2.2) *holds, and $\hat{\Sigma}_m$ denotes a consistent estimator of the long-run variance Σ_F defined in* (3.3)*, then as $m \to \infty$*

$$\max_{k=1}^{Tm} \frac{\hat{D}_m(k)}{w(k/m)} \overset{\mathcal{D}}{\Longrightarrow} \sup_{t\in[1,T+1]} \sup_{s\in[1,t]} \frac{B(s,t)^\top B(s,t)}{w(t-1)} \ .$$

for any threshold function $w : [0,T] \to \mathbb{R}^+$, which is increasing.

By the result in Corollary 3.5 it is reasonable to choose for a given level α a threshold function $w_\alpha(\cdot)$, such that

$$\mathbb{P}\bigg(\sup_{t\in[1,T+1]} \sup_{s\in[1,t]} \frac{B(s,t)^\top B(s,t)}{w_\alpha(t-1)} > 1\bigg) = \alpha \tag{3.14}$$

and to reject the null hypothesis H_0 in (2.2) at time k, if

$$\hat{D}_m(k) > w_\alpha(k/m) \ . \tag{3.15}$$

By Corollary 3.5 this test has asymptotic level α, that is

$$\lim_{m\to\infty} \mathbb{P}_{H_0}\Big(\max_{k=1}^{Tm} \frac{\hat{D}_m(k)}{w_\alpha(k/m)} > 1\Big) = \alpha$$

(if the assumptions of Theorem 3.3 hold and w_α satisfies (3.14)). The choice of of $w_\alpha(\cdot)$ has been investigated by several authors [see Chu et al. (1996), Aue et al. (2009b) and Wied and

Galeano (2013) among others] and we will compare different options by means of a simulation study in Section 5. Note that one can take any function (which is increasing and bounded from below by a positive constant) and multiply an appropriate constant such that (3.14) is fulfilled. Next we discuss the consistency of the monitoring scheme (3.15). For this purpose we consider the alternative hypothesis in (2.3), where the location of the change point is increasing with the length of the training sample, that is $m + k^* = \lfloor mc \rfloor$ for some $1 < c < T + 1$. Recalling the definition of $\hat{D}_m(k)$ in (2.9) and observing the inequality

$$\max_{k=1}^{Tm} \hat{D}_m(k) \geq \frac{\lfloor mc \rfloor^2 \big(T(m+1) - \lfloor mc \rfloor\big)^2}{m^3} \cdot \frac{\big(\hat{\theta}_1^{\lfloor mc \rfloor} - \hat{\theta}_{\lfloor mc \rfloor + 1}^{T(m+1)}\big)^\top \hat{\Sigma}_m^{-1} \big(\hat{\theta}_1^{\lfloor mc \rfloor} - \hat{\theta}_{\lfloor mc \rfloor + 1}^{T(m+1)}\big)}{w_\alpha(T)}$$

it is intuitively clear that the statistic $\max_{k=1}^{Tm} \hat{D}_m(k)$ converges to infinity, provided that $\hat{\theta}_1^{\lfloor mc \rfloor}$ and $\hat{\theta}_{\lfloor mc \rfloor + 1}^{T(m+1)}$ are consistent estimates of the parameter θ before and after the change point $m + k^* = \lfloor mc \rfloor$ and $\hat{\Sigma}_m$ converges to a positive definite $p \times p$ matrix. The following Theorem 3.8 makes these heuristic arguments more precise. Its proof requires several assumptions, which are stated first. The result might be even correct under slightly weaker assumptions. However, in the form stated below we can also prove consistency of a sequential scheme based on a self-normalized version of $\hat{D}_m(k)$ (see Theorem 3.10 in Section 3.2).

Assumption 3.6 *If the alternative hypothesis H_1 defined in (2.3) holds we assume that the change occurs at position $m + k^* = \lfloor mc \rfloor$ for some $c \in (1, T+1)$. Moreover, let $\{Z_t(1)\}_{t\in\mathbb{Z}}$ and $\{Z_t(2)\}_{t\in\mathbb{Z}}$ denote (strictly) stationary $\mathbb{R}^d$-valued processes with marginal distribution functions $F^{(1)}$ and $F^{(2)}$, respectively, such that*

$$\theta(F^{(1)}) \neq \theta(F^{(2)}) \ ,$$

and that for each $m \in \mathbb{N}$

$$(X_1, X_2, \ldots, X_{\lfloor mc \rfloor}) \stackrel{\mathcal{D}}{=} (Z_1(1), \ldots, Z_{\lfloor mc \rfloor}(1)), \tag{3.16}$$

$$(X_{\lfloor mc \rfloor + 1}, \ldots, X_{\lfloor mT \rfloor}) \stackrel{\mathcal{D}}{=} (Z_{\lfloor mc \rfloor + 1}(2), \ldots, Z_{\lfloor mT \rfloor}(2)) \ . \tag{3.17}$$

Note, that formally the process $\{X_t\}_{t=1,\ldots \lfloor mT \rfloor}$ is a triangular array, that is $X_t = X_{m,t}$, but we do not reflect this in our notation. Further, we assume that there exist two (standard) Brownian motions W_1 and W_2 such that the joint weak convergence

$$\begin{pmatrix} \big\{\frac{1}{\sqrt{m}} \sum_{t=1}^{\lfloor ms \rfloor} \mathcal{IF}(Z_t(1), F^{(1)}, \theta)\big\}_{s\in[0,c]} \\ \big\{\frac{1}{\sqrt{m}} \sum_{t=\lfloor mc \rfloor+1}^{\lfloor ms \rfloor} \mathcal{IF}(Z_t(2), F^{(2)}, \theta)\big\}_{s\in[c,T+1]} \end{pmatrix} \stackrel{\mathcal{D}}{\Longrightarrow} \begin{pmatrix} \big\{\sqrt{\Sigma_{F^{(1)}}} W_1(s)\big\}_{s\in[0,c]} \\ \big\{\sqrt{\Sigma_{F^{(2)}}} \big(W_2(s) - W_2(c)\big)\big\}_{s\in[c,T+1]} \end{pmatrix}$$

holds, where $\Sigma_{F^{(1)}}$ and $\Sigma_{F^{(2)}}$ denote positive definite matrices defined in the same way as (3.3), that is

$$\Sigma_{F^{(\ell)}} = \sum_{t\in\mathbb{Z}} \mathrm{Cov}\left(\mathcal{IF}(Z_0(\ell), F^{(\ell)}, \theta),\ \mathcal{IF}(Z_t(\ell), F^{(\ell)}, \theta)\right) \in \mathbb{R}^{p\times p} \ , \ \ \ell = 1, 2 \tag{3.18}$$

and both phases in (3.16) and (3.17) fulfill Assumption 3.2 for the corresponding expansion.

Remark 3.7

(a) The interpretation of Assumption 3.6 is as follows: There exist two regimes and the process under consideration switches from one regime to the other.

(b) The assumption of two stationary phases before and after the change point is commonly made in the literature to analyze change point tests under the alternative [see for example Aue et al. (2009b), Dette and Wied (2016), Kirch and Weber (2017) among others]. Note, that we do not assume that the two limiting processes W_1 and W_2 are independent.

(c) Often Assumption 3.6 is directly implied by the underlying change point problem. For example, in the situation of the mean vector introduced in (2.1) it is usually assumed that $X_t = \mu_t + \varepsilon_t$, where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a stationary process and $\mu_t = \mu^{(1)}$ if $t \leq \lfloor mc \rfloor$ and $\mu_t = \mu^{(2)}$ if $t \geq \lfloor mc \rfloor + 1$. In this case Assumption 3.6 is obviously satisfied. Further examples are discussed in Section 4.

Theorem 3.8 *Let Assumption 3.6 be satisfied and let the threshold function w_α satisfy (3.14). Further assume that $\hat{\Sigma}_m$ denotes a consistent estimator of the long-run variance $\Sigma_{F^{(1)}}$ based on the observations $X_1, \ldots, X_m$. Under the alternative hypothesis H_1 we have*

$$\lim_{m\to\infty} \mathbb{P}\left(\max_{k=0}^{Tm} \frac{\hat{D}_m(k)}{w_\alpha(k/m)} > 1 \right) = 1 \ .$$

Remark 3.9 We can establish similar results for the statistics (2.13) and (2.14) proposed by Wied and Galeano (2013) and Fremdt (2014), respectively. For example, if the assumptions of Theorem 3.3 are satisfied we obtain the weak convergence

$$\max_{k=1}^{Tm} \frac{\hat{Q}_m(k)}{w(k/m)} \overset{\mathcal{D}}{\Longrightarrow} \sup_{t\in[1,T+1]} \frac{B(t,1)^\top B(t,1)}{w(t-1)} \ ,$$
$$\max_{k=1}^{Tm} \frac{\hat{P}_m(k)}{w(k/m)} \overset{\mathcal{D}}{\Longrightarrow} \sup_{t\in[1,T+1]} \sup_{s\in[1,t]} \frac{\big(B(s,1)+B(t,1)\big)^\top \big(B(s,1)+B(t,1)\big)}{w(t-1)}$$

under the null hypothesis, which can be used to construct an asymptotic level α monitoring scheme based on the statistics $\hat{Q}_m$ and $\hat{P}_m$, respectively. Consistency of the corresponding tests follows along the arguments given in the proof of Theorem 3.8. The details are omitted for the sake of brevity. The finite sample properties of the three different tests will be investigated by means of a simulation study in Section 5.

3.2 Self-Normalization

The test proposed in Section 3.1 requires an estimator of the long-run variance $\hat{\Sigma}_m$, and we discuss commonly used estimates for this purpose in Section 4. However, it has been pointed

out by several authors that this problem is not an easy one as the common estimates depend sensitively on a regularization parameter (for example a bandwidth), which might be difficult to select in practice. An alternative to long-run variance estimation is the concept of self-normalization, which will be investigated in this section. This approach has been studied intensively for a-posteriori change point analysis [see Shao and Zhang (2010) and Shao (2015) among many others], but - to our best knowledge - self-normalization has not been considered in the context of sequential monitoring. In the following we discuss a self-normalized version of the statistic $\hat{D}_m$ proposed in this paper (see equation (2.9)). Self-normalization of the statistic $\hat{P}_m$ in (2.14) will be briefly discussed in Remark 3.12.
To be precise, we define a self-normalizing matrix

$$\begin{aligned}\mathbb{V}(z,u) = &\sum_{j=1}^{z} j^2(z-j)^2\big(\hat{\theta}_1^j - \hat{\theta}_{j+1}^z\big)\big(\hat{\theta}_1^j - \hat{\theta}_{j+1}^z\big)^\top \\ &+ \sum_{j=z+1}^{u} (u-j)^2(j-z)^2\big(\hat{\theta}_{z+1}^j - \hat{\theta}_{j+1}^u\big)\big(\hat{\theta}_{z+1}^j - \hat{\theta}_{j+1}^u\big)^\top .\end{aligned} \tag{3.19}$$

and replace the estimate $\hat{\Sigma}_m$ of the long-run variance in (2.9) by the matrix $\frac{1}{m^4}\mathbb{V}(m+j,m+k)$. This yields the self-normalized statistic

$$\hat{D}^{\mathrm{SN}}(k) = m \max_{j=0}^{k-1} (m+j)^2(k-j)^2\big(\hat{\theta}_1^{m+j} - \hat{\theta}_{m+j+1}^{m+k}\big)^\top \mathbb{V}^{-1}(m+j,m+k)\big(\hat{\theta}_1^{m+j} - \hat{\theta}_{m+j+1}^{m+k}\big) , \tag{3.20}$$

for which we prove the following result in Section A.

Theorem 3.10 *Let $w : [0,T] \to \mathbb{R}^+$ denote any threshold function, which is increasing and let the assumptions of Theorem 3.3 be satisfied. If the null hypothesis in (2.2) holds, then as $m \to \infty$*

$$\max_{k=1}^{Tm} \frac{\hat{D}_m^{\mathrm{SN}}(k)}{w(k/m)} \overset{\mathcal{D}}{\Longrightarrow} \sup_{t\in[1,T+1]} \sup_{s\in[1,t]} \frac{|\tilde{B}(s,t)|}{w(t-1)} , \tag{3.21}$$

where

$$\tilde{B}(s,t) = B^\top(s,t)\big(N_1(s) + N_2(s,t)\big)^{-1}B(s,t) ,$$

the process $\{B(s,t)\}_{s\in\Delta_2}$ is defined in (3.11) and $\{N_1(s)\}_{s\in[0,T+1]}$ and $\{N_2(s,t)\}_{(s,t)\in\Delta_2}$ are given by

$$\begin{aligned}N_1(s) &= \int_0^s B(r,s)B^\top(r,s)dr , \\ N_2(s,t) &= \int_s^t \big(B(r,t)+B(s,r)-B(s,t)\big)\big(B(r,t)+B(s,r)-B(s,t)\big)^\top dr .\end{aligned} \tag{3.22}$$

The monitoring rule is now defined in the same way as described in Section 3.1 determining a threshold function $w_\alpha(\cdot)$, such that

$$\mathbb{P}\Big(\sup_{t\in[1,T+1]} \sup_{s\in[1,t]} \frac{|\tilde{B}(s,t)|}{w_\alpha(t-1)} > 1 \Big) = \alpha \tag{3.23}$$

for a given level α, and rejecting the null hypothesis H_0 in (2.2) at the time k, if

$$\hat{D}_m^{\mathrm{SN}}(k) > w_\alpha(k/m) \ . \tag{3.24}$$

By Theorem 3.10 this test has asymptotic level α and our next result shows that this procedure is also consistent.

Theorem 3.11 *Let Assumption 3.6 be satisfied and let the threshold function w_α satisfy (3.23). Under the alternative hypothesis H_1 we have*

$$\lim_{m\to\infty} \mathbb{P}\Big(\max_{k=0}^{Tm} \frac{\hat{D}_m^{\mathrm{SN}}(k)}{w_\alpha(k/m)} > 1 \Big) = 1.$$

Remark 3.12 The statistic $\hat{P}_m(k)$ defined (2.14) can be self-normalized in a similar manner, that is

$$\hat{P}_m^{\mathrm{SN}}(k) = m^3 \max_{j=0}^{k} (k-j)^2 \big(\hat{\theta}_1^m - \hat{\theta}_{m+j+1}^{m+k}\big)^\top \mathbb{V}^{-1}(m+j, m+k) \big(\hat{\theta}_1^m - \hat{\theta}_{m+j+1}^{m+k}\big) \ . \tag{3.25}$$

If the the null hypothesis holds and the assumptions of Theorem 3.3 are satisfied it can be shown using (3.12) and similar arguments as given in the proof of Theorem 3.10 that

$$\max_{k=1}^{Tm} \frac{\hat{P}_m^{\mathrm{SN}}(k)}{w(k/m)} \overset{\mathcal{D}}{\Longrightarrow} \sup_{t\in[1,T+1]} \sup_{s\in[1,t]} \frac{|\tilde{B}^{\mathrm{SN}}(s,t)|}{w(t-1)} \ , \tag{3.26}$$

where the process $\tilde{B}^{\mathrm{SN}}$ is defined by

$$\tilde{B}^{\mathrm{SN}}(s,t) = \big(B(s,1) + B(t,1)\big)^\top \big(N_1(s) + N_2(s,t)\big)^{-1} \big(B(s,1) + B(t,1)\big) \ .$$

Similarly, consistency follows along the lines given in the proof of Theorem 3.11. The details are omitted for the sake of brevity.
On the other hand a statistic like $\hat{Q}_m(k)$ defined in (2.13) cannot be self-normalized in a straightforward manner as it does not employ a maximum, which is necessary to separate points before and after the (possible) change point. In particular one cannot use the matrix $\mathbb{V}$ in (3.19), which is based on such a separation. Obviously, a self-normalization approach without separation could be constructed but this would lead to a severe loss in power and is therefore not discussed here. We refer the reader to Shao and Zhang (2010) for a comprehensive discussion of this problem. The finite sample properties of both self-normalized methods $\hat{D}_m^{\mathrm{SN}}(k)$ and $\hat{P}_m^{\mathrm{SN}}(k)$ will be compared by means of a simulation study in Section 5 .

3.3 Implementation

We will close this section with a description of the algorithm to detect changes in the functional $\theta(F)$ employing self-normalization. In the following paragraph $\hat{S}_m(k)$ denotes any of the statistics $\hat{D}_m(k), \hat{D}_m^{\mathrm{SN}}(k), \hat{P}_m(k), \hat{P}_m^{\mathrm{SN}}(k)$ and $\hat{Q}_m(k)$ discussed in Section 3.1 and 3.2.

Algorithm 3.13 Let $\{X_1, \dots, X_m\}$ denote the "stable" training data satisfying (2.1).

> (*Initialization*) Choose a threshold function w_α such that the probability of type I error is asymptotically α. Further choose the factor T to determine how much longer the monitoring can be performed.
>
> (*Monitoring*) If X_{m+k} has been observed, compute the the statistic $\hat{S}_m(k)$ and reject the null hypothesis of no change in the parameter $\theta(F)$ if $\hat{S}_m(k) > w_\alpha(k/m)$. In this case stop monitoring. Otherwise, repeat this comparison with the next observation X_{m+k+1}.
>
> (*Stop*) If there has been no rejection at time $m + mT$, stop monitoring with observation X_{m+mT} and conclude that no change has occurred within the monitoring period.

4 Some specific change point problems

In this section we illustrate how the assumptions of Section 3 can be verified for concrete functionals. Exemplarily we consider the mean and variance, but similar arguments could be given for other functionals under consideration. To be precise consider a time series $\{X_t\}_{t\in\mathbb{Z}}$, which forms a physical system in the sense of Wu and Rosenblatt (2005), that is

$$X_t = \begin{cases} g(\varepsilon_t, \varepsilon_{t-1}, \dots) & \text{if } t < \lfloor mc \rfloor \\ h(\varepsilon_t, \varepsilon_{t-1}, \dots) & \text{if } t \geq \lfloor mc \rfloor \end{cases}, \tag{4.1}$$

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ denotes a sequence of i.i.d. random variables with values in some measure space $\mathbb{S}$ such that the functions $g, h : \mathbb{S}^{\mathbb{N}} \to \mathbb{R}^d$ are measurable. The functions g and h determine the phases of the physical system before and after the change at position $\lfloor mc \rfloor$ with $c > 1$, respectively. Under the null hypothesis we will always assume that g and h coincide, which yields that the (whole) times series $\{X_t\}_{t\in\mathbb{Z}}$ is strictly stationary. In the case $g \neq h$ the random variables X_t form a triangular array, but for the sake of readability we do not reflect this in our notation. In order to adapt the concept of physical dependence to the situation considered in this paper, let ε_0' be an independent copy of ε_0 and define

$$X_t' = \begin{cases} g(\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_1, \varepsilon_0', \varepsilon_{-1}, \dots) & \text{if } t < \lfloor mc \rfloor \\ h(\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_1, \varepsilon_0', \varepsilon_{-1}, \dots) & \text{if } t \geq \lfloor mc \rfloor \end{cases}. \tag{4.2}$$

The distance

$$\delta_{t,q} := \big(\mathbb{E}[|X_t - X_t'|^q]\big)^{1/q}, \tag{4.3}$$

between X_t and its counterpart X_t' is used to quantify the (temporal) dependence of the physical system and $\delta_{t,q}$ measures the influence of ε_0 on the random variable X_t. Further let

$$\Theta_{m,q} = \sum_{t=m}^{\infty} \delta_{t,q} \tag{4.4}$$

denote the tail sum of the coefficients (which might diverge). Additionally, we define the (ordinary) long-run variance matrix of the phases before and after the change by

$$\begin{aligned} \Gamma(g) &= \sum_{t \in \mathbb{Z}} \operatorname{Cov}\left(g(\varepsilon_0, \varepsilon_{-1}, \dots), g(\varepsilon_t, \varepsilon_{t-1}, \dots)\right) , \\ \Gamma(h) &= \sum_{t \in \mathbb{Z}} \operatorname{Cov}\left(h(\varepsilon_0, \varepsilon_{-1}, \dots), h(\varepsilon_t, \varepsilon_{t-1}, \dots)\right) . \end{aligned} \tag{4.5}$$

4.1 Sequential testing for changes in the mean vector

In this section we are interested in detecting changes in the mean

$$\mu(F_t) = \mathbb{E}[X_t] = \int_{\mathbb{R}^d} x dF_t(x) \ , \quad t = 1, 2, \dots . \tag{4.6}$$

of a d-dimensional time series $\{X_t\}_{t \in \mathbb{Z}}$. Sequential detection schemes for a change in the mean have been been investigated by Chu et al. (1996), Horváth et al. (2004) and Aue et al. (2009b) among others. We consider the closed-end-procedure developed in Section 3 and assume that the first m observations $X_1, \dots, X_m$ to be mean-stable.

As the mean functional (4.6) is linear the influence function is given by

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

and therefore Assumption 3.2 is obviously satisfied (note that $R_{i,j} = 0$ for all i, j). Assumption 3.1 reduces to Donsker's invariance principle, that is

$$\left\{ \frac{1}{\sqrt{m}} \sum_{t=1}^{\lfloor ms \rfloor} (X_t - \mathbb{E}[X_t]) \right\}_{s \in [0, T+1]} \overset{\mathcal{D}}{\Longrightarrow} \left\{ \Sigma_F^{1/2} W(s) \right\}_{s \in [0, T+1]} \tag{4.7}$$

in $\ell^\infty([0, T+1], \mathbb{R}^d)$ with $\Sigma_F = \sum_{t \in \mathbb{Z}} \operatorname{Cov}\left(X_0, X_t\right)$, which has been derived by Wu and Rosenblatt (2005) for physical systems under the assumption that $\Theta_{m,2} < \infty$ (see See Theorem 3 in this reference). Note that for this functional the (ordinary) long-run variance matrix $\Gamma(g)$ and Σ_F coincide (under stationarity).

If the alternative of a change in the mean at position $\lfloor mc \rfloor$ for some $c \in (1, T+1)$ holds, we may assume that

$$h = g + \Delta\mu \ , \tag{4.8}$$

where $\Delta\mu = \mathbb{E}[X_{\lfloor mc \rfloor}] - \mathbb{E}[X_{\lfloor mc \rfloor - 1}]$. Consequently, if $\Theta_{m,2} < \infty$, Assumption 3.6 is also satisfied with $W_1 = W = W_2$ (see also the discussion at the end of Remark 3.7). We summarize these observations in the following proposition.

Proposition 4.1 *Assume that* (4.1) *holds with* $\Theta_{m,2} < \infty$ *and further let* $\hat{\Sigma}_m$ *denote a consistent estimator of the (positive definite) long-run variance matrix* $\Sigma_{F^{(1)}}$ *(before the change) based on the observations* $X_1, \ldots, X_m$*.*

(a) If $g = h$*, then the assumptions of Theorem 3.5 and 3.10 are satisfied for the functional* (4.6)*. In other words: The sequential tests for a change in the mean based on the statistics* $\hat{D}$ *or* $\hat{D}^{\mathrm{SN}}$ *with* $\theta(F_t) = \int_{\mathbb{R}^d} x dF_t(x)$ *have asymptotic level* α*.*

(b) Let representation (4.8) *hold with* $\Delta\mu \neq 0$*, then the assumptions of Theorem 3.8 and 3.11 are satisfied for the functional* (4.6)*. In other words: The sequential tests for a change in the mean based on the statistics* $\hat{D}$ *or* $\hat{D}^{\mathrm{SN}}$ *are consistent.*

The finite sample properties of this test will be investigated in Section 5.1.

4.2 Sequential testing for changes in the variance

In this section, we focus on detecting changes in the variance. Following Aue et al. (2009a), who investigated this problem in the non-sequential case, we consider a time series $\{X_t\}_{t\in\mathbb{Z}}$ with common mean $\mu = \mathbb{E}_F[X]$ and define the functional

$$V(F) = \int_{\mathbb{R}^d} x x^\top dF(x) - \int_{\mathbb{R}^d} x dF(x) \int_{\mathbb{R}^d} x^\top dF(x) \ . \tag{4.9}$$

A careful but straightforward calculation shows that the corresponding influence function is given by

$$\begin{aligned} \mathcal{IF}(x, F, V) &= -\mathbb{E}_F[XX^\top] + 2\mathbb{E}_F[X]\mathbb{E}_F[X^\top] + xx^\top - \mathbb{E}_F[X]x^\top - x\mathbb{E}_F[X]^\top \\ &= (x - \mathbb{E}_F[X])(x - \mathbb{E}_F[X])^\top - V(F) \ . \end{aligned} \tag{4.10}$$

Hence the remainder term (under stationarity with $X_1 \sim F$) in equation (3.4) is given by

$$\begin{aligned} R_{i,j} &= V(\hat{F}_i^j) - V(F) - \frac{1}{j-i+1}\sum_{t=i}^{j} \mathcal{IF}(X_t, F, V) \\ &= \int_{\mathbb{R}^d} x x^\top d\hat{F}_i^j(x) - \int_{\mathbb{R}^d} x d\hat{F}_i^j(x) \int_{\mathbb{R}^d} x^\top d\hat{F}_i^j(x) - \frac{1}{j-i+1}\sum_{t=i}^{j}(X_i - \mathbb{E}[X_1])(X_i - \mathbb{E}[X_1])^\top \\ &= -\int_{\mathbb{R}^d} x d\hat{F}_i^j(x) \int_{\mathbb{R}^d} x^\top d\hat{F}_i^j(x) + \int_{\mathbb{R}^d} x\mathbb{E}[X_1^\top] d\hat{F}_i^j(x) + \int_{\mathbb{R}^d} \mathbb{E}[X_1] x^\top d\hat{F}_i^j(x) - \mathbb{E}[X_1]\mathbb{E}[X_1^\top] \\ &= -\left(\int_{\mathbb{R}^d} x - \mathbb{E}[X_1] d\hat{F}_i^j(x)\right)\left(\int_{\mathbb{R}^d} x^\top - \mathbb{E}[X_1^\top] d\hat{F}_i^j(x)\right) \ . \end{aligned} \tag{4.11}$$

Define vech$(\cdot)$ to be the operator that stacks the columns of a symmetric $d \times d$-matrix above the diagonal as a vector of dimension $d(d+1)/2$. As this operator is linear, it is obvious that expansion (4.11) is equivalent to

$$\text{vech}(R_{i,j}) = \text{vech}(V(\hat{F}_i^j)) - \text{vech}(V) - \frac{1}{j-i+1}\sum_{t=i}^{j}\mathcal{IF}_v(X_t, F, V) \ , \tag{4.12}$$

where $\mathcal{IF}_v$ is defined as

$$\mathcal{IF}_v(X_t, F, V) = \text{vech}\left(\mathcal{IF}(X_t, F, V)\right) = \mathcal{IF}\big(X_t, F, \text{vech}(V)\big) \tag{4.13}$$

$$= \begin{pmatrix} (X_{t,1} - \mathbb{E}[X_{t,1}])^2 - \mathbb{E}\big[(X_{t,1} - \mathbb{E}[X_{t,1}])^2\big] \\ (X_{t,1} - \mathbb{E}[X_{t,1}])(X_{t,2} - \mathbb{E}[X_{t,2}]) - \mathbb{E}\big[(X_{t,1} - \mathbb{E}[X_{t,1}])(X_{t,2} - \mathbb{E}[X_{t,2}])\big] \\ (X_{t,2} - \mathbb{E}[X_{t,2}])^2 - \mathbb{E}\big[(X_{t,2} - \mathbb{E}[X_{t,2}])^2\big] \\ (X_{t,1} - \mathbb{E}[X_{t,1}])(X_{t,3} - \mathbb{E}[X_{t,3}]) - \mathbb{E}\big[(X_{t,1} - \mathbb{E}[X_{t,1}])(X_{t,3} - \mathbb{E}[X_{t,3}])\big] \\ (X_{t,2} - \mathbb{E}[X_{t,2}])(X_{t,3} - \mathbb{E}[X_{t,3}]) - \mathbb{E}\big[(X_{t,2} - \mathbb{E}[X_{t,2}])(X_{t,3} - \mathbb{E}[X_{t,3}])\big] \\ (X_{t,3} - \mathbb{E}[X_{t,3}])^2 - \mathbb{E}\big[(X_{t,3} - \mathbb{E}[X_{t,3}])^2\big] \\ \vdots \end{pmatrix} ,$$

and $X_{t,h}$ denotes the h-th component of the vector X_i. We now provide sufficient conditions such that the general theory in Section 3 is applicable for the functional vech(V). Assumption 3.1 is satisfied if the time series $\{X_t\}_{t\in\mathbb{Z}}$ is stationary and the invariance principle

$$\frac{1}{\sqrt{m}}\sum_{t=1}^{\lfloor ms \rfloor}\mathcal{IF}_v(X_t, F, V) \overset{\mathcal{D}}{\Longrightarrow} \sqrt{\Sigma_F}W(s) \ , \tag{4.14}$$

holds in the space $\ell^\infty(\mathbb{R}, \mathbb{R}^{d^*})$, where W is a $d^* = d(d+1)/2$-dimensional Brownian motion and Σ_F is defined in (3.3) with $\mathcal{IF} = \mathcal{IF}_v$. Invariance principles of the form (4.14) are well known for many classes of weakly dependent time series. The required assumptions for the underlying time series $\{X_t\}_{t\in\mathbb{Z}}$ are typically the same as for the mean - except for some extra moment conditions to cover the product structure of the random variables in (4.13). Condition (3.5) in Assumption 3.2 reads as follows

$$\sup_{1\le i<j\le n}\frac{1}{j-i+1}\Big|\sum_{t=i}^{j}X_{t,k} - \mathbb{E}[X_{t,k}]\Big|\Big|\sum_{t=i}^{j}X_{t,\ell} - \mathbb{E}[X_{t,\ell}]\Big| = o_{\mathbb{P}}(n^{1/2}) \tag{4.15}$$

$(1 \le k, \ell \le d^*)$. The validity of this assumption depends on the underlying dependence structure, in particular of the properties of the functions g and h in (4.1), and exemplarily we give sufficient conditions in the following result, which is proved in the appendix.

Proposition 4.2 *Assume that* (4.1) *holds with bounded functions g and h and $\delta_{t,4} = \mathcal{O}(\rho^t)$ for some $\rho \in (0,1)$. Let $\hat{\Sigma}_m$ denote a consistent estimator of the long-run variance $\Sigma_{F^{(1)}}$ (before* *the change) based on the observations $X_1, \dots, X_m$. Further assume, that there exists a constant $\nu > 0$, such that $\Gamma(g) - \nu \cdot I_d$ and $\Gamma(h) - \nu \cdot I_d$ are positive definite matrices, where $\Gamma(g)$ and $\Gamma(h)$ are defined in (4.5).*

(a) If $g = h$, then the assumptions of Theorem 3.5 and 3.10 are satisfied for the functional (4.9). In other words: The sequential tests for a change in the variance based on the statistics $\hat{D}$ or $\hat{D}^{\mathrm{SN}}$ have asymptotic level α.

(b) If $h = A \cdot g$ for some non-singular matrix $A \in \mathbb{R}^{d \times d}$ with $A \cdot V(F^{(1)}) \cdot A^\top \neq V(F^{(1)})$, then the assumptions of Theorem 3.8 and 3.11 are satisfied for the mean functional (4.6). In other words: The sequential tests for a change in the variance based on the statistics $\hat{D}$ or $\hat{D}^{\mathrm{SN}}$ are consistent.

Remark 4.3 The assumption of bounded observations in Proposition 4.2 is crucial to prove the estimate (4.15). Essentially a proof of such a statement requires a version of Theorem 1 in Shao (1995) for dependent random variables. The main ingredient for a proof of Shao's result is an Erdös-Renyi-Law of large numbers in the case of dependent random variables, which - to the authors best knowledge - is only known for bounded random variables [see Kifer (2017) for example]. On the other hand the assumption of bounded functions g and h in (4.1) is not necessary for the functional (4.9) in the case of M-dependent time series.

5 Finite sample properties

In this section, we investigate the finite sample properties of the new detection schemes based on the statistics $\hat{D}$ and $\hat{D}^{\mathrm{SN}}$ in (2.9) and (3.20) and also provide a comparison to the detection schemes based on $\hat{Q}$, $\hat{P}$ and $\hat{P}^{\mathrm{SN}}$, which are defined in (2.13), (2.14) and (3.25), respectively. For the choice of the threshold function we follow the ideas of Horváth et al. (2004), Aue et al. (2009b) and Wied and Galeano (2013) and consider the parametric family

$$w(t) = (t+1)^2 \cdot \max\left\{ \left(\frac{t}{t+1} \right)^{2\gamma}, \delta \right\}, \tag{5.1}$$

where the parameter γ varies in the interval $[0, 1/2)$ and $\delta > 0$ is a small constant introduced to avoid problems in the denominator of the ratio considered in (3.14). For the statistic $\hat{Q}$ these threshold functions are motivated by the law of iterated logarithm and are used to reduce the stopping delay under the alternative hypothesis see Aue et al. (2009b) or Wied and Galeano (2013)].

Note that we use the squared versions of the thresholds from the cited references, since we consider statistics in terms of quadratic forms. To be precise consider three different threshold functions

(T1) $w(t) = c_\alpha$,

(T2) $w(t) = c_\alpha (t+1)^2$,

(T3) $w(t) = c_\alpha (t+1)^2 \cdot \max\left\{ \left(\frac{t}{t+1}\right)^{1/2}, 10^{-10} \right\}$,

where the constant c_α is chosen by Monte-Carlo simulations, such that

$$\mathbb{P}\left(\sup_{t\in[1,T+1]} \frac{L(t)}{w(t-1)} \right) = \alpha \ . \tag{5.2}$$

Here L denotes the limit process corresponding to the test statistic under consideration. For the estimation of the long-run variance we employ the well-known quadratic spectral kernel [see Andrews (1991)] and its implementation contained in the R-package 'sandwich' [see Zeileis (2004)]. For the sake of brevity, we will only display situations where the parameter T is fixed as $T = 1$, i.e the monitoring period will always have the same size as the historical data set. For the case $T \geq 2$ we obtained a similar picture, which will not be displayed here. All results that are presented below are based on 1000 simulation runs.

5.1 Changes in the mean

For the analysis of the new procedures in the problem of detecting changes in the mean we look at independent data, a MA(3)- and an AR(1)-process, that is

(M1) $X_t \sim \varepsilon_t$,

(M2) $X_j = 0.1X_{t-1} + \varepsilon_t$,

(M3) $X_j = \varepsilon_t + 0.3\varepsilon_{t-1} + 0.1\varepsilon_{t-2}$,

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of independent standard Gaussian random variables. In the case of the alternative hypothesis we consider the sequence

$$X_t^\mu = \begin{cases} X_t & \text{if } \ t < m + \lfloor \frac{m}{2} \rfloor \\ X_t + \mu & \text{if } \ t \geq m + \lfloor \frac{m}{2} \rfloor \end{cases}$$

for various values of μ. For all discussed detection schemes the empirical rejection probabilities for the models (M1) - (M3) and threshold functions (T1) - (T3) are shown in Figure 1 and 2 corresponding to the choice $m = 50$ and $m = 100$ as initial sample size. The results can be summarized as follows. The statistic $\hat{D}$ outperforms $\hat{P}$ and $\hat{Q}$ with respect to the power for all combinations of the model and threshold function. Further the statistic $\hat{P}$ shows a better performance with respect to power as $\hat{Q}$ in all cases under consideration. For example, the plot in the left-upper corner of Figure 1 shows, that $\hat{D}$ already has empirical power close to 1 (0.94) for a change of size 1, while $\hat{P}$ and $\hat{Q}$ have only empirical power of 0.85 and 0.73, respectively. This relation is basically the same in all plots contained in Figures 1 and 2. Moreover the results show no substantial differences between the different threshold functions and for this reason we will only consider the constant threshold (T1) in the remaining part of this paper.

In Figures 3 we compare the power of the tests based on the self-normalized statistics $\hat{D}^{\mathrm{SN}}$ and $\hat{P}^{\mathrm{SN}}$. The results are similar as before and the empirical power obtained by the use of $\hat{D}^{\mathrm{SN}}$ is considerably higher. Only for model (M3) and $m = 50$ the statistic based on $\hat{P}^{\mathrm{SN}}$ is slightly more powerful for small values of μ, but this superiority is achieved by a to high nominal level (see the right panel in Figure 3). A comparison of the results with the corresponding rejection probabilities in Figures 1 and 2 shows, that self-normalization yields a substantial loss of power in the sequential detection schemes and we also observe this loss in the examples considered in Section 5.2 and 5.3 (these results are not displayed for the sake of brevity).

On the other hand the approximation of the nominal level is more stable with respect to different dependence structures for self-normalized methods. To illustrate this fact we display in Table 1 the type-I error for all five sequential monitoring schemes based on the statistics $\hat{D}$, $\hat{P}$ and $\hat{Q}$. The results provide some empirical evidence that the self-normalized statistics yield a more stable approximation of the nominal level.

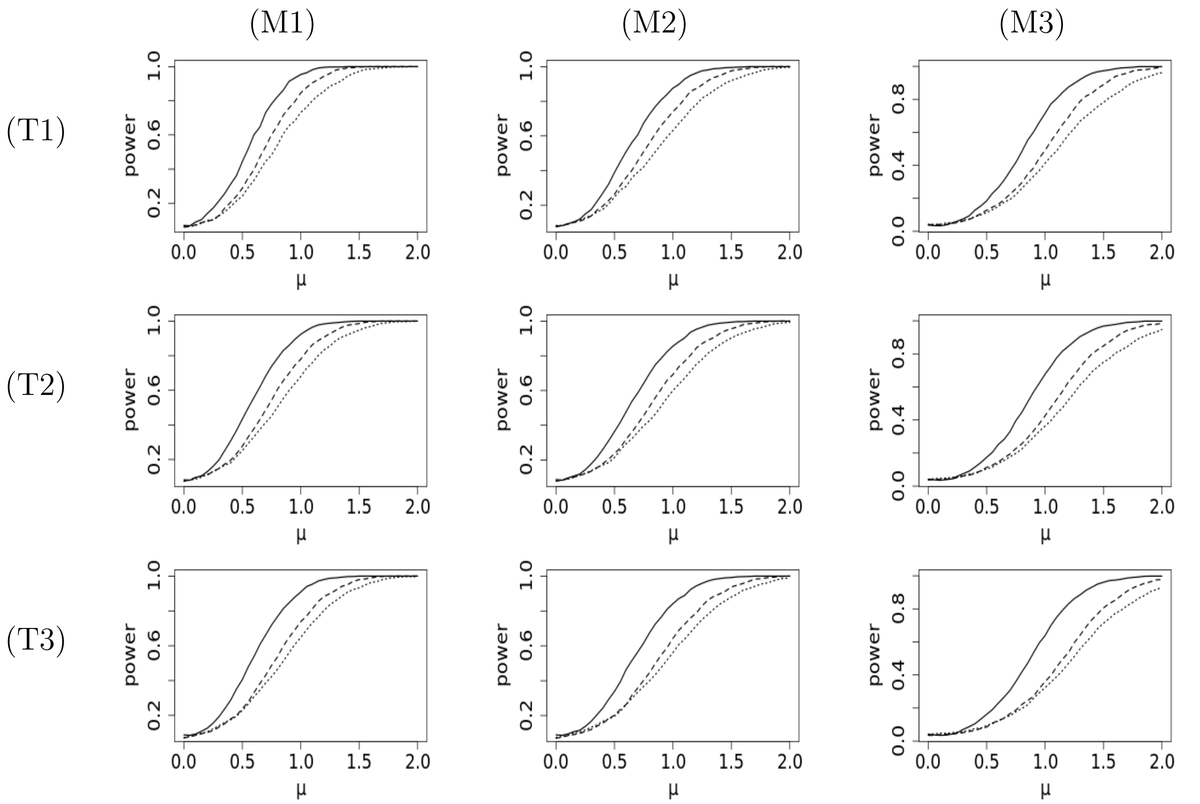

Figure 1: *Empirical rejection probabilities of the sequential tests for a change in the mean based on the statistics $\hat{D}$ (solid line), $\hat{P}$ (dashed line) , $\hat{Q}$ (dotted line). The initial and total sample size are $m = 50$ and $m(T + 1) = 100$, respectively, and the change occurs at observation 75. The level is $\alpha = 0.05$. Different rows correspond to different threshold functions, while different columns correspond to different models.*

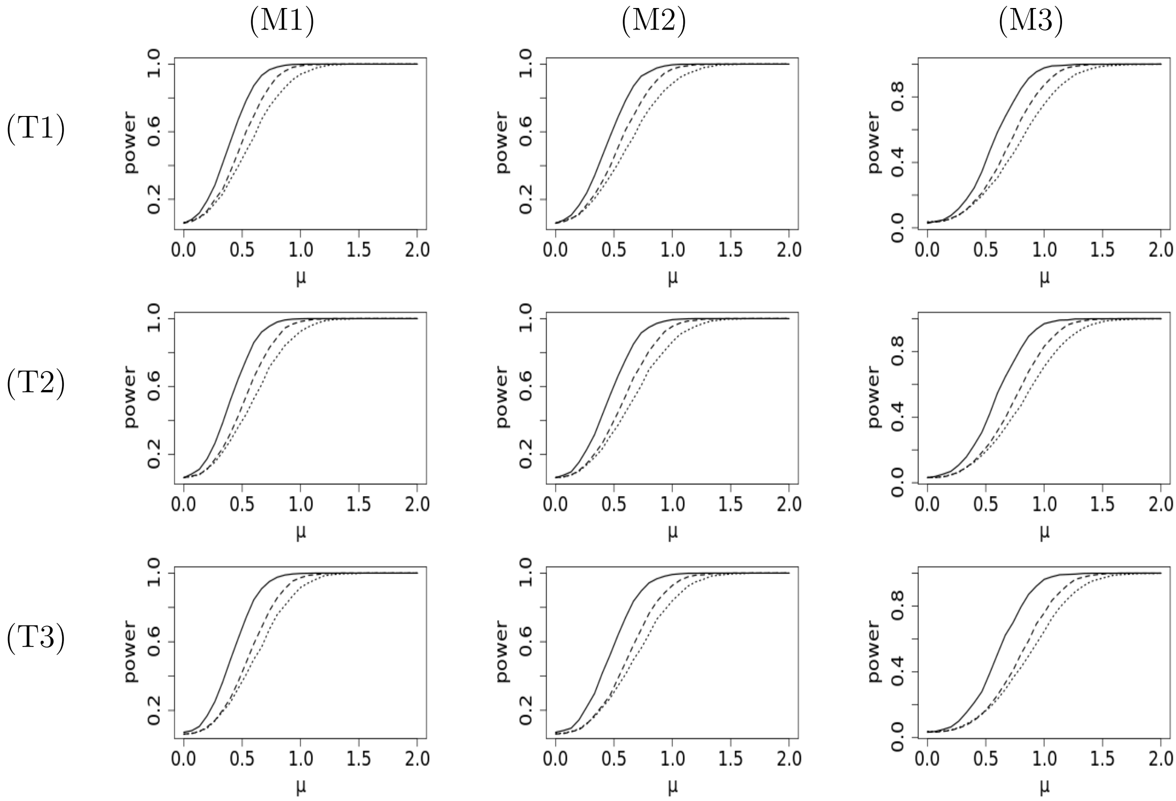

Figure 2: *Empirical rejection probabilities of the sequential tests for a change in the mean based on the statistics* $\hat{D}$ *(solid line),* $\hat{P}$ *(dashed line),* $\hat{Q}$ *(dotted line). The initial and total sample size are* $m = 100$ *and* $m(T+1) = 200$*, respectively, and the change occurs at observation* 150*. The level is* $\alpha = 0.05$*. Different rows correspond to different threshold functions, while different columns correspond to different models.*

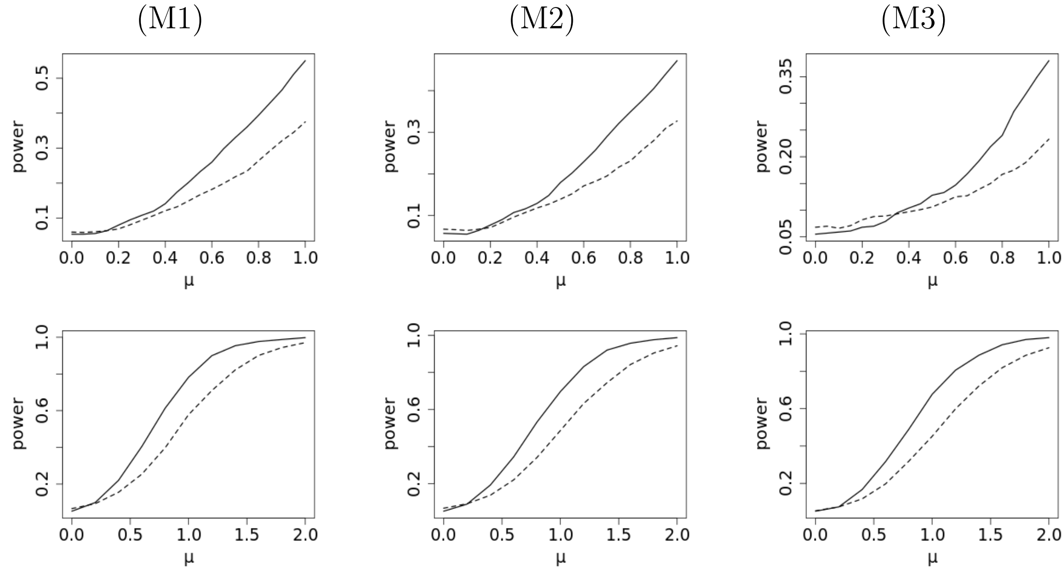

Figure 3: *Empirical rejection probabilities of the sequential tests for a change in the mean based on the self-normalized tests statistics* $\hat{D}^{\mathrm{SN}}$ *(solid line),* $\hat{P}^{\mathrm{SN}}$ *(dashed line). The initial and total sample size are* $m = 50$ *and* $m(T+1) = 100$ *(upper panel, change at observation* 75*) and* $m = 100$ *and* $m(T+1) = 200$ *(lower panel, change at observation* 150*). The level is* $\alpha = 0.05$ *and the threshold function is given by* (T1).

m	model \ statistic	$\hat{D}$	$\hat{P}$	$\hat{Q}$	$\hat{D}^{\mathrm{SN}}$	$\hat{P}^{\mathrm{SN}}$
50	(M1)	7.0%	6.0%	6.8%	5.4%	6.0%
	(M2)	8.1%	7.5%	8.2%	5.7%	6.7%
	(M3)	3.9%	4.0%	4.3%	5.5%	6.8%
100	(M1)	5.8%	6.0%	6.2%	5.2%	6.6%
	(M2)	5.9%	6.0%	6.0%	5.2%	6.8%
	(M3)	2.8%	3.4%	3.8%	5.1%	5.3%

Table 1: *Simulated type I error (level* $\alpha = 0.05$*) of the sequential tests for a change in the mean based on the statistics* $\hat{D}$, $\hat{P}$, $\hat{Q}$, $\hat{D}^{SN}$ *and* $\hat{P}^{SN}$. *The threshold function is (T1).*

5.2 Changes in the variance

In this subsection we present a small simulation study investigating the performance of the detection schemes for a change in the variance matrix. We consider the following models

(V1) $X_j = \varepsilon_j$,

(V2) $X_j = A_1 X_{j-1} + \varepsilon_j$,

(V3) $X_j = \varepsilon_j + A_2\varepsilon_{j-1} + A_3\varepsilon_{j-2}$.

(V1) (V2) (V3)

(T1)

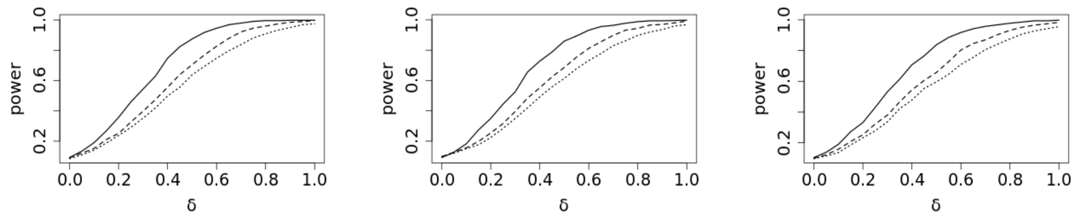

Figure 4: *Empirical rejection probabilities of the sequential tests for a change in the variance matrix based on the statistics $\hat{D}$ (solid line), $\hat{P}$ (dashed line), $\hat{Q}$ (dotted line). The initial and total sample size are $m = 200$ and $m(T+1) = 400$, respectively, and the change occurs at observation* 300. *The level is* $\alpha = 0.05$.

where $\{\varepsilon_j\}_{j\in\mathbb{Z}} = \{(\varepsilon_{j,1}, \varepsilon_{j,2})^\top\}_{j\in\mathbb{Z}}$ denotes a sequence of centered and independent two-dimensional Gaussian distributed random variables and the involved matrices are defined as

$$A_1 = \begin{pmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0.3 & 0.1 \\ 0.1 & 0.3 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 0.1 & 0.05 \\ 0.05 & 0.1 \end{pmatrix}. \tag{5.3}$$

For the alternative, we proceed similarly as in Aue et al. (2009a) and define

$$\mathrm{Cov}(\varepsilon_j, \varepsilon_j) = \varepsilon_j \cdot \varepsilon_j^\top = \begin{cases} I_2 & \text{if } j \le m + \lfloor \frac{m}{2} \rfloor \\ I_2 + \delta \cdot I_2 & \text{if } j > m + \lfloor \frac{m}{2} \rfloor \end{cases}, \tag{5.4}$$

where I_2 denotes the two dimensional identity matrix (the case $\delta = 0$ corresponds to the null hypothesis of no change). For the sake of brevity we will focus on the non-self-normalized statistics $\hat{D}$, $\hat{P}$ and $\hat{Q}$ here. In Figure 4 we display the empirical power for the three data generation processes and the threshold function (T1). The results are similar to those presented in Section 5.1. The test based on the statistic $\hat{D}$ is more powerful than the tests based on $\hat{P}$ and $\hat{Q}$.

5.3 Changes in the correlation

We conclude this paper with a brief comparison of the three methods for the detection of a change in the correlation, which has been considered in Wied and Galeano (2013). For the definition of the data generating processes, we use the models (V1) - (V3) introduced in Section 5.2 but with a different process $\{\varepsilon_j\}_{j\in\mathbb{Z}} = \{(\varepsilon_{j,1}, \varepsilon_{j,2})^\top\}_{j\in\mathbb{Z}}$. In this section $\{\varepsilon_j\}_{j\in\mathbb{Z}}$ is a sequence of independent two-dimensional Gaussian random variables such that

$$\mathrm{Cor}(\varepsilon_{j,1}, \varepsilon_{j,2}) = \begin{cases} c_1 & \text{if} \quad j \le m + \lfloor \frac{m}{2} \rfloor , \\ c_2 & \text{if} \quad j > m + \lfloor \frac{m}{2} \rfloor \end{cases}$$

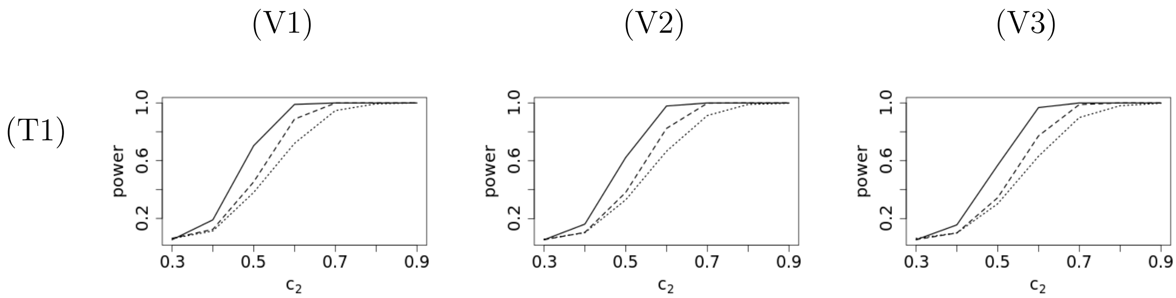

Figure 5: *Empirical rejection probabilities of the sequential tests for a change in the correlation based on the statistics* $\hat{D}$ *(solid line),* $\hat{P}$ *(dashed line) ,* $\hat{Q}$ *(dotted line). The initial and total sample size are* $m = 500$ *and* $n = 1000$*, respectively, and the change occurs at observation* 750*. The level is* $\alpha = 0.05$.

and $\mathrm{Var}(\epsilon_{j,1}) = \mathrm{Var}(\epsilon_{j,2}) = 1$. We use $c_1 = 0.3$ for the correlation before the change and consider different values of c_2. For estimation of the long-run variance matrix we use the estimator proposed in Wied and Galeano (2013) (the explicit formula for the estimator is given in the appendix of this paper and omitted here for the sake of brevity). Figure 5 now compares the power of the non-self-normalized methods for the three models defined above. As in the previous sections the sequential detection scheme based on $\hat{D}$ yields substantially better results.

Acknowledgments. 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). The authors would like to thank Claudia Kirch and Dominik Wied for some helpful discussions on this subject.

References

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

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):804837.

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

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

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

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.

Davis, R. A., Huang, D., and Yao, Y.-C. (1995). Testing for a change in the parameter values and order of an autoregressive model. *The Annals of Statistics*, 23(1):282–304.

Dette, H. and Wied, D. (2016). Detecting relevant changes in time series models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(2):371–394.

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

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.

Jandhyala, V., Fotopoulos, S., MacNeill, I., and Liu, P. (2013). Inference for single and multiple change-points in time series. *Journal of Time Series Analysis*, 34(4):423–446.

Jirak, M. (2015). Uniform change point tests in high dimension. *The Annals of Statistics*, 43(6):2451–2483.

Kifer, Y. (2017). Functional Erdös-Rényi law of large numbers for nonconventional sums under weak dependence. *Electronic Journal of Probability*, 22(23):1–17.

Kirch, C. and Weber, S. (2017). Modified sequential change point procedures based on estimating functions. Preprint, Otto-von-Guericke-Universitaet Magdeburg.

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.

Moustakides, G. V. (1986). Optimal stopping times for detecting changes in distributions. *The Annals of Statistics*, 14(4):1379–1387.

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

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

Shao, Q.-M. (1995). On a conjecture of Révész. *Proceedings of the American Mathematical Society*, 123(2):575–582.

Shao, X. (2015). Self-normalization for time series: A review of recent developments. *Journal of the American Statistical Association*, 110(512):1797–1817.

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

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.

Wu, W. B. and Rosenblatt, M. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America*, 102(40):14150–14154.

Wu, W. B. and Zhou, Z. (2011). Gaussian approximations for non-stationary multiple time series. *Statistica Sinica*, 21(3):1397–1413.

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

A Technical details

Proof of Theorem 3.3. From (3.4), (3.6) we obtain the representation

$$
\begin{aligned}
& m^{-3/2}\tilde{\mathbb{U}}(\lfloor mr\rfloor, \lfloor ms\rfloor, \lfloor mt\rfloor) = m^{-3/2}(\lfloor mt\rfloor - \lfloor ms\rfloor)(\lfloor ms\rfloor - \lfloor mr\rfloor)\left(\hat{\theta}_{\lfloor mr+1\rfloor}^{\lfloor ms\rfloor} - \hat{\theta}_{\lfloor ms\rfloor+1}^{\lfloor mt\rfloor}\right) \\
&= \frac{\lfloor mt\rfloor - \lfloor ms\rfloor}{m^{3/2}} \sum_{i=\lfloor mr\rfloor+1}^{\lfloor ms\rfloor} \mathcal{IF}(X_i, F, \theta) - \frac{\lfloor ms\rfloor - \lfloor mr\rfloor}{m^{3/2}} \sum_{t=\lfloor ms\rfloor+1}^{\lfloor mt\rfloor} \mathcal{IF}(X_i, F, \theta) \\
&+ \frac{(\lfloor mt\rfloor - \lfloor ms\rfloor)(\lfloor ms\rfloor - \lfloor mr\rfloor)}{m^{3/2}} \left(R_{\lfloor mr\rfloor+1, \lfloor ms\rfloor} - R_{\lfloor ms\rfloor+1, \lfloor mt\rfloor}\right) .
\end{aligned}
$$

By Assumption 3.1 we have

$$
\begin{aligned}
&\left\{\frac{\lfloor mt\rfloor - \lfloor ms\rfloor}{m^{3/2}} \sum_{i=\lfloor mr\rfloor+1}^{\lfloor ms\rfloor} \mathcal{IF}(X_i, F, \theta) - \frac{\lfloor ms\rfloor - \lfloor mr\rfloor}{m^{3/2}} \sum_{i=\lfloor ms\rfloor+1}^{\lfloor mt\rfloor} \mathcal{IF}(X_i, F, \theta)\right\}_{(r,s,t)\in\Delta_3} \\
&\qquad \overset{\mathcal{D}}{\Longrightarrow} \Sigma_F^{1/2}\Big\{(t-s)\big(W(s) - W(r)\big) - (s-r)\big(W(t) - W(s)\big)\Big\}_{(r,s,t)\in\Delta_3} \\
&\qquad = \Sigma_F^{1/2}\big\{B(s,t) + B(r,s) - B(r,t)\big\}_{(r,s,t)\in\Delta_3} ,
\end{aligned}
$$

where we use the definition of the process B in (3.11) and the fact

$$
\sup_{(s,t)\in\Delta_2} \left|\frac{\lfloor mt\rfloor - \lfloor ms\rfloor}{m} - (t-s)\right| \leq \frac{2}{m} = o(1) .
$$

Finally, Assumption 3.2 yields

$$
\frac{(\lfloor mt\rfloor - \lfloor ms\rfloor)(\lfloor ms\rfloor - \lfloor mr\rfloor)}{m^{3/2}} \left(R_{\lfloor mr\rfloor+1, \lfloor ms\rfloor} - R_{\lfloor ms\rfloor+1, \lfloor mt\rfloor}\right) = o_p(1) ,
$$

uniformly with respect to $(r,s,t) \in \Delta_3$ so that Theorem 3.3 follows. □

Proof of Corollary 3.5. Define

$$
D_m(k) = m^{-3} \max_{j=0}^{k-1} |\mathbb{U}^\top(m+j, m+k)\Sigma_F^{-1}\mathbb{U}(m+j, m+k)| . \tag{A.1}
$$

Using the fact, that the detection scheme D_m is piecewise constant and the monotonicity of the threshold function we obtain the representation

$$
\max_{k=0}^{Tm} \frac{D_m(k)}{w(k/m)} = \sup_{t\in[0,T]} \frac{D_m(\lfloor mt\rfloor)}{w(t)} = \sup_{t\in[1,T+1]} \sup_{s\in[1,t]} \frac{m^{-3}\left|\mathbb{U}^\top(\lfloor ms\rfloor, \lfloor mt\rfloor)\Sigma_F^{-1}\mathbb{U}(\lfloor ms\rfloor, \lfloor mt\rfloor)\right|}{w(t-1)} .
$$

By Remark 3.4 it and the continuous mapping theorem we have

$$
\max_{k=1}^{Tm} \frac{D_m(k)}{w(k/m)} \overset{\mathcal{D}}{\Longrightarrow} \sup_{t\in[1,T+1]} \sup_{s\in[1,t]} \frac{B(s,t)^\top B(s,t)}{w(t-1)} ,
$$

where the process B is defined in (3.11). The result now follows from Remark 3.4, the fact, that w_α has a lower bound and that $\hat{\Sigma}_m$ is a consistent estimate of the matrix Σ_F, which implies (observing the definition of $\hat{D}$ in (A.1))

$$\begin{aligned}\Big|\max_{k=1}^{Tm}\frac{D_m(k)}{w(k/m)}-\max_{k=1}^{Tm}\frac{\hat{D}_m(k)}{w(k/m)}\Big| &\le \max_{k=1}^{Tm}\big|D_m(k)-\hat{D}_m(k)\big|\\ &\le \|\hat{\Sigma}_m^{-1}-\Sigma_F^{-1}\|_{op}\sup_{t\in[1,T+1]}\sup_{s\in[1,t]}|m^{-3/2}\mathbb{U}(\lfloor ms\rfloor,\lfloor mt\rfloor)|^2\\ &= o_{\mathbb{P}}(1)\ .\end{aligned}$$

Here $\|\cdot\|_{op}$ denotes the operator norm and we have used the estimate $\|\hat{\Sigma}_m^{-1}-\Sigma_F^{-1}\|_{op}=o_{\mathbb{P}}(1)$, which is a consequence of the Continuous Mapping Theorem. □

Proof of Theorem 3.8. By the definition of the statistic $\hat{D}$ in (2.9), we obtain

$$\max_{k=0}^{Tm}\frac{\hat{D}_m(k)}{w_\alpha(k/m)}\ge m^{-3}\frac{\big|\mathbb{U}^\top(\lfloor mc\rfloor,m(T+1))\hat{\Sigma}_m^{-1}\mathbb{U}(\lfloor mc\rfloor,m(T+1)\big|}{w_\alpha(T)}\ , \tag{A.2}$$

where $\lfloor mc\rfloor$ denotes the (unknown) location of the change. We can apply expansion (3.4) to $X_1,\dots,X_{\lfloor mc\rfloor}$ and $X_{\lfloor mc\rfloor+1},\dots,X_{\lfloor mT\rfloor}$ and obtain

$$\begin{aligned}m^{-3/2}\mathbb{U}(\lfloor mc\rfloor,m(T+1)) &= \frac{\lfloor mc\rfloor\big(m(T+1)-\lfloor mc\rfloor\big)}{m^{3/2}}\Big(\hat{\theta}_1^{\lfloor mc\rfloor}-\hat{\theta}_{\lfloor mc\rfloor+1}^{m(T+1)}\Big)\\ &= \frac{m(T+1)-\lfloor mc\rfloor}{m^{3/2}}\sum_{i=1}^{\lfloor mc\rfloor}\mathcal{IF}(X_i,F^{(1)},\theta_{F^{(1)}})\\ &-\frac{\lfloor mc\rfloor}{m^{3/2}}\sum_{i=\lfloor mc\rfloor+1}^{m(T+1)}\mathcal{IF}(X_i,F^{(2)},\theta_{F^{(2)}})\\ &+\frac{\lfloor mc\rfloor\big(m(T+1)-\lfloor mc\rfloor\big)}{m^{3/2}}\Big(\theta_{F^{(1)}}-\theta_{F^{(2)}}+R_{1,\lfloor mc\rfloor}^{(F^{(1)})}-R_{\lfloor mc\rfloor+1,m(T+1)}^{(F^{(2)})}\Big)\ ,\end{aligned}$$

where $\theta_{F^{(\ell)}}=\theta(F^{(\ell)})$ $(\ell=1,2)$. Using Assumption 3.6 we obtain the joint convergence of

$$\frac{1}{m^{3/2}}\begin{pmatrix}\big(m(T+1)-\lfloor mc\rfloor\big)\sum_{i=1}^{\lfloor mc\rfloor}\mathcal{IF}(X_i,F^{(1)},\theta_{F^{(1)}})\\ \lfloor mc\rfloor\sum_{i=\lfloor mc\rfloor+1}^{m(T+1)}\mathcal{IF}(X_i,F^{(2)},\theta_{F^{(2)}})\end{pmatrix}\overset{\mathcal{D}}{\Longrightarrow}\begin{pmatrix}(T+1-c)\sqrt{\Sigma_{F^{(1)}}}W_1(c)\\ c\sqrt{\Sigma_{F^{(2)}}}\Big(W_2(T+1)-W_2(c)\Big)\end{pmatrix}$$

and

$$\frac{\lfloor mc\rfloor m(T+1)}{m^{3/2}}\Big(R_{1,\lfloor mc\rfloor}^{(F^{(1)})}-R_{\lfloor mc\rfloor+1,m(T+1)}^{(F^{(2)})}\Big)\xrightarrow{\mathbb{P}}0.$$

As $\theta_{F^{(1)}}\ne\theta_{F^{(2)}}$ this directly implies $m^{-3/2}|\mathbb{U}(\lfloor mc\rfloor,m(T+1))|\xrightarrow{\mathbb{P}}\infty$, and the assertion follows from (A.2) and the assumption that $\hat{\Sigma}_m$ is a consistent estimate for $\Sigma_{F^{(1)}}$. □

Proof of Theorem 3.10. Recalling the definition of $\tilde{\mathbb{U}}$ and $\mathbb{U}$ in (3.6) and (3.7), respectively, we obtain for the normalizing process $\mathbb{V}$ in (3.19) the representation

$$\begin{aligned}
m^{-4}\mathbb{V}(\lfloor ms\rfloor,\lfloor mt\rfloor) &= m^{-4}\sum_{j=1}^{\lfloor ms\rfloor} j^2(\lfloor ms\rfloor - j)^2\Big(\hat\theta_1^j - \hat\theta_{j+1}^{\lfloor ms\rfloor}\Big)\Big(\hat\theta_1^j - \hat\theta_{j+1}^{\lfloor ms\rfloor}\Big)^\top \\
&+ m^{-4}\sum_{j=\lfloor ms\rfloor+1}^{\lfloor mt\rfloor} (\lfloor mt\rfloor - j)^2(j-\lfloor ms\rfloor)^2\Big(\hat\theta_{\lfloor ms\rfloor+1}^j - \hat\theta_{j+1}^{\lfloor mt\rfloor}\Big)\Big(\hat\theta_{\lfloor ms\rfloor+1}^j - \hat\theta_{j+1}^{\lfloor mt\rfloor}\Big)^\top \\
&= m^{-4}\sum_{j=1}^{\lfloor ms\rfloor}\mathbb{U}(j,\lfloor ms\rfloor)\mathbb{U}^\top(j,\lfloor ms\rfloor) \\
&+ m^{-4}\sum_{j=\lfloor ms\rfloor+1}^{\lfloor mt\rfloor}\tilde{\mathbb{U}}(\lfloor ms\rfloor,j,\lfloor mt\rfloor)\tilde{\mathbb{U}}^\top(\lfloor ms\rfloor,j,\lfloor mt\rfloor) \\
&= m^{-3}\int_0^s \mathbb{U}(\lfloor mr\rfloor,\lfloor ms\rfloor)\mathbb{U}^\top(\lfloor mr\rfloor,\lfloor ms\rfloor)dr \\
&+ m^{-3}\int_s^t \tilde{\mathbb{U}}(\lfloor ms\rfloor,\lfloor mr\rfloor,\lfloor mt\rfloor)\tilde{\mathbb{U}}^\top(\lfloor ms\rfloor,\lfloor mr\rfloor,\lfloor mt\rfloor)dr\ .
\end{aligned}$$

By Theorem 3.3 we have

$$\big\{m^{-3/2}\tilde{\mathbb{U}}_m(\lfloor mr\rfloor,\lfloor ms\rfloor,\lfloor mt\rfloor)\big\}_{(r,s,t)\in\Delta_3} \overset{\mathcal{D}}{\Longrightarrow} \Sigma_F^{1/2}\big\{B(s,t)+B(r,s)-B(r,t)\big\}_{(r,s,t)\in\Delta_3} \tag{A.3}$$

in the space $\ell^\infty(\Delta_3,\mathbb{R}^p)$, where the process B is defined in (3.11). Consequently, the Continuous Mapping Theorem yields

$$\left\{\begin{pmatrix} m^{-3/2}\cdot\mathbb{U}(\lfloor ms\rfloor,\lfloor mt\rfloor) \\ m^{-4}\cdot\mathbb{V}(\lfloor ms\rfloor,\lfloor mt\rfloor)\end{pmatrix}\right\}_{(s,t)\in\Delta_2} \overset{\mathcal{D}}{\Longrightarrow} \left\{\begin{pmatrix} \Sigma_F^{1/2}B(s,t) \\ \Sigma_F^{1/2}\big(N_1(s)+N_2(s,t)\big)\Sigma_F^{1/2}\end{pmatrix}\right\}_{(s,t)\in\Delta_2}, \tag{A.4}$$

where N_1, N_2 are defined in (3.22). Now the assertion of Theorem 3.10 follows by a further application of the Continuous Mapping Theorem. □

Proof of Theorem 3.11. By definition of the self-normalized statistic $\hat D^{\mathrm{SN}}$ in (3.20), we obtain

$$\max_{k=0}^{Tm}\frac{\hat D_m^{\mathrm{SN}}(k)}{w_\alpha(k/m)} \ge m\cdot\frac{\big|\mathbb{U}^\top(\lfloor mc\rfloor,m(T+1))\mathbb{V}^{-1}(\lfloor mc\rfloor,m(T+1))\mathbb{U}(\lfloor mc\rfloor,m(T+1)\big|}{w_\alpha(T)}\ , \tag{A.5}$$

where $\lfloor mc\rfloor$ denotes the (unknown) location of the change. The discussion in the proof of Theorem 3.8 shows

$$m^{-3/2}\mathbb{U}(\lfloor mc\rfloor,m(T+1)) \xrightarrow{\mathbb{P}} \infty\ .$$

The proof will be completed by inspecting the random variable $\mathbb{V}^{-1}(\lfloor mc\rfloor, m(T+1))$ in the lower bound in (A.5). Repeating again the arguments from the proof of Theorem 3.3 we can rewrite

$$\begin{aligned} m^{-4}\cdot\mathbb{V}(\lfloor mc\rfloor, m(T+1)) = m^{-3}\int_0^c \mathbb{U}(\lfloor mr\rfloor,\lfloor ms\rfloor)\mathbb{U}^\top(\lfloor mr\rfloor,\lfloor ms\rfloor)dr \\ + m^{-3}\int_c^{T+1}\tilde{\mathbb{U}}(\lfloor ms\rfloor,\lfloor mr\rfloor,\lfloor mt\rfloor)\tilde{\mathbb{U}}^\top(\lfloor ms\rfloor,\lfloor mr\rfloor,\lfloor mt\rfloor)dr\ . \end{aligned} \tag{A.6}$$

Using Assumption 3.6 and employing the arguments from the proof of Theorem 3.3 we obtain weak convergence of

$$\begin{pmatrix} \{\mathbb{U}(\lfloor mr\rfloor,\lfloor ms\rfloor)\}_{0\le r\le s\le c} \\ \{\tilde{\mathbb{U}}(\lfloor ms\rfloor,\lfloor mr\rfloor,\lfloor mt\rfloor)\}_{c\le s\le r\le t\le c}\end{pmatrix} \overset{\mathcal{D}}{\Longrightarrow} \begin{pmatrix} \{B^{(1)}(r,s)\}_{0\le r\le s\le c} \\ \{B^{(2)}(r,t)+B^{(2)}(s,r)-B^{(2)}(s,t)\}_{c\le s\le r\le t\le T+1}\end{pmatrix}\ ,$$

where we use the extra definition

$$B^{(\ell)}(s,t) = tW_\ell(s) - sW_\ell(t) \qquad \ell = 1,2$$

and W_1 and W_2 are defined in Assumption 3.6. By the Continuous Mapping Theorem and the representation in (A.6) this implies

$$m^{-4}\cdot\mathbb{V}(\lfloor mc\rfloor, m(T+1)) \overset{\mathcal{D}}{\Longrightarrow} \Sigma_{F^{(1)}}^{1/2}\big(N_1^{(1)}(c)\big)\Sigma_{F^{(2)}}^{1/2} + \Sigma_{F^{(2)}}^{1/2}\big(N_2^{(2)}(c,T+1)\big)\Sigma_{F^{(2)}}^{1/2}\ ,$$

where the processes $N_1^{(1)}$ and $N_2^{(2)}$ are distributed like N_1 and N_2 in (3.22) but with respect to the processes $B^{(1)}$ and $B^{(2)}$, respectively. □

Proof of Proposition 4.2. For the sake of readability, we will give the proof only for the case $d = 2$. The arguments presented here can be easily extended to higher dimension. In view of the representation in (4.13), we may also assume without loss of generality that $\mu = \mathbb{E}[X_t] = 0$.

Part (a) of the proposition is a consequence of the discussion after Corollary 3.5 provided that Assumptions 3.1 and 3.2 can be established. For this purpose we introduce the notation

$$Z_t := \mathcal{IF}_v(X_t, F, V) = \begin{pmatrix} X_{t,1}^2 - \mathbb{E}[X_{t,1}^2] \\ X_{t,1}X_{t,2} - \mathbb{E}[X_{t,1}X_{t,2}] \\ X_{t,2}^2 - \mathbb{E}[X_{t,2}^2]\end{pmatrix}$$

and note that the time series $\{Z_t\}_{t\in\mathbb{Z}}$ can be represented as a physical system, that is

$$Z_t = \begin{pmatrix} g_1^2(\varepsilon_t,\dots) - \mathbb{E}[X_{1,1}^2] \\ g_1(\varepsilon_t,\dots)g_2(\varepsilon_t,\dots) - \mathbb{E}[X_{1,1}X_{1,2}] \\ g_2^2(\varepsilon_t,\dots) - \mathbb{E}[X_{1,2}^2]\end{pmatrix} := G(\varepsilon_t,\varepsilon_{t-1},\dots)\ , \tag{A.7}$$

where g_i denotes the i-th component of the function g in (4.1). The corresponding physical dependence coefficients $\delta_{t,2}^Z$ in (4.3) are given by

$$\begin{aligned}\delta_{t,2}^Z &= \Big\| \sqrt{(X_{t,1}^2 - (X_{t,1}')^2)^2 + (X_{t,2}^2 - (X_{t,2}')^2)^2 + (X_{t,1}X_{t,2} - X_{t,1}'X_{t,2}')^2} \Big\|_2 \\ &\leq C \cdot \Big(\|X_{t,1}^2 - (X_{t,1}')^2\|_2 + \|X_{t,2}^2 - (X_{t,2}')^2\|_2 + \|X_{t,1}X_{t,2} - X_{t,1}'X_{t,2}'\|_2 \Big) \\ &\leq C \cdot \max \Big\{ \|X_{t,1}^2 - (X_{t,1}')^2\|_2 \,, \|X_{t,2}^2 - (X_{t,2}')^2\|_2 \,, \|X_{t,1}X_{t,2} - X_{t,1}'X_{t,2}'\|_2 \Big\} \,,\end{aligned}$$

where $C > 0$ denotes a sufficiently large constant. Now Hölder's inequality yields (for another appropriate constant C)

$$\begin{aligned}\|X_{t,1}^2 - (X_{t,1}')^2\|_2 &\leq \|X_{t,1} + X_{t,1}'\|_4 \|X_{t,1} - X_{t,1}'\|_4 \leq C \cdot \delta_{t,4} \,, \\ \|X_{t,1}X_{t,2} - X_{t,1}'X_{t,2}'\|_2 &\leq \big\|X_{t,1}\big(X_{t,2} - X_{t,2}'\big)\big\|_2 + \big\|X_{t,2}'\big(X_{t,1} - X_{t,1}'\big)\big\|_2 \\ &\leq \big\|X_{t,1}\big\|_4 \big\|X_{t,2} - X_{t,2}'\big\|_4 + \big\|X_{t,2}'\big\|_4 \big\|X_{t,1} - X_{t,1}'\big\|_4 \leq C \cdot \delta_{t,4} \,.\end{aligned}$$

Combining these results gives $\sum_{t=m}^{\infty} \delta_{t,2}^Z \leq C \cdot \Theta_{m,4} < \infty$ and Theorem 3 from Wu and Rosenblatt (2005) implies the weak convergence

$$\frac{1}{\sqrt{m}} \sum_{t=1}^{\lfloor ms \rfloor} \mathcal{IF}_v(X_t, F, V) = \frac{1}{\sqrt{m}} \sum_{t=1}^{\lfloor ms \rfloor} Z_t \overset{\mathcal{D}}{\Longrightarrow} \sqrt{\Sigma_F} W(s) \,,$$

in the space $\ell^\infty([0, T+1], \mathbb{R}^3)$ as $m \to \infty$, where Σ_F is the long-run variance matrix defined in (3.3). Therfore Assumption 3.1 is satisfied.

To finish part (a) it remains to show that Assumption 3.2 holds. Due to (4.15) this is a consequence of

$$\sup_{1 \leq i < j \leq n} \frac{1}{\sqrt{j - i + 1}} \Big| \sum_{t=i}^{j} X_{t,\ell} - \mathbb{E}[X_{t,\ell}] \Big| = o_{\mathbb{P}}(n^{1/4}) \tag{A.8}$$

for $\ell = 1, 2, 3$. Since the arguments are exactly the same, we will only elaborate the case $\ell = 1$. For this purpose let

$$S_i = \sum_{t=1}^{i} X_{t,1} - \mathbb{E}[X_{t,1}] \,,$$

and note that the left-hand side of (A.8) can be rewritten as

$$\max_{1 \leq j \leq n} \max_{1 \leq k \leq n-j} \frac{1}{\sqrt{k}} |S_{j+k} - S_j| = \max \Big\{ \max_{1 \leq j \leq n} \max_{1 \leq k \leq n-j} \frac{1}{\sqrt{k}} (S_{j+k} - S_j) \,, \max_{1 \leq j \leq n} \max_{1 \leq k \leq n-j} \frac{-1}{\sqrt{k}} (S_{j+k} - S_j) \Big\} \,.$$

Thus it suffices to show that both terms inside the (outer) maximum are of order $o_p(n^{1/4})$. For the sake of brevity, we will only prove that

$$\max_{1 \leq j \leq n} \max_{1 \leq k \leq n-j} \frac{1}{\sqrt{k}} (S_{j+k} - S_j) = o_p(n^{1/4}) \tag{A.9}$$

and the other term can be treated in the same way. Assertion (A.9) follows obviously from the two estimates

$$\max_{1\le j\le n}\ \max_{1\le k\le (n-j)\wedge\lfloor \log^2(n)\rfloor} \frac{S_{j+k}-S_j}{\sqrt{k}n^{1/4}} = o_p(1)\ , \tag{A.10}$$

$$\max_{1\le j\le n}\ \max_{\lfloor \log^2(n)\rfloor\le k\le n-j} \frac{S_{j+k}-S_j}{\sqrt{k}n^{1/4}} = o_p(1)\ . \tag{A.11}$$

Since the function g is bounded, one directly obtains that there exists a constant C such that $|X_{j,1}-\mathbb{E}[X_{j,1}]|\le C$. This gives

$$\Big|\max_{1\le j\le n}\ \max_{1\le k\le (n-j)\wedge\lfloor \log^2(n)\rfloor} \frac{S_{j+k}-S_j}{\sqrt{k}n^{1/4}}\Big| \le \max_{1\le j\le n}\ \max_{1\le k\le (n-j)\wedge\lfloor \log^2(n)\rfloor} \frac{\sqrt{k}C}{n^{1/4}} = o(1)$$

and so (A.10) is shown. To establish (A.11) we will use Corollary 1 from Wu and Zhou (2011), which implies, that (on a richer probability space) there exists a process $\{\hat{S}_i\}_{i=1}^n$ and independent Gaussian distributed random variables $Y_1,\dots,Y_n\sim\mathcal{N}(0,\Gamma(g))$ such that

$$(\hat{S}_1,\dots,\hat{S}_n)\overset{\mathcal{D}}{=}(S_1,\dots,S_n)$$

and the centered Gaussian process

$$\{\hat{G}_i\}_{i=1}^n = \Big\{\sum_{t=1}^i Y_t\Big\}_{i=1}^n$$

satisfies

$$\max_{1\le i\le n}|\hat{S}_i-\hat{G}_i| = \mathcal{O}_{\mathbb{P}}\big(n^{1/4}(\log(n))^{3/2}\big)\ .$$

Therefore we obtain

$$\begin{aligned}
&\left|\max_{1\le j\le n}\ \max_{\lfloor \log^2(n)\rfloor\le k\le n-j} \frac{\hat{S}_{j+k}-\hat{S}_j}{\sqrt{k}n^{1/4}} - \max_{1\le j\le n}\ \max_{\lfloor \log^2(n)\rfloor\le k\le n-j} \frac{\hat{G}_{j+k}-\hat{G}_j}{\sqrt{k}n^{1/4}}\right| \\
&\le \left|\max_{1\le j\le n}\ \max_{\lfloor \log^2(n)\rfloor\le k\le n-j} \frac{\hat{S}_{j+k}-\hat{S}_j-\hat{G}_{j+k}+\hat{G}_j}{\sqrt{k}n^{1/4}}\right| \le 2\max_{1\le j\le n}\frac{|\hat{S}_j-\hat{G}_j|}{\log^2(n)n^{1/4}} = o_p(1)\ .
\end{aligned}$$

Now Theorem 1 in Shao (1995) gives

$$\lim_{n\to\infty}\max_{1\le j\le n}\ \max_{\lfloor \log^2(n)\rfloor\le k\le n-j} \frac{\hat{G}_{j+k}-\hat{G}_j}{\sqrt{k}n^{1/4}} = 0$$

with probability 1, which completes the proof of Part (a).

For a proof of part (b) of Proposition 4.2 let $F^{(1)}$, $\Sigma_{F^{(1)}}$ and $F^{(2)}$, $\Sigma_{F^{(2)}}$ denote the distribution function and corresponding long-run variances in equation (3.18) before and after the change point, respectively. Note that $h = A \cdot g$ and consider the time series

$$\tilde{X}_t = \begin{cases} g(\varepsilon_t, \varepsilon_{t-1}, \dots) & \text{if } \ t < \lfloor mc \rfloor \\ A^{-1} \cdot h(\varepsilon_t, \varepsilon_{t-1}, \dots) & \text{if } \ t \geq \lfloor mc \rfloor \end{cases} ,$$

which is strictly stationary with distribution function $F^{(1)}$. Using similar arguments as in the proof of part (a), one easily verifies that

$$\Big\{ \frac{1}{\sqrt{m}} \sum_{t=1}^{\lfloor ms \rfloor} \mathcal{IF}_v(\tilde{X}_t, F^{(1)}, V) \Big\}_{s \in [0, T+1]} \overset{\mathcal{D}}{\Longrightarrow} \{ \sqrt{\Sigma_{F^{(1)}}} W(s) \}_{s \in [0, T+1]} \ . \tag{A.12}$$

Next, observe that there exists a matrix $A^{(v)} \in \mathbb{R}^{3 \times 3}$, such that for all symmetric matrices $M \in \mathbb{R}^{2 \times 2}$, the following identity holds

$$\operatorname{vech}(A \cdot M \cdot A^\top) = A^{(v)} \cdot \operatorname{vech}(M) \ .$$

Further, using (4.10) one observes

$$A \cdot \mathcal{IF}(\tilde{X}_t, F^{(1)}, V) \cdot A^\top = A \tilde{X}_t \tilde{X}_t^\top A^\top - A \cdot V(F^{(1)}) \cdot A^\top = \mathcal{IF}(X_t, F^{(2)}, V)$$

whenever $t \geq \lfloor mc \rfloor$, which yields

$$A^{(v)} \mathcal{IF}_v(\tilde{X}_t, F^{(1)}, V) = \mathcal{IF}_v(X_t, F^{(2)}, V) \quad \text{for} \quad t \geq \lfloor mc \rfloor \ . \tag{A.13}$$

Similar arguments give

$$A^{(v)} \Sigma_{F^{(1)}} (A^{(v)})^\top = \Sigma_{F^{(2)}} \ . \tag{A.14}$$

Now consider the mapping

$$\Phi_A : \begin{cases} \ell^\infty([0, T+1], \mathbb{R}^3) \to \ell^\infty([0, c], \mathbb{R}^3) \times \ell^\infty([c, T+1], \mathbb{R}^3) \\ \{f(s)\}_{s \in [0, T+1]} \mapsto \begin{pmatrix} \{f(s)\}_{s \in [0, c]} \\ \big\{ A^{(v)}(f(s) - f(c)) \big\}_{s \in [c, T+1]} \end{pmatrix} \end{cases} ,$$

then the Continuous Mapping, (A.12) and (A.13) yield

$$\begin{aligned} \begin{pmatrix} \big\{ \frac{1}{\sqrt{m}} \sum_{t=1}^{\lfloor ms \rfloor} \mathcal{IF}(X_t, F^{(1)}, V) \big\}_{s \in [0, c]} \\ \big\{ \frac{1}{\sqrt{m}} \sum_{t=\lfloor mc \rfloor + 1}^{\lfloor ms \rfloor} \mathcal{IF}(X_t, F^{(2)}, V) \big\}_{s \in [c, T+1]} \end{pmatrix} &\overset{\mathcal{D}}{\Longrightarrow} \begin{pmatrix} \big\{ \sqrt{\Sigma_{F^{(1)}}} W(s) \big\}_{s \in [0, c]} \\ \big\{ A^{(v)} \sqrt{\Sigma_{F^{(1)}}} \big(W(s) - W(c) \big) \big\}_{s \in [c, T+1]} \end{pmatrix} \\ &\overset{\mathcal{D}}{=} \begin{pmatrix} \big\{ \sqrt{\Sigma_{F^{(1)}}} W(s) \big\}_{s \in [0, c]} \\ \big\{ \sqrt{\Sigma_{F^{(2)}}} \big(W(s) - W(c) \big) \big\}_{s \in [c, T+1]} \end{pmatrix} , \end{aligned}$$

where the identity in distribution follows from the fact that both components are independent and the identity

$$\left(A^{(v)}\sqrt{\Sigma_{F^{(1)}}}\right)\left(A^{(v)}\sqrt{\Sigma_{F^{(1)}}}\right)^{\top} = A^{(v)}\Sigma_{F^{(1)}}\left(A^{(v)}\right)^{\top} = \Sigma_{F^{(2)}} \ .$$

For the verification of Assumption 3.6 it suffices to show that both, the phase before and after the change point satisfy Assumption 3.2. This can be done using similar arguments as in the proof of part (a) of Proposition 4.2 and the details are omitted. □