Sequential change point detection in high dimensional time series

Josua Gösmann, Christina Stoehr and Holger Dette

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

Abstract

Change point detection in high dimensional data has found considerable interest interest in recent years. Most of the literature designs methodology for a retrospective analysis, where the whole sample is already available when the statistical inference begins. This paper takes a different point of view and develops monitoring schemes for the online scenario, where high dimensional data arrives steadily and the goal is to detect changes as fast as possible controlling at the same time the probability of a type I error of a false alarm. We develop sequential procedures capable of detecting changes in the mean vector of a successively observed high dimensional time series with spatial and temporal dependence. The statistical properties of the methods are analyzed in the case where both, the sample size and dimension converge to infinity. In this scenario it is shown that the new monitoring schemes have asymptotic level alpha under the null hypothesis of no change and are consistent under the alternative of a change in at least one component of the high dimensional mean vector. Moreover, we also prove that the new detection scheme identifies all components affected by a change. The finite sample properties of the new methodology are illustrated by means of a simulation study and in the analysis of a data example.

Our approach is based on a new type of monitoring scheme for one-dimensional data which turns out to be often more powerful than the usually used CUSUM and Page-CUSUM methods, and the component-wise statistics are aggregated by the maximum statistic. From a mathematical point of view we use Gaussian approximations for high dimensional time series to prove our main results and derive extreme value convergence for the maximum of the maximal increment of dependent Brownian motions. In particular we show that the range of a Brownian motion on a given interval is in the domain of attraction of the Gumbel distribution.

Keywords and phrases: high dimensional time series, change point analysis, sequential monitoring, Gaussian approximation, bootstrap

AMS Subject classification: Primary 62M10, 62H15; Secondary 62G20; 60G70

1 Introduction

As digital transformation processes have accelerated during the last decades, new technologies like smartphones or car sensors are able to gather large amounts of data. Due to this development companies, states, research institutes etc. face the problem to manage, monitor and examine huge data sets, which regularly exceed the means of traditional tools. Thus the demand of so-called big data technology is steadily growing and thereby the requirement for theoretical foundation has put a lot of attention at the topic of high dimensional statistics in recent years.

Especially, the topic of change-point analysis or detection of structural breaks has regained attraction and numerous authors have started to embed commonly used multivariate methods into a high dimensional framework or even develop new methodology from scratch. Among many others, high dimensional change point problems have been considered by Cho and Fryzlewicz (2015), Wang and Samworth (2018), who develop methodology to identify multiple change points by a (wild) binary segmentation algorithm under sparsity assumptions. Jirak (2015) and Dette and Gösmann (2018) aggregate component-wise CUSUM-statistics by the maximum functional to detect structural breaks in a sequence of means of a high dimensional time series. Lévy-Leduc and Roueff (2009) analyze internet traffic data, by applying a compenent-wise CUSUM-test to dimension-reduced censored data. Enikeeva and Harchaoui (2019) employ the Euclidean norm of the CUSUM-process to obtain a linear and a scan statistic of χ^2-type, that is minimax-optimal under the regime of independent Gaussian observations. Change point problems in high dimensional covariance matrices are studied by Wang et al. (2017), Avanesov and Buzun (2018) and Dette et al. (2018) using (wild) binary segmentation, multiscale methods and U-statistics, respectively. U-statistics are also used by Wang et al. (2019) and Wang and Shao (2020) to develop testing and estimation methodology for a structural break in the mean.

All listed references on high dimensional change point problems have in common that the proposed methods are designed for a *retrospective analysis*, where the whole sample is already available when the statistical inference is commenced. In contrast to this, sequential change point detection deals with methods that are applicable for monitoring data in a so-called *online scenario*. In such a setup, data arrives steadily and methods are constructed to detect changes as fast as possible, while the problem is reevaluated with each new data point. The literature distinguishes between the *open-end* and *closed-end scenario*. A *closed-end scenario* is associated with a fixed endpoint, where monitoring has to be eventually stopped if no change was detected before. An *open-end scenario* does not postulate an endpoint meaning that monitoring can (theoretically) continue forever if no change is detected. For many applications an online scenario appears considerably more appropriate as operators do not have to wait until data collection is completed or the collection never stops, which can be the case for instance in meteorology or economics among many other fields.

The problem of sequential change point detection has received a lot of attention in the last century since considered early by the seminal paper of Page (1954). In the major part of the 20th century it was tackled by the use of so-called *control charts*, which have their focus on detecting a change as quickly as possible after it occurs [see for example Woodall and Montgomery (1999)]. Usually control charts do not offer the feature to control the type I error of a false alarm, i.e. deciding for a change although it is absent. Chu et al. (1996) propose a sequential paradigm where such a control is possible. It is based on the premise of an initial stable data set, which is unaffected by changes. By the help of invariance principles monitoring procedures can be derived, which have power under the alternative of a structural break and control the type I error (asymptotically, i.e. with increasing length of the stable data set). Since its first introduction, this paradigm has found considerable attention in the literature on change point detection. For example, Horváth et al. (2004), Hušková and Koubková (2005) and Aue et al. (2006) consider changes in the parameters of linear models with statistics based on residuals. For independent identically distributed (i.i.d.) data, and Kirch (2008) and Hušková and Kirch (2012) propose several bootstrap procedures for sequential change point detection in the mean and in the parameters of a linear regression models. A MOSUM-approach, which employs a moving monitoring window in linear models was introduced by Chen and Tian (2010), while Ciuperca (2013) proposes a generalization of the *sequential* CUSUM statistic to non-linear models. Fremdt (2014b) uses the so-called Page-CUSUM, which scans for changes through the already available monitoring data and is more efficient to detect later changes than the classical sequential CUSUM scheme. Hoga (2017) proposes an ℓ_1-norm to detect structural breaks in the mean and variance of a multivariate time series and Dette and Gösmann (2019) develop an amplified scanning method combined with self-normalization. Otto and Breitung (2019) define a Backward CUSUM statistic based on recursive residuals in a linear model. Unifying frameworks are provided in Kirch and Kamgaing (2015) and Kirch and Weber (2018) and a theory based on U-statistic is established in Kirch and Stoehr (2019). We also refer to the recent review of sequential procedures in Section 1 of Anatolyev and Kosenok (2018).

This list is by no means complete, but - to the authors best knowledge - a common feature of the literature is that it does not consider sequential change point detection in the high dimensional scenario. The purpose of the present paper is to address this problem in the context of detecting changes in the mean, where the dimension is allowed to grow with the sample size at any polynomial order. For this purpose we will consider the paradigm of Chu et al. (1996) and develop sequential algorithms in the high dimensional regime. Our approach is based on aggregating component-wise sequential detection schemes by the maximum statistic. For the individual components we use a novel monitoring procedure, which screens for all possible positions of the change point and takes into account that the change does not necessarily occur in the first observations after the initial sample - see Section 2 for more details. A nice feature of this approach consists in the fact that the limiting distribution of the statistic used to monitor each component (after appropriate standardization) is given by the *range of the Brownian motion*, that is

$$\mathbb{M} = \max_{0 \le t \le q} W(t) - \min_{0 \le t \le q} W(t)$$

where W is a Brownian motion on the interval $[0, q]$ and $0 < q < 1$ a known constant. The distribution of the random variable $\mathbb{M}$ appears as the weak limit of the range of cumulative sums of i.i.d. random variables with variance 1 [see Feller (1951)], and we will show that it belongs to the domain of attraction of the Gumbel distribution. This result is of independent interest in extreme value theory and allows us to aggregate component-wise statistics by the maximum and develop a sequential monitoring scheme using the quantiles of the Gumbel distribution. Thus our approach differs substantially from the methods proposed by Soh and Chandrasekaran (2017), who consider high dimensional sparse signals, and Chu and Chen (2018), who develop an algorithm based on nearest neighbor information.

As the rates of most convergence results in extreme value theory are known to be rather slow, we also propose a simple bootstrap procedure, which improves the finite sample performance of the sequential monitoring scheme substantially. We would like to point out that the development of resampling procedures in the sequential regime is a difficult problem. On the one hand critical values can be computed only from the initial stable sample, but this set can be too small to obtain reliable values. On the other hand one can compute new critical values with each new data point, which is computationally expensive and can be corrupted by an undetected structural break. Therefore both approaches have natural advantages and disadvantages. For i.i.d. data Kirch (2008) proposes a bootstrap procedure for sequential detection of a structural break in the mean of a one-dimensional sequence by a combination of both methods following ideas of Steland (2006). However, the construction of bootstrap methodology for sequential change point detection in the high dimensional regime remains challenging.

The remaining part of this paper is organized as follows. In Section 2 we introduce the specific testing problem under consideration and present the new monitoring procedure for structural breaks in the sequence of means from a high dimensional time series. Section 3 is devoted to our main results and to the analysis of the asymptotic properties of the new procedure. In particular we prove that the maximum of the individual test statistics converges weakly (with increasing dimension and initial sample size) to a Gumbel distribution. These results are used to show that the monitoring scheme has asymptotic level α and is consistent.

For this purpose we combine recently developed Gaussian approximations tools for high dimensional statistics [see Chernozhukov et al. (2013) and Zhang and Cheng (2018)] with classical extreme value theory. We are able to control the distance between our statistic and a counterpart computed from Gaussian random variables, which share the long-run correlation structure of the observed time series. To analyze the Gaussian statistic, we employ Gaussian comparison and anti-concentration inequalities to show that this statistic is sufficiently close to the maximum of ranges of dependent Brownian motions. Finally, we prove that the range of a Brownian motion on a given interval is in the domain of attraction of the Gumbel distribution in the case of independence and use a Poisson approximation to eliminate the independence condition.

The asymptotic theory, in particular the Gaussian approximation used in the proofs, provides a clear way how to develop a bootstrap procedure to detect structural breaks under a controlled type I error, for which we also prove its consistency. In Section 4 we investigate the finite sample properties of the new procedures by means of a simulation study and illustrate potential applications in a data example. Finally, all proofs which are techically demanding are deferred to an appendix in Section A.

2 Sequential monitoring of high dimensional time series

Let $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ denote a time series of random vectors in $\mathbb{R}^d$ with mean vectors

$$\boldsymbol{\mu}_t := (\mu_{t,1},\ldots,\mu_{t,d})^\top = \mathbb{E}[\boldsymbol{X}_t] := \mathbb{E}[(X_{t,1},\ldots,X_{t,d})^\top] \ .$$

We take the sequential point of view and are interested to monitor for changes of the vectors $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots$. Following Chu et al. (1996) we assume that a historic or initial data set, say $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$, is available, which is known to be mean stable. Starting with observation $\boldsymbol{X}_{m+1}$ we will sequentially test for a change of the mean vector in the monitoring period. The corresponding testing problem is therefore given by the hypotheses

$$\begin{aligned} &H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_m = \boldsymbol{\mu}_{m+1} = \boldsymbol{\mu}_{m+2} = \cdots \\ \text{versus} \quad &H_1 : \exists k^* \in \mathbb{N}\ , s.t.\ \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_m = \cdots = \boldsymbol{\mu}_{m+k^*-1} \neq \boldsymbol{\mu}_{m+k^*} = \cdots \ . \end{aligned} \tag{2.1}$$

In the present paper we consider a closed-end scenario where the procedure stops after $m+Tm$ observations even if no change has been detected [see Aue et al. (2012), Wied and Galeano (2013) among many others]. The factor T determines the length of the monitoring period compared to the size of the initial training set m and so the hypotheses in (2.1) read as follows

$$\begin{aligned} &H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_m = \boldsymbol{\mu}_{m+1} = \boldsymbol{\mu}_{m+2} = \cdots = \boldsymbol{\mu}_{m+Tm} \\ \text{versus} \quad &H_1 : \exists k^* \in \{1,\ldots,Tm\}, \text{ such that} \\ &\qquad \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_{m+k^*-1} \neq \boldsymbol{\mu}_{m+k^*} = \cdots = \boldsymbol{\mu}_{m+Tm} \ . \end{aligned} \tag{2.2}$$

In the following we will develop a sequential detection scheme which is capable to distinguish between the hypotheses given in (2.2) in a high dimensional setting where the dimension d of the mean vector is increasing with the initial sample size m. To be precise we denote by

$$\widehat{\mu}_i^j(h) = \frac{1}{j-i+1}\sum_{t=i}^{j} X_{t,h} \tag{2.3}$$

the estimator of the mean in component $h \in \{1,\ldots,d\}$ from the sample $X_{i,h},\ldots,X_{j,h}$. Following Gösmann et al. (2019) we consider the statistic

$$\hat{E}_{m,h}(k) = \max_{j=0}^{k-1} \frac{k-j}{\sqrt{m}\hat{\sigma}_h}\left|\widehat{\mu}_{m+j+1}^{m+k}(h) - \widehat{\mu}_1^{m+j}(h)\right| \tag{2.4}$$

at time point $m+k$ in a single component h, where $\hat{\sigma}_h^2$ denotes an appropriate estimator of the unknown long-run variance

$$\sigma_h^2 = \sum_{t\in\mathbb{Z}} \mathrm{Cov}(X_{0,h}, X_{t,h})$$

in the hth component (explicit conditions for the existence of the long-run variance are given in Section 3). Note that $\hat{E}_{m,h}(k)$ is a weighted CUSUM statistic to detect a change point in the sequence of means corresponding to the data $X_{m+1,h}, \ldots, X_{m+k,h}$. A structural break in the sequence of means $\mu_{m+1,h}, \mu_{m+2,h}, \ldots$ is detected as soon as the sequence

$$w(1/m)\hat{E}_{m,h}(1), w(2/m)\hat{E}_{m,h}(2), \ldots$$

exceeds a given threshold, that is

$$w(k/m)\hat{E}_{m,h}(k) > c_{\alpha,h} \ , \tag{2.5}$$

where w is a suitable weight function and the critical value $c_{\alpha,h}$ is chosen based on the desired test level α. Following Aue and Horváth (2004), Wied and Galeano (2013) and Fremdt (2014b) we will work with the commonly used weight function

$$w(t) = 1/(1+t) \ , \tag{2.6}$$

throughout this paper.

Remark 2.1

(1) Note that most of the literature investigates sequential detectors based on the differences

$$\left|\widehat{\mu}_{m+1}^{m+k}(h) - \widehat{\mu}_1^{m}(h)\right| \tag{2.7}$$

and the corresponding detection schemes are usually called (ordinary) CUSUM tests [see Chu et al. (1996), Horváth et al. (2004), Aue et al. (2006)]. Another part of the literature focuses on detectors based on the differences

$$\left|\widehat{\mu}_{m+j+1}^{m+k}(h) - \widehat{\mu}_1^{m}(h)\right| \ \text{ for } \ j = 0, \ldots, k-1 \tag{2.8}$$

and the corresponding detection schemes are usually called Page-CUSUM tests [see Fremdt (2014a,b), Kirch and Weber (2018)]. The use of the differences $\left|\widehat{\mu}_{m+j+1}^{m+k}(h) - \widehat{\mu}_1^{m+j}(h)\right|$ is motivated by the likelihood principle [see Dette and Gösmann (2019)]. Compared to the differences in (2.7) it avoids the problem that the estimator $\widehat{\mu}_{m+1}^{m+k}(h)$ may be corrupted by observations before the change point, which could lead to a loss of power. Compared to the differences in (2.8) the use of $\widehat{\mu}_1^{m+j}(h)$ instead of $\widehat{\mu}_1^{m}(h)$ may avoid a loss in power in cases with a small initial sample and a rather late change point. The advantages of detection schemes based on the differences $\left|\widehat{\mu}_{m+j+1}^{m+k}(h) - \widehat{\mu}_1^{m+j}(h)\right|$ against ordinary sequential CUSUM and the Page-CUSUM procedures have been recently demonstrated by Gösmann et al. (2019).

(2) Several authors consider the more general class of weight functions

$$w_\gamma(t) = (t+1)^{-1}\Big(\frac{t}{t+1}\Big)^{-\gamma}$$

for $\gamma \in [0, 1/2)$ [see for instance Horváth et al. (2004), Aue et al. (2006) or Kirch and Weber (2018)]. The weight function in (2.6) is obtained for $\gamma = 0$ and has proven to be preferable to $\gamma > 0$ in many situations except for changes that occur almost immediately [see Kirch and Weber (2018)]. It is most likely that the theoretical results of this paper remain correct in the case $\gamma > 0$.

In order to control the probability of erroneously deciding for a structural break in the component h during the monitoring period one has to determine the probability

$$\mathbb{P}\Big(\max_{k=1}^{Tm} w(k/m)\hat{E}_{m,h}(k) > c_{\alpha,h} \Big) \ .$$

For fixed $h \in \{1, \dots, d\}$ we can use a result of Gösmann et al. (2019) who showed (that under appropriate assumptions)

$$\max_{k=1}^{Tm} w(k/m)\hat{E}_{m,h}(k) \overset{\mathcal{D}}{\Longrightarrow} \mathbb{M} = \max_{0 \le t \le q(T)} W(t) - \min_{0 \le t \le q(T)} W(t) \tag{2.9}$$

where the symbol $\overset{\mathcal{D}}{\Longrightarrow}$ denotes weak convergence, $q(T) = T/(T+1)$ and W is a standard one-dimensional Brownian motion. $\mathbb{M}$ is known in the probability literature as the range of the Brownian motion on the interval $[0, q(T)]$ and its distribution appears as the weak limit of the range of cumulative sums of i.i.d. random variables with variance 1 [see Feller (1951)].

For a detection of a change point in the complete mean vector we propose to aggregate the statistics for the different spatial dimensions $h = 1, \dots, d$. More precisely, we consider the maximum of the different components, that is

$$\max_{h=1}^{d} w(k/m)\hat{E}_{m,h}(k) \tag{2.10}$$

and reject the (closed-end) null hypothesis

$$H_0 : \ \boldsymbol{\mu}_1 = \dots = \boldsymbol{\mu}_m = \boldsymbol{\mu}_{m+1} = \dots = \boldsymbol{\mu}_{m+Tm}$$

of no structural break in the high dimensional means $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_{m+Tm}$ if this quantity exceeds a given threshold, that is

$$\max_{k=1}^{Tm} \max_{h=1}^{d} w(k/m)\hat{E}_m(k) > q \ . \tag{2.11}$$

Here the critical value q is chosen appropriately such that (asymptotically) the probability of erroneously deciding for a change point is controlled. In the following section we investigate the weak convergence of the statistic $\max_{k=1}^{Tm} \max_{h=1}^{d} w(k/m)\hat{E}_{m,h}(k)$. These results will be used to define critical values $q_{d,m,\alpha}$ in (2.11) (one by asymptotic theory and one by bootstrap), such that the monitoring procedure is consistent and at the same time controls the probability of the type I error, that is

$$\limsup_{m,d \to \infty} \mathbb{P}_{H_0}\Big(\max_{k=1}^{Tm} \max_{h=1}^{d} w(k/m)\hat{E}_{m,h}(k) > q_{d,m,\alpha} \Big) \le \alpha \ , \tag{2.12}$$

$$\lim_{m,d \to \infty} \mathbb{P}_{H_1}\Big(\max_{k=1}^{Tm} \max_{h=1}^{d} w(k/m)\hat{E}_{m,h}(k) > q_{d,m,\alpha} \Big) = 1 \ . \tag{2.13}$$

3 Main results

In this section, we derive the asymptotic properties of the proposed detector defined in (2.10) in the high dimensional setting where sample size and dimension converge to infinity and we allow for temporal as well as spatial dependencies in the data. In particular we establish in Theorem 3.6 below under the null hypothesis of no change in the mean vector the weak convergence

$$a_d\Big(\max_{k=1}^{Tm}\max_{h=1}^{d} w(k/m)\hat{E}_{m,h}(k) - b_d\Big) \overset{\mathcal{D}}{\Longrightarrow} G \quad \text{as } m,d\to\infty, \tag{3.1}$$

where a_d, b_d are suitable sequences and G is a standard Gumbel random variable with c.d.f. $F_G(x) = \exp\big(-\exp(-x)\big)$. As inevitable in high dimensional time series analysis we require assumptions on the relation between the (initial) sample size and the dimension as well as assumptions on the dependence structure to control the dependence between components at different time points uniformly.

Throughout this paper we assume the location model

$$X_{t,h} = \mu_{t,h} + e_{t,h}\ , \quad h = 1,\dots,d;\ t = 1,\dots,m+Tm\ , \tag{3.2}$$

where $\mu_{t,h} = \mathbb{E}[X_{t,h}]$ is the expectation of the h component and the centered sequence $\{e_{t,h}\}$ is given as a physical system [see e.g. Wu (2005)], that is

$$e_{t,h} = g_h(\varepsilon_t, \varepsilon_{t-1},\dots)\ . \tag{3.3}$$

The underlying innovation sequence $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ consists of i.i.d. random variables with values in some arbitrary measure space $\mathbb{S}$ and the functions $g_h : \mathbb{S}^{\mathbb{N}} :\to \mathbb{R}$ are assumed to be measurable. Note that by the definition above the random variables $\{e_{t,h}\}_{t\in\mathbb{Z},h\in\mathbb{N}}$ are (strictly) stationary with respect to the time index t under the null hypothesis such that for any fixed dimension d the multivariate time series

$$\{\boldsymbol{e}_t\}_{t\in\mathbb{Z}} = \{(e_{t,1}, e_{t,2}, e_{t,3},\dots,e_{t,d})^\top\}_{t\in\mathbb{Z}}$$

is stationary. The data generating model defined by formula (3.3) has received a lot of attention in recent years [see for example Wu and Zhou (2011), Liu et al. (2013), El Machkouri et al. (2013), Berkes et al. (2014) among many others]. It covers the major part of prevalent time series models like autoregressive or moving average processes. Furthermore, it also allows for a natural measurement of temporal dependence which is constructed as follows. Let ε_0' be an independent copy of ε_0 and define

$$X_{t,h}' = \mu_{t,h} + g_h(\varepsilon_t, \varepsilon_{t-1},\dots\varepsilon_1, \varepsilon_0', \varepsilon_{-1},\dots)$$

as a counterpart of $X_{t,h}$ where ε_0 is replaced by ε_0'. If $p \geq 1$ we denote by $\|X\|_p = \mathbb{E}[|X|^p]^{1/p}$ the ordinary L_p-norm of a real-valued random variable X (assuming its existence). If $\|e_{t,h}\|_p < \infty$ the coefficients

$$\vartheta_{t,h,p} := \big\|X_{t,h} - X_{t,h}'\big\|_p$$

measure the influence of innovation ε_0 on $X_{t,h}$ and thereby quantify the (temporal) dependence within the system $\{X_{t,h}\}_{t\in\mathbb{Z},h\in\mathbb{N}}$ defined by (3.3). If $\|e_{t,h}\|_p < \infty$ for some $p \geq 2$ we define the cross-components covariances by

$$\phi_{t,h_1,h_2} := \mathrm{Cov}(X_{0,h_1}, X_{t,h_2}) \ , \quad \phi_{t,h} := \phi_{t,h,h}$$

and the long-run covariances and variances by

$$\gamma_{h_1,h_2} := \sum_{t\in\mathbb{Z}} \phi_{t,h_1,h_2} \quad \text{and} \quad \sigma_h^2 := \gamma_{h,h} \ , \tag{3.4}$$

respectively. If $\sigma_{h_1}, \sigma_{h_2} > 0$ let additionally

$$\rho_{h_1,h_2} := \frac{\gamma_{h_1,h_2}}{\sigma_{h_1}, \sigma_{h_2}} \tag{3.5}$$

denote the long-run correlations. It will be crucial for the asymptotic considerations to control the coefficients $\vartheta_{t,h,p}$ and the correlations ρ_{h_1,h_2} for increasing t and $|h_1 - h_2|$, respectively. This will be formulated in Assumptions 3.3 and 3.4 below. Before we state these precisely we begin with two assumptions on the relation between sample size and dimension and on the tail behaviour of the errors in model (3.2).

Assumption 3.1 (Assumption on the dimension) There exists constants $D > 0$ and $C_D > 0$, such that

(D1) $d = C_D \cdot m^D$.

Assumption 3.2 (Structural assumptions) Assume that exponential moments exist and are bounded, that is: there exists a constant $B < 3/8$, such that

(S1) $\max_{h=1}^{d} \mathbb{E}\big[\exp\big(|e_{1,h}|\big)/B_m\big] \leq C_e$, where $B_m \leq C'_e \cdot m^B$,

where $C_e > 1$ and $C'_e > 0$ are constants.

Assumption 3.3 (Temporal dependence) There exist constants $p > 2D+4$, $\beta \in [0,1)$, $C_\vartheta > 0$ such that

(TD1) $\max_{h=1}^{d} \vartheta_{t,h,p} \leq C_\vartheta \beta^t$.

Further assume that for two positive constants c_σ and C_σ the long-run variances defined in (3.4) are uniformly bounded, that is

(TD2) $c_\sigma \leq \max_{h=1}^{d} \sigma_h \leq C_\sigma$.

Assumption 3.4 (Spatial dependence) Assume that there exists a sequence L_d such that

(SD1) $L_d = o(\sqrt{\log d})$.

and the long-run correlations defined in (3.5) fulfill

(SD2) $\sup_{i,j:\,|i-j|>L_d} |\rho_{i,j}| = o\big(\log^{-2}(d)\big)$,

(SD3) $\sup_{i,j:\,|i-j|\geq 1} |\rho_{i,j}| \leq \rho_+ < 1$.

Note that assumptions of the type (D1) are quite common in high dimensional change point problems. For example, Jirak (2015), Wang and Samworth (2018) and Dette and Gösmann (2018) also assume a polynomial growth of the dimension with the sample size. Conditions like Assumption 3.2 are indispensable ingredients for Gaussian approximation results in high dimensional statistics [see Chernozhukov et al. (2013), Zhang and Cheng (2018) or Chernozhukov et al. (2019)] and will be used in the proofs of our main results. Assumption (TD1) controls the component-wise temporal dependence and guarantees, that the quantities defined in (3.4) exist. Assumption 3.4 controls the long-run correlations between different components.

3.1 Monitoring using asymptotic quantiles

In this section we develop the asymptotic theory to define a quantile in the monitoring scheme (2.11). Our first result provides the basis for the proof of the main Theorem 3.6 of this section. It is stated here, because it is of independent interest and shows that the distribution of the random variable $\mathbb{M}$ in (2.9) is in the domain of attraction of the Gumbel distribution. Throughout this paper Φ denotes the cumulative distribution function of a standard normal distribution.

Theorem 3.5 *Let $(W_1, \ldots, W_d)^\top$ be a d-dimensional Brownian motion with correlation matrix $\Sigma_d = (\rho_{i,j})_{1\leq i,j\leq d}$ such that $\rho_{i,j} = 0$ if $|i-j| > L_d$, and assume that (SD1) and (SD3) of Assumption 3.4 are satisfied. Further denote for $h = 1, \ldots, d$ by*

$$M_h = \max_{0\leq t\leq q(T)} W_h(t) - \min_{0\leq t\leq q(T)} W_h(t) \tag{3.6}$$

the range of the Brownian motion W_h in the interval $[0, q(T)]$. Then we obtain for $d \to \infty$

$$a_d\Big(\max_{h=1}^{d} M_h - b_d\Big) \overset{\mathcal{D}}{\Longrightarrow} G \ ,$$

where G denotes the Gumbel distribution with cumulative distribution function $F_G(x) = \exp(-\exp(-x))$. The scaling sequences a_d, b_d are implicitly given by

$$b_d = U_{\mathbb{M}}(d) \ , \quad a_d = \frac{F'_{\mathbb{M}}(U_{\mathbb{M}}(d))}{1 - F_{\mathbb{M}}(U_{\mathbb{M}}(d))} \ , \tag{3.7}$$

where $U_{\mathbb{M}}$ is the inverse function of $1/(1-F_{\mathbb{M}})$ and $F_{\mathbb{M}}$ is the distribution function of the random variable $\mathbb{M}$ defined in (2.9), which is given by

$$F_{\mathbb{M}}(x) := \mathbb{P}\big(\mathbb{M} \leq x\big) = 1 + 8\sum_{k=1}^{\infty}(-1)^k k\Phi\Big(-\frac{kx}{\sqrt{q(T)}}\Big) \tag{3.8}$$

for $x > 0$ and $F_{\mathbb{M}}(x) = 0$ for $x < 0$.

The explicit representation (3.8) of the distribution function of $F_{\mathbb{M}}$ can be found in Borodin and Salminen (1996) and is crucial in proof of Theorem 3.5. In the Appendix it will be shown that

$$\lim_{d\to\infty} \frac{b_d}{\sqrt{\log(d)}} = \sqrt{2q(T)} \ , \qquad \lim_{d\to\infty} \frac{a_d}{\sqrt{\log(d)}} = \sqrt{\frac{2}{q(T)}}.$$

In the proof of Theorem 3.6 below, we will use a Gaussian approximation which leads to the maximum of the increments M_h defined in (3.6) such that Theorem 3.5 can be applied. As indicated by (2.12), the limit distribution of the statistic

$$\widehat{\mathcal{T}}_{m,d} = \max_{k=1}^{Tm} \max_{h=1}^{d} w(k/m) \hat{E}_{m,h}(k) \tag{3.9}$$

has to be derived for the case $m, d \to \infty$ in order to determine an appropriate asymptotic critical value. For this purpose recall the definition of $\hat{E}_{m,h}(k)$ in (2.4) and define by

$$\begin{aligned} \mathcal{T}_{m,d} &= \max_{h=1}^{d} \max_{k=1}^{Tm} w(k/m) \frac{\hat{\sigma}_h}{\sigma_h} \hat{E}_{m,h}(k) \\ &= \max_{h=1}^{d} \max_{k=1}^{Tm} w(k/m) \max_{j=0}^{k-1} \frac{k-j}{\sqrt{m}\sigma_h} \Big| \widehat{\mu}_{m+j+1}^{m+k}(h) - \widehat{\mu}_1^{m+j}(h) \Big| \end{aligned} \tag{3.10}$$

a version of the statistic $\widehat{\mathcal{T}}_{m,d}$, where all component-wise long-run variance estimators $\hat{\sigma}_h$ have been replaced by the (unknown) true long-run variances σ_h. The following Theorem yields the asymptotic distribution of $\mathcal{T}_{m,d}$ as $m, d \to \infty$.

Theorem 3.6 *Suppose that the null hypothesis H_0 defined in (2.2) holds. Under the Assumptions 3.1 - 3.4 it follows that*

$$a_d \big(\mathcal{T}_{m,d} - b_d \big) \stackrel{\mathcal{D}}{\Longrightarrow} G$$

as $m, d \to \infty$, where G denotes a standard Gumbel random variable and the sequences a_d, b_d are defined in (3.7).

Given Theorem 3.6 our final task is to identify suitable long-run variance estimators to obtain the asymptotic distribution of $\widehat{\mathcal{T}}_{m,d}$. We will identify a general condition on the estimators in Assumption 3.7, which guarantees that all true long-run variances $\{\sigma_h^2\}_{h=1,\dots,d}$ in the statistic $\mathcal{T}_{m,d}$ can be replaced by their corresponding estimators. Explicit estimators satisfying this assumption are constructed in Remark 3.8.

Assumption 3.7 Suppose that there exists a long-run variance estimator $\hat{\sigma}_h = \hat{\sigma}_h(m)$ based only on the stable initial set, such that

$$\mathbb{P}\Big(\max_{h=1}^{d} |\hat{\sigma}_h - \sigma_h| \geq m^{-\delta_\sigma} \Big) \lesssim m^{-C_\sigma} \ ,$$

where $C_\sigma > 0$ and $\delta_\sigma > 0$ are sufficiently small constants.

Remark 3.8 In the field of sequential change point detection it is common to use only the initial stable data for the estimation of the long-run variance as this ensures that the estimate cannot be corrupted by a change [see for instance Aue et al. (2012), Wied and Galeano (2013) or Fremdt (2014b) among many others]. It follows from Jirak (2015) that Assumption 3.7 holds for the standard long-run variance estimators

$$\hat{\sigma}^2_{h,strd} = \hat{\phi}_{0,h} + 2\sum_{t=1}^{B_m} \hat{\phi}_{t,h} \ (h = 1, \dots, d) \ , \tag{3.11}$$

where the bandwidth parameter B_m grows polynomially in m and $\hat{\phi}_{t,h}$ denotes the lag t autocovariance estimator in component h, that is

$$\hat{\phi}_{t,h} := \frac{1}{m-t} \sum_{i=t+1}^{m} \big(X_{i,h} - \widehat{\mu}_1^m(h)\big)\big(X_{i-t,h} - \widehat{\mu}_1^m(h)\big) \ . \tag{3.12}$$

Corollary 3.9 *If the null hypothesis H_0 defined in (2.2) holds and the Assumptions 3.1 - 3.4 and 3.7 are satisfied, it follows that*

$$a_d\big(\widehat{\mathcal{T}}_{m,d} - b_d\big) \overset{\mathcal{D}}{\Longrightarrow} G \ .$$

If $g_{1-\alpha}$ is the $(1-\alpha)$ quantile of the Gumbel distribution, we obtain from Corollary 3.9 that the sequential procedure defined by (2.11) with $q = c_{d,\alpha} := g_{1-\alpha}/a_d + b_d$ has asymptotic size α, i.e.

$$\lim_{m,d\to\infty} \mathbb{P}_{H_0}\Big(\max_{h=1}^{d} w(k/m)\hat{E}_{m,h}(k) > c_{d,\alpha} \quad \text{for some} \ \ k \in \{1, \dots, Tm\} \Big) = \lim_{m,d\to\infty} \mathbb{P}_{H_0}\Big(\widehat{\mathcal{T}}_{m,d} > \frac{g_{1-\alpha}}{a_d} + b_d\Big) = \alpha \ .$$

The next theorem yields consistency of this monitoring scheme under the alternative hypothesis of a change in the mean vector.

Theorem 3.10 *Under the alternative hypothesis H_A defined in (2.2) assume that there is a component h^* and a time point $k^* = k^*(m)$ such that*

$$\sqrt{\frac{m}{\log m}} \cdot \big|\mu_{m+k^*-1,h^*} - \mu_{m+k^*,h^*}\big| \to \infty \quad \text{and} \quad \limsup_{m\to\infty} \frac{k^*}{m} < T \ . \tag{3.13}$$

If Assumptions 3.1 - 3.4 and 3.7 are satisfied, it follows that

$$\lim_{m,d\to\infty} \mathbb{P}_{H_1}\Big(\widehat{\mathcal{T}}_{m,d} > \frac{g_{1-\alpha}}{a_d} + b_d\Big) = 1 \ .$$

Condition (3.13) shows that the test is able to detect alternatives which converge to the null at the rate of $m^{-1/2}$ up to a factor $c_m\sqrt{\log m}$ with a sequence $(c_m)_{m\in\mathbb{N}}$ tending to ∞ at an arbitrary slow rate. This factor is needed to address for the high dimensional setting. Note also the time $m + k^*$ of the change is not permitted to be close to the end $m + mT$, which reflects the necessity to have a reasonable large sample after the change point such that the corresponding means can be estimated with sufficient precision.

3.2 Bootstrap quantiles

Theorem 3.6 and 3.10 show that the new sequential testing procedure (2.11) with $q = g_{1-\alpha}/a_d + b_d$ has asymptotic level α and is consistent. However, the approach so far is based on an approximation of the distribution of the statistic by a Gumbel distribution featuring the well-known disadvantage that the convergence rates in such limiting results are rather slow. As a consequence these quantiles may yield to imprecise approximations in practical applications. To tackle this problem, we will propose a resampling procedure, which is motivated by an essential intermediate step in the proof of Theorem 3.6. More precisely, by Lemma A.3 and A.4 from the Appendix we obtain the approximation

$$\mathbb{P}\Big(a_d\big(\mathcal{T}_{m,d} - b_d\big) \leq x\Big) - \mathbb{P}\Big(a_d\big(\mathcal{T}^{(Z)}_{m,d} - b_d\big) \leq x\Big) = o(1)\ . \tag{3.14}$$

Here the statistic $\mathcal{T}^{(Z)}_{m,d}$ is the counterpart of $\mathcal{T}_{m,d}$ computed from standard Gaussian random variables $Z_{t,h}$, which are independent in time and have spatial dependence structure

$$\text{Cov}\left(Z_{0,h}, Z_{0,i}\right) = \rho_{h,i}\ , \tag{3.15}$$

where $\rho_{h,i}$ are the long-run correlations defined in (3.5). By the approximation (3.14) it is therefore reasonable to obtain the quantiles for the statistic $\mathcal{T}_{m,d}$ from those of the statistic $\mathcal{T}^{(Z)}_{m,d}$, which can easily be simulated if the the correlation structure in (3.15) would be known. These parameters can be straightforwardly estimated from the initial stable data set $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$. Compared to a bootstrap procedure continuously performed during monitoring, this idea exhibits two important advantages. Firstly, it ensures that the correlation estimates cannot be corrupted by a mean change, that may occur during the monitoring period. Secondly, it requires less computational effort, as the quantile is only computed once before monitoring is commenced. This is of vital importance in a high dimensional setup, where the method on its own is already quite expensive and resampling and/or repeated estimation during monitoring may quickly exceed the computational resources.

Before discussing the technical details of this resampling procedure, we state a necessary assumption regarding the precision of the estimates of the long-run covariances.

Assumption 3.11 Suppose that there exists a long-run covariance estimator $\hat{\gamma}_{h,i} = \hat{\gamma}_{h,i}(m)$ based on the stable initial set $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$, such that

$$\mathbb{P}\left(\max_{h,i=1}^{d} \big|\hat{\gamma}_{h,i} - \gamma_{h,i}\big| \geq m^{-\delta_\gamma}\right) \lesssim m^{-C_\gamma}\ , \tag{3.16}$$

where $C_\gamma > 0$ and $\delta_\gamma > 0$ are sufficiently small constants.

Remark 3.12 A canonical choice for a long-run covariance estimator that is capable to fulfill (3.16), is the standard estimator

$$\hat{\gamma}_{h,i,strd} = \hat{\phi}_{0,h,i} + \sum_{t=1}^{B_m} \hat{\phi}_{t,h,i} + \sum_{t=1}^{B_m} \hat{\phi}_{t,i,h}\ , \tag{3.17}$$

where B_m is an appropriate bandwidth and the involved cross-components covariance estimators are given by

$$\hat{\phi}_{t,h,i} := \frac{1}{m-t} \sum_{j=t+1}^{m} \big(X_{j,h} - \widehat{\mu}_1^m(h)\big)\big(X_{j-t,i} - \widehat{\mu}_1^m(i)\big) \ . \tag{3.18}$$

Note that these definitions are natural extensions of the long-run variance and auto-covariance estimators in (3.11) and (3.12), respectively. Therefore one can use similar arguments as given in Jirak (2015) (for the verification of Assumption 3.7) to prove the consistency stated in (3.16).

In the following denote by $\mathcal{X} = \sigma(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m)$ the σ-algebra generated by the initial sample and let $\mathbb{P}_{|\mathcal{X}}, \mathrm{Cov}_{|\mathcal{X}}$ denote the conditional probability and covariance with respect to $\mathcal{X}$. To define the bootstrap statistic we use centered Gaussian random variables $\{\widehat{Z}_{t,h}\}_{t=1,\ldots,m+mT}^{h=1,\ldots d}$ with (conditional) covariance structure

$$\mathrm{Cov}_{|\mathcal{X}}\left(\widehat{Z}_{t,h},\, \widehat{Z}_{s,i}\right) = \hat{\rho}_{h,i} I\{t=s\} \ , \tag{3.19}$$

where $\hat{\rho}_{h,i}$ are canonically defined by

$$\hat{\rho}_{h,i} = \frac{\hat{\gamma}_{h,i}}{\hat{\sigma}_h \hat{\sigma}_i} \ .$$

Note that by the definition in (3.19) the random variables are independent in time but preserve the (estimated) spatial correlation structure of the time series. Denote the component-wise mean estimators for subsamples of $\{\widehat{Z}_{t,h}\}_{t=i}^{j}$ by

$$\hat{z}_i^j(h) = \frac{1}{j-i+1} \sum_{t=i}^{j} \widehat{Z}_{t,h} \tag{3.20}$$

and the final bootstrap statistic by

$$\widehat{\mathcal{T}}_{m,d}^{(Z)} = \max_{k=1}^{Tm} \max_{h=1}^{d} \max_{j=0}^{k-1} \frac{w(k/m)(k-j)}{\sqrt{m}} \left| \hat{z}_{m+j+1}^{m+k}(h) - \hat{z}_1^{m+j+1}(h) \right| \ . \tag{3.21}$$

Once the correlation estimates $\hat{\rho}_{h,i}$ are computed, the conditional distribution of $\widehat{\mathcal{T}}_{m,d}^{(Z)}$ with respect to the σ-algebra $\mathcal{X}$ can be approximated by Monte-Carlo simulation with arbitrary precision, generating replicates of $\{\widehat{Z}_{t,h}\}_{t=1,\ldots,m+mT}^{h=1,\ldots d}$. Thus provided with a batch of realizations of the statistic $\widehat{\mathcal{T}}_{m,d}^{(Z)}$ one can compute the corresponding empirical quantile for the desired test level and launch the sequential procedure with this bootstrap quantile instead of the (probably less precise) Gumbel quantile.

The following result yields the validity of the above proposed bootstrap procedure.

Theorem 3.13 (Bootstrap consistency) *Under the Assumptions 3.1-3.4 and 3.11 we have*

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}_{|\mathcal{X}}\Big(a_d\big(\widehat{\mathcal{T}}_{m,d}^{(Z)} - b_d\big) \le x\Big) - \mathbb{P}_{H_0}\Big(a_d\big(\widehat{\mathcal{T}}_{m,d} - b_d\big) \le x\Big) \right| = o_{\mathbb{P}}(1) \ ,$$

where $\mathbb{P}_{H_0}$ denotes the probability under the null hypothesis of no change in any component.

Combining Theorem 3.13 with Corollary 3.9 and Theorem 3.10, it follows that the use of the quantiles of the bootstrap distribution in (2.11) yields a consistent monitoring scheme, which keeps it pre-specified nominal level.

We conclude this section stating a detailed algorithm to monitor for a change point in the mean vector of a high dimensional time series. For this purpose we denote the set of all components without a change in the mean by

$$\mathcal{S}_d := \big\{ h \in \{1, \dots, d\} \;\big|\; \mu_{1,h} = \mu_{2,h} = \cdots = \mu_{m+Tm,h} \big\} \; . \tag{3.22}$$

An interesting feature of the following algorithm is its capability to identify the sets $\mathcal{S}_d$ and $\mathcal{S}_d^c = \{1, \dots, d\} \setminus \mathcal{S}_d$, see Theorem 3.15 below.

Algorithm 3.14

Step 1: Either choose the quantile q using the approximation by the Gumbel distribution, or obtain the quantile from the bootstrap as follows:

Step 1.1: Compute the long-run correlation estimates $\big(\hat{\rho}_{i,j}\big)_{i,j=1}^{d}$ from the initial set $\boldsymbol{X}_1, \dots, \boldsymbol{X}_m$.

Step 1.2: Based on the estimates, generate B independent realizations of the Gaussian vectors $(\widehat{Z}_{t,1}, \dots, \widehat{Z}_{t,d})^\top$ with covariance structure matrix $\big(\hat{\rho}_{i,j}\big)_{i,j=1}^{d}$ for $t = 1, \dots, m + Tm$ and compute the corresponding bootstrap statistics

$$a_d\big(\widehat{\mathcal{T}}_{m,d}^{(Z)}(1) - b_d\big), a_d\big(\widehat{\mathcal{T}}_{m,d}^{(Z)}(2) - b_d\big), \dots, a_d\big(\widehat{\mathcal{T}}_{m,d}^{(Z)}(B) - b_d\big)$$

defined in (3.21).

Step 1.3: Compute q as the empirical $(1-\alpha)$-quantile of the sample

$$\big\{ a_d\big(\widehat{\mathcal{T}}_{m,d}^{(Z)}(b) - b_d\big) \big\}_{b=1,\dots,B} \; .$$

Step 2: Monitoring: Initialize $\widehat{\mathcal{S}}_{d,\alpha} = \{1, \dots, d\}$. For $k = 1, \dots, Tm$ compute the statistics $\hat{E}_{m,h}(k)$. If the inequality

$$\max_{h \in \widehat{\mathcal{S}}_{d,\alpha}} w(k/m) \hat{E}_{m,h}(k) > \frac{q}{a_d} + b_d$$

holds, reject the null hypothesis in favor of the alternative. Eliminate the components that led to the rejection, i.e.

$$\widehat{\mathcal{S}}_{d,\alpha} \longleftarrow \widehat{\mathcal{S}}_{d,\alpha} \setminus \Big\{ h \in \widehat{\mathcal{S}}_{d,\alpha} \mid w(k/m) \hat{E}_{m,h}(k) > \frac{q}{a_d} + b_d \Big\} \tag{3.23}$$

and continue monitoring with the remaining components in $\widehat{\mathcal{S}}_{d,\alpha}$.

Step 3: If there was no rejection during monitoring, decide for the null hypothesis of no change in the mean vector. In case of rejections, decide for the alternative of a change in at least one component. The components remaining in the set

$$\widehat{\mathcal{S}}_{d,\alpha} = \Big\{ h \in \{1, \dots, d\} \;\Big|\; \max_{k=1}^{Tm} w(k/m)\hat{E}_{m,h}(k) \leq \frac{q}{a_d} + b_d \Big\} \tag{3.24}$$

after termination are assumed as mean stable.

The following theorem states, that Algorithm 3.14 is able to separate the sets $\mathcal{S}_d$ and $\mathcal{S}_d^c$ correctly. For its precise statement, let $k_h^* \in \{1, \dots, mT\}$ denote the time of change in the component $h \in \mathcal{S}_d^c$ within the monitoring period, that is

$$\mu_{1,h} = \cdots = \mu_{m,h} = \cdots = \mu_{m+k_h^*-1,h} \neq \mu_{m+k_h^*,h} = \dots \mu_{m+Tm,h} \ .$$

Theorem 3.15 *Let Assumptions 3.1 - 3.4 and 3.11 be satsified and assume that Algorithm 3.14 was launched either with the Gumbel or with the Bootstrap quantile. The set $\widehat{\mathcal{S}}_{d,\alpha}$ defined in (3.24) satisfies*

$$\limsup_{m,d\to\infty} \mathbb{P}\Big(\widehat{\mathcal{S}}_{d,\alpha} \subset \mathcal{S}_d\Big) \geq 1 - \alpha \ . \tag{3.25}$$

If further

$$\sqrt{\frac{m}{\log(m)}} \cdot \min_{h \in \mathcal{S}_d^c} |\mu_{m+k_h^*-1,h} - \mu_{m+k_h^*,h}| \to \infty \quad \text{and} \quad \limsup_{m,d\to\infty} \max_{h \in \mathcal{S}_d^c} \frac{k_h^*}{m} < T \tag{3.26}$$

then

$$\lim_{m,d\to\infty} \mathbb{P}\Big(\mathcal{S}_d^c \subset \widehat{\mathcal{S}}_{d,\alpha}^c\Big) = 1 \ . \tag{3.27}$$

4 Finite Sample Properties

In this section we investigate the finite sample properties of the new monitoring schemes by means of a simulation study and illustrate potential applications in a data example.

4.1 Simulation study

In our simulation study we consider the following models:

(M1) $X_{t,h} = \varepsilon_{t,h}$,

(M2) $X_{t,h} = 0.1X_{t-1,h} + \varepsilon_{t,h}$,

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

(M4) $X_{t,h} = e_{t,h}$,

where $\{\varepsilon_{t,h}\}_{t\in\mathbb{N},h\in\mathbb{N}}$ is an array of i.i.d. standard Gaussian random variables and $\{e_{t,h}\}_{t\in\mathbb{N},h\in\mathbb{N}}$ are Gaussian random variables, such that $\{\boldsymbol{e}_t = (e_{t,1},\dots,e_{t,d})^\top\}_{t\in\mathbb{Z}}$ are i.i.d. d-dimensional random vectors with covariance structure

$$\mathrm{Cov}(e_{1,j}, e_{1,i}) = \frac{1}{|j-i|+1} \ . \tag{4.1}$$

For the alternative hypothesis we also consider the models (M1)-(M4) and add a shift in the mean, at some point $m + k^*$, that is

$$X_{t,h}^{(\delta,A)} = \begin{cases} X_{t,h} & \text{if } \ t < m + k^* \ , \\ X_{t,h} + \delta \cdot I\{h \in A\} & \text{if } \ t \geq m + k^* \ , \end{cases} \tag{4.2}$$

where I denotes the indicator function, A is the set of spatial components affected by the change and δ is the size of the change. In order to examine the influence of both parameters on the procedure, we will consider different values of δ and three different choices of the set A below.

For the long-run variance estimation we use the quadratic spectral kernel estimator [see Andrews (1991)] in each component, that is

$$\hat{\sigma}_h^2 := \sum_{|t|\leq m-1} k\Big(\frac{t}{B_m}\Big)\hat{\phi}_{t,h} \tag{4.3}$$

where the empirical auto-covariances $\hat{\phi}_{t,h}$ are given by

$$\hat{\phi}_{t,h} := \frac{1}{m}\sum_{i=t+1}^{m} \big(X_{i,h} - \widehat{\mu}_1^m(h)\big)\big(X_{i-t,h} - \widehat{\mu}_1^m(h)\big) \ .$$

and the underlying kernel is given by

$$k(x) = \frac{25}{12\pi^2x^2}\bigg(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\bigg) \ .$$

In particular we employ the implementation of the estimator (4.3) provided by the R-package 'sandwich' [see Zeileis (2004)] and select the bandwidth parameter as $B_m = \log_{10}(m)$. Note that we only use the stable set $\boldsymbol{X}_1,\dots,\boldsymbol{X}_m$ for the estimation of the long-run variance, which avoids corruption from observations after the potential change point under the alternative [see the discussion in Remark 3.8]. All results presented in this section are based on 1000 simulation runs and for the AR(1)-process (M2) we employ a burn-in sample of 200 observations. The test level is always fixed at $\alpha = 0.05$.

		m=100		m=200		m=500	
T	model	d=100	d=200	d=200	d=500	d=200	d=500
1	(M1)	7.9%	8.7%	6.3%	7.6%	3.0%	4.6%
	(M2)	10.8%	14.0%	9.9%	11.4%	5.5%	6.4%
	(M3)	5.7%	7.4%	5.9%	6.8%	3.4%	4.7%
	(M4)	6.0%	9.4%	5.8%	7.0%	4.2%	4.6%
2	(M1)	9.0%	10.7%	6.0%	8.4%	4.5%	6.9%
	(M2)	13.5%	14.4%	10.0%	11.3%	6.7%	5.7%
	(M3)	9.6%	10.4%	6.1%	6.5%	4.2%	5.1%
	(M4)	6.3%	9.5%	6.8%	8.0%	3.6%	4.8%
4	(M1)	7.6%	10.1%	6.5%	8.6%	3.7%	5.1%
	(M2)	13.9%	15.7%	10.0%	12.0%	5.5%	6.3%
	(M3)	8.7%	10.5%	6.2%	6.4%	3.4%	4.5%
	(M4)	9.8%	9.2%	6.5%	8.1%	3.8%	6.6%

Table 1: *Approximation of the nominal level by the detector defined in* (2.11) *for different choices of initial sample size m, dimension d and monitoring duration $m \cdot T$. Critical values are obtained from Corollary 3.9 (approximation by Gumbel distribution).*

In Table 1 and 2 we illustrate the finite sample properties of the detection scheme (2.11) under the null hypothesis for different choices of the sample size m, the dimension d and the length of the monitoring period determined by T. The results in Table 1 are based on the weak convergence in Corollary 3.9 and therefore we use the critical value $q = c_{d,\alpha} = g_{1-\alpha}/a_d + b_d$ in (2.11), where $g_{1-\alpha}$ is the quantile of the Gumbel distribution.

The results in Table 2 are obtained by the bootstrap procedure as described in Step 2 of Algorithm 3.14. We note that the use of (any) sequential monitoring scheme in a simulation study is computationally demanding in particular in a high dimensional setup. In our case we have to estimate the spatial correlation structure in each simulation run and then simulate the quantile of the distribution of the statistic $\widehat{\mathcal{T}}^{(Z)}_{m,d}$ defined in (3.21). Of course, this is no problem in data analysis as in this case the monitoring procedure has only to be used once, but it requires large computational resources in a simulation, where the same procedure is repeated 1000 times. Therefore, in order to reduce the computational complexity, of the bootstrap approach in the simulation study we do not estimate the spatial correlation structure for the bootstrap but employ temporal and spatial independent Gaussian random variables to generate the bootstrap statistics defined in (3.21). With this adaption, the quantiles are fixed within each column of Table 2, which makes the simulation study practicable. Moreover, it can easily be seen from the theory developed in Section 3 that the use of these quantiles also yields a consistent test in (2.11) (note that in Lemma A.1 the Gumbel limit distribution is derived for independent Brownian motions).

In Table 1, we observe a reasonable approximation of the nominal level by the asymptotic test in many cases, which becomes more accurate with larger initial sample size m and dimension d. For instance consider the model (M2) for the choice $T = 2$, where we have obtained a type I error of 13.5% for $m = d = 100$ and 10.0% for $m = d = 200$. This finally reduces to an appropriate approximation of 5.7% for the choice $m = d = 500$. As common for high dimensional procedures, the relation of sample size m and dimension d has a severe impact on the performance of the monitoring procedure. For example, an empirical type I error of 6.0% was measured for model (M4) with $m = d = 100$ and $T = 1$, which increases to 9.4% if the dimension is set to $d = 200$. This effect becomes weaker, when the sample size is generally increased.

In Table 2 we display the type I error for the method where the quantiles are calculated by bootstrap as described above. For the sake of brevity we focus on the case $T = 1$, as the results obtained for different choices of T are similar. We observe a very reasonable approximation of the desired test level and - compared to the results in Table 1 - a substantial improvement by the bootstrap procedure in nearly all cases under consideration. In particular, the bootstrap should be taken into account when m and d are relatively small, where the approximation of the nominal level using the quantiles from the Gumbel distribution is rather imprecise, while the extra computational costs for the bootstrap are still tolerable.

		m=100		m=200		m=500	
T	model	d=100	d=200	d=200	d=500	d=200	d=500
1	(M1)	4.7%	5.2%	5.1%	4.9%	5.0%	5.3%
	(M2)	6.9%	11.1%	8.1%	7.0%	8.1%	6.8%
	(M3)	3.7%	5.0%	4.6%	4.4%	6.1%	5.5%
	(M4)	4.0%	7.0%	4.1%	4.9%	6.4%	5.5%

Table 2: *Approximation of the nominal level using the detector defined in* (2.11) *for different choices of initial sample size m and dimension d. Critical values are computed by the bootstrap with spatial independence.*

To analyze the performance of the sequential procedure under the alternative hypothesis we consider the model (4.2), where the processes $X_{t,h}$ are defined by (M1)-(M4). Here we distinguish between the following three scenarios:

(A1) The change occurs only in one component. This corresponds to the choice $A = \{1\}$.

(A2) The change occurs in 50% of the components, i.e. $A = \{1, \dots, d/2\}$.

(A3) The change occurs in all components, i.e. $A = \{1, \dots, d\}$.

For the sake of brevity and readability, we focus on the case $T = 1$ under the alternative and only consider change positions in the middle of the monitoring period, i.e. we fix $k^* = m + m/2$.

In Figures 1 and 2 we display the rejection probabilities of the detection rule (2.11) for these scenarios, different values of the change, different sample size and dimensions. The critical values in (2.11) have been obtained by the spatial independent bootstrap (as they are usually more accurate than the quantiles derived from the Gumbel distribution).

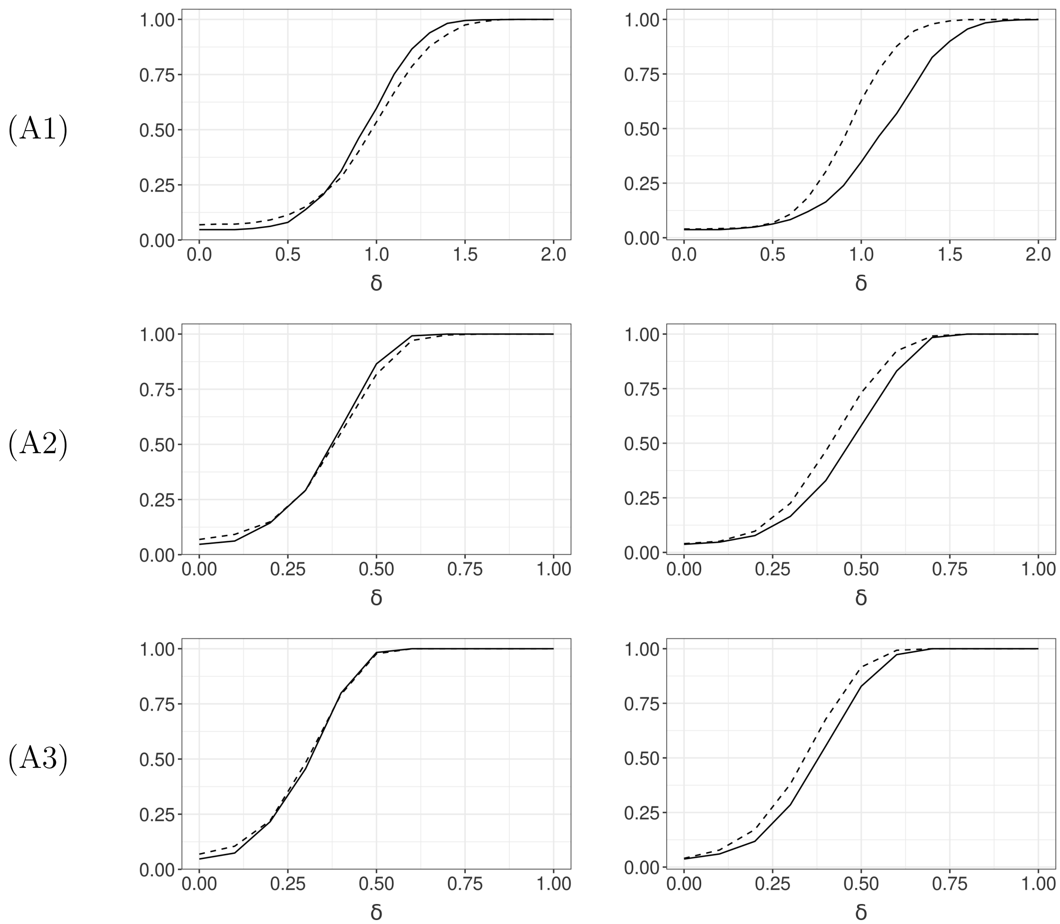

Figure 1: *Simulated power of the monitoring scheme* (2.11) *for different size* δ *of the change, sample size* $m = 100$ *and dimension* $d = 100$*. Left panels: Solid line (M1), dashed line (M2). Right panels: Solid line (M3), dashed line (M4).*

The results can be sumarized as follows. In all considered scenarios the new monitoring procedure (2.11) for a change in the high dimensional mean vector has reasonable power under the alternative, and in all cases the type II error approaches zero for an increasing size δ of the change. As expected, the power is lower under alternative (A1), where the change occurs in only one coordinate. To give an example, consider model (M1) and (M2) corresponding to the left columns in Figures 1 and 2. The results for the different alternatives (A1), (A2) and (A3) can be found in the first, second and third rows of the figures, respectively. If the sample size and dimension are given by $m = 100$ and $d = 100$ we observe from Figure 1 that for $\delta = 0.75$ the power for model (M1) and (M2) under alternative (A1) is approximately given by 0.25, while it is 1 under alternative (A2). Interestingly, the differences between alternatives (A2) and (A3) are not so strong, but they are still clearly visible. For instance, a comparison between the left parts of the second and third row of Figure 2 shows that the power of the detection scheme (2.11) in model (M1) for $\delta = 0.3$ is approximately 0.72 for alternative (A2) and 0.92 for alternative (A3).

The differences between the four data generating models are in general not substantial with one exception. In model (M3) under alternative (A1) the power of the detection scheme (2.11) is considerably smaller [see the first rows in Figure 1 and Figure 2].

We summarize the discussion of the finite sample properties emphasizing that our numerical results have supported the theoretical findings developed in Section 3 in all cases under consideration.

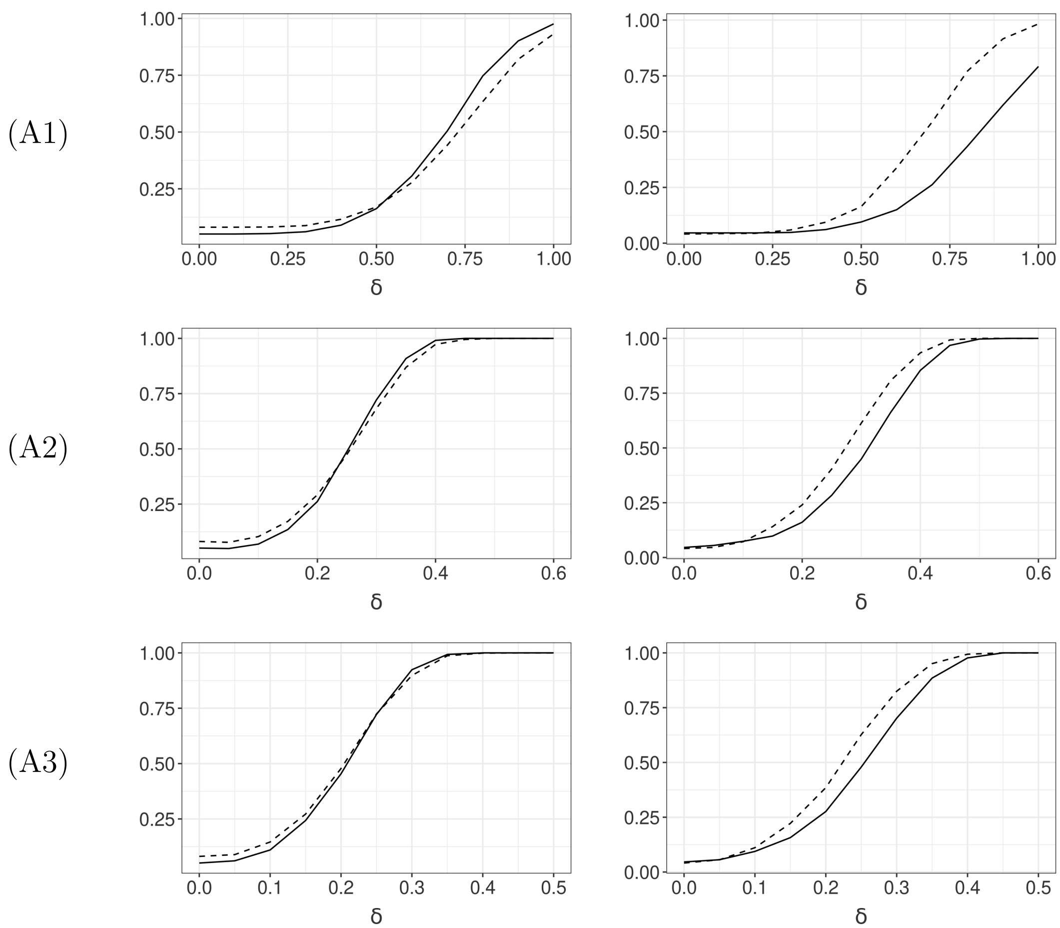

Figure 2: *Simulated power of the monitoring scheme* (2.11) *for different size* δ *of the change, sample size* $m = 200$ *and dimension* $d = 200$*. Left panels: Solid line (M1), dashed line (M2) Right panels: Solid line (M3), dashed line (M4).*

4.2 Data example

In this section we illustrate potential applications of the new monitoring scheme in a data example. For this purpose we consider a data set sampled in hydrology, which consists of the average daily flows measured in m^3/sec of the river Chemnitz at Göritzhain in Saxony, Germany, for the years 1909-2013.

This data has been previously analyzed in the (retrospective) change point literature by Sharipov et al. (2016), who developed methodology for detecting change points in functional data. The data set consists of a sample of $n = 105$ observations with dimension $d = 365$, such that each vector $\boldsymbol{X}_t = (X_{t,1}, \ldots, X_{t,365})^\top$ contains the daily average flows of one (German) hydrological year, which lasts from 1st of November to 31st of October. For instance, the data point $X_{1,1}$ represents the daily average flow of the 1st of November 1909, while $X_{105,365}$ is the same key figure for the 31st of October 2014. By this transformation, Sharipov et al. (2016) located a change in the annual flow curves in the year 1964. Dette et al. (2018) propose a retrospective test for relevant changes in a high dimensional time series. They consider the same data and locate 4 different mean changes that exceed a test threshold of 0.63 and are traced back to the dates 10th of July 1950, 18th of March 1956, 23rd of December 1965, 7th of February 1979, which correspond to spatial components 252, 138, 53, 99 respectively.

Based on these prior analyses, we consider the first 35 observations as our initial, stable data set and will use the remaining 70 observations as the monitoring period corresponding to a choice of $m = 35$ and $T = 2$. From the initial set $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{35}$ the spatial correlation structure is estimated via the implementation of the estimator based on the quadratic spectral kernel provided in the R-package 'sandwich' [see Zeileis (2004)]. By the bootstrap in Algorithm 3.14 critical values are obtained as $q_{0.99} = 7.43$ and $q_{0.95} = 4.93$, for which we conduct our monitoring method. During the monitoring period, we proceed as described in Algorithm 3.14: If the detection scheme rejects the null hypothesis of no change at a certain time point, we report the corresponding component(s) as instable and remove it/them from the sample. Afterwards, we continue monitoring with the remaining components until there's is another rejection or the end of the monitoring period is attained.

The results of this procedure are displayed in Table 4.2 for the test levels $\alpha = 0.01$ and $\alpha = 0.05$ and can summarized as follows. For a test level of 0.05 more instable components (33) are identified than for 0.01 (17 components). Naturally, all breaks identified with the lower test level, are also detected by the other one, while the time of detection is sometimes earlier in the latter case. As the data exhibits (positive) spatial correlation, breaks partially occur in clusters, for example consider components 285, 286 and 287, for which both test levels detect changes or components 214, 215 and 216, for which changes are found at test level 0.05. It is worth to mention that our findings match three out of four instable components identified by the threshold procedure of Dette et al. (2018). Namely, we refer to components 53, 99, 252, which are likewise identified to contain a break by our sequential analysis. To illustrate the data set, we finally display the average daily flow over the years for these three components in Figure 4.2. The plots indicate that the break in component 252 (10th of July) is most probably caused by a huge outlier in the year 1953, which leads to an immediate rejection, while there seems to be *actual* structural changes in the components 53 (23rd of December) and 99 (7th of February).

$\alpha = .05$		$\alpha = .01$	
component	year	component	year
101	1945	101	1945
252, 253	1953	252, 253	1953
249, 251	1954	249	1957
247	1957	105	1960
105	1960	189	1977
189, 191	1977	191	1979
104	1979	100	1980
100, 190	1980	104	1986
102	1986	53	1995
53, 57	1993	285, 286, 287	2001
54, 209	1995	280	2009
264	1996	54, 215	2012
285, 286, 287	2001	209	2013
92	2002		
99,192	2003		
138	2004		
280, 283	2009		
44	2010		
55, 214, 215, 216	2012		
199	2013		

Table 3: *Structural breaks detected in the river flow data for a test level of* $\alpha = 0.05$ *(left column) and* $\alpha = 0.01$ *(right column). The column 'year' specifies the (hydrological) year after which the rejection occurred. For instance 1945 means that the data from the hyrodological year 01st November 1945 to 31st October 1946 was already under consideration.*

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 also like to thank Johannes Heiny and Florian Heinrichs for extremely helpful discussions during the preparation of this manuscript.

Moreover we are grateful to Andreas Schuhmann and Svenja Fischer from the Institute of Hydrology in Bochum, who provided hydrological data analyzed in Section 4.2.

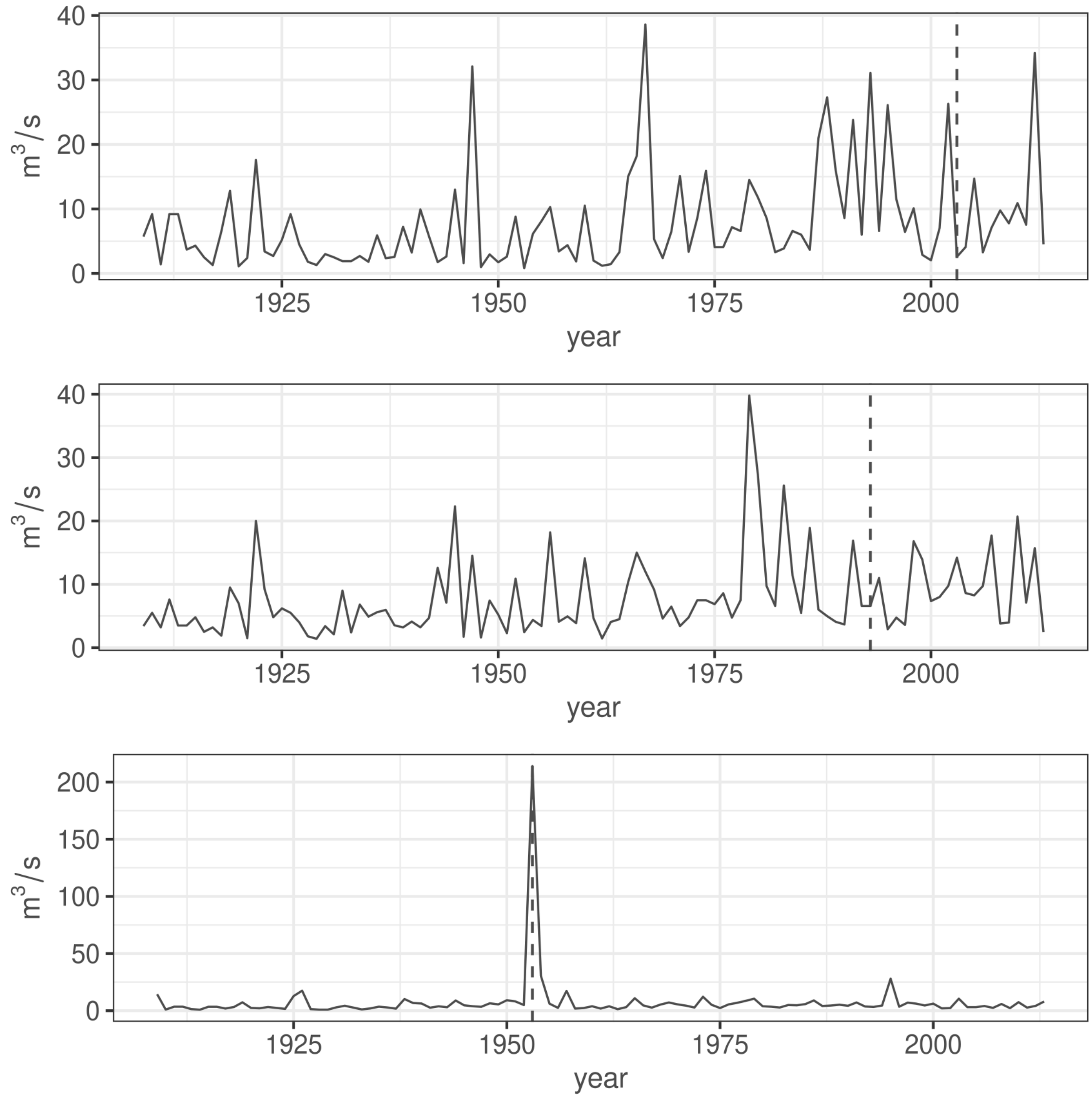

Figure 3: *Average daily flows for the dates 23rd of December (spatial component 53, upper row), 7th of February (spatial component 99, middle row) and 10th of July (spatial component 252, lower row). Vertical dashed lines indicate the time points at which a break was detected by the sequential method with a test level of* 5%.

References

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

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

Arratia, R., Goldstein, L., and Gordon, L. (1989). Two moments suffice for Poisson approximations: The Chen-Stein method. *The Annals of Probability*, 17(1):9–25.

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

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

Aue, A., Horváth, L., Hušková, M., and Kokoszka, P. (2006). Change-point monitoring in linear models. *The Econometrics Journal*, 9(3):373–403.

Avanesov, V. and Buzun, N. (2018). Change-point detection in high-dimensional covariance structure. *Electronic Journal of Statistics*, 12(2):3254–3294.

Berkes, I., Liu, W., and Wu, W. B. (2014). Komlós-major-tusnády approximation under dependence. *The Annals of Probability*, 42(2):794–817.

Billingsley, P. (1999). *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. Wiley.

Borodin, A. and Salminen, P. (1996). *Handbook of Brownian Motion - Facts and Formulae*. Probability and its applications. Springer.

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

Chernozhukov, V., Chetverikov, D., and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, 41(6):2786–2819.

Chernozhukov, V., Chetverikov, D., Kato, K., and Koike, Y. (2019). Improved central limit theorem and bootstrap approximations in high dimensions. *Preprint:https://arxiv.org/pdf/1912.10529.pdf*.

Cho, H. and Fryzlewicz, P. (2015). Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(2):475–507.

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

Chu, L. and Chen, H. (2018). Sequential change-point detection for high-dimensional and non-euclidean data. *Preprint: https://arxiv.org/abs/1810.05973*.

Ciuperca, G. (2013). Two tests for sequential detection of a change-point in a nonlinear model. *Journal of Statistical Planning and Inference*, 143(10):1719–1743.

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. Springer New York.

Dette, H. and Gösmann, J. (2018). Relevant change points in high dimensional time series. *Electronic Journal of Statistics*, 12(2):2578–2636.

Dette, H. and Gösmann, J. (2019). A likelihood ratio approach to sequential change point detection for a general class of parameters. *Journal of the American Statistical Association*.

Dette, H., Pan, G. M., and Yang, Q. (2018). Estimating a change point in a sequence of very high-dimensional covariance matrices.

El Machkouri, M., Volný, D., and Wu, W. B. (2013). A central limit theorem for stationary random fields. *Stochastic Processes and their Applications*, 123(1):1–14.

Enikeeva, F. and Harchaoui, Z. (2019). High-dimensional change-point detection under sparse alternatives. *The Annals of Statistics*, 47(4):2051–2079.

Feller, W. (1951). The asymptotic distribution of the range of sums of independent random variables. *The Annals of Mathematical Statistics*, 22(3):427–432.

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

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

Gösmann, J., Kley, T., and Dette, H. (2019). A new approach for open-end sequential change point monitoring. *Preprint*.

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

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

Hušková, M. and Kirch, C. (2012). Bootstrapping sequential change-point tests for linear regression. *Metrika*, 75(5):673–708.

Hušková, M. and Koubková, A. (2005). Monitoring jump changes in linear models. *Journal of Statistical Research*, 39(2):51–70.

Jiang, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *The Annals of Applied Probability*, 14(2):865–880.

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

Karatzas, I. and Shreve, S. (1991). *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer New York, second edition.

Kirch, C. (2008). Bootstrapping sequential change-point tests. *Sequential Analysis*, 27(3):330–349.

Kirch, C. and Kamgaing, J. T. (2015). On the use of estimating functions in monitoring time series for change points. *Journal of Statistical Planning and Inference*, 161:25–49.

Kirch, C. and Stoehr, C. (2019). Sequential change point tests based on U-statistics. *arXiv preprint arXiv:1912.08580*.

Kirch, C. and Weber, S. (2018). Modified sequential change point procedures based on estimating functions. *Electronic Journal of Statistics*, 12(1):1579–1613.

Lévy-Leduc, C. and Roueff, F. (2009). Detection and localization of change-points in high-dimensional network traffic data. *Annals of Applied Statistics*, 3(2):637–662.

Liu, W., Xiao, H., and Wu, W. B. (2013). Probability and moment inequalities under dependence. *Statistica Sinica*, 23(3):1257–1272.

Otto, S. and Breitung, J. (2019). Backward CUSUM for testing and monitoring structural change. *Preprint: http://www.wisostat.uni-koeln.de/sites/statistik/BackwardCUSUM.pdf*.

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

Serfling, R. J. (2009). *Approximation Theorems of Mathematical Statistics. Wiley Series in Probability and Statistics.* John Wiley & Sons, New York.

Sharipov, O., Tewes, J., and Wendler, M. (2016). Sequential block bootstrap in a Hilbert space with application to change point analysis. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 44(3):300–322.

Soh, Y. S. and Chandrasekaran, V. (2017). High-dimensional change-point estimation: Combining filtering with convex optimization. *Applied and Computational Harmonic Analysis*, 43(1):122 – 147.

Steland, A. (2006). A Bootstrap view on Dickey-Fuller control charts for AR(1) series. *Austrian Journal of Statistics*, 35:339–346.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

Wang, D., Yu, Y., and Rinaldo, A. (2017). Optimal covariance change point localization in high dimension.

Wang, R. and Shao, X. (2020). Dating the break in high-dimensional data. *Preprint: https://arxiv.org/abs/2002.04115*.

Wang, R., Volgushev, S., and Shao, X. (2019). Inference for change points in high dimensional data. *Preprint: https://arxiv.org/abs/1905.08446*.

Wang, T. and Samworth, R. J. (2018). High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1):57–83.

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.

Woodall, W. H. and Montgomery, D. C. (1999). Research issues and ideas in statistical process control. *Journal of Quality Technology*, 31(4):376–386.

Wu, W. B. (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.

Zhang, X. and Cheng, G. (2018). Gaussian approximation for high dimensional vector under physical dependence. *Bernoulli*, 24(4A):2640–2675.

A Proofs of main results

Throughout the appendix the symbol $\lesssim$ denotes an inequality up to a constant, which does not depend on size m of the training sample and the dimension d.

A.1 Some preliminary results

We will begin with two auxiliary results. In Lemma A.1 we investigate the weak convergence of the maximum of independent identically distributed random variables with the same distribution as the random variable $\mathbb{M}$ defined in (2.9) and in Lemma A.2 we analyze the asymptotic behavior of the scaling sequences defined in (3.7).

Lemma A.1 *Let $M_1', M_2', \dots$ be independent identically distributed random variables with*

$$M_1' \overset{\mathcal{D}}{=} \max_{0\le t\le q(T)} \max_{0\le s\le t} |W(t) - W(s)| \ ,$$

where W denotes a standard Brownian motion. Then, it holds that

$$a_d\Big(\max_{h=1}^{d} M_h' - b_d\Big) \overset{\mathcal{D}}{\Longrightarrow} G \ .$$

as $d \to \infty$ with a_d and b_d as in (3.7).

Proof of Lemma A.1. First note, that

$$\begin{aligned} M_h' \overset{\mathcal{D}}{=} \max_{0\le t\le q(T)} \max_{0\le s\le t} |W_h(t) - W_h(s)| &= \max_{0\le t\le q(T)} \max_{0\le s\le q(T)} |W_h(t) - W_h(s)| \\ &= \max_{0\le t\le q(T)} \max_{0\le s\le q(T)} W_h(t) - W_h(s) \\ &= \max_{0\le t\le q(T)} W_h(t) - \min_{0\le t\le q(T)} W_h(t) \ . \end{aligned}$$

By Borodin and Salminen (1996), page 146, the distribution function of M_h' is given by

$$F_{\mathbb{M}}(x) = \begin{cases} 1 + 4\sum_{k=1}^{\infty} (-1)^k k \operatorname{Erfc}\Big(\frac{kx}{\sqrt{2q(T)}}\Big) & \text{if } x > 0 \ , \\ 0 & \text{otherwise} \ , \end{cases}$$

where $\operatorname{Erfc} = 1 - \operatorname{Erf}$ denotes the complementary error function. Using the elementary property $\operatorname{Erf}(x) = 2\Phi(x\sqrt{2}) - 1$ we obtain for $x > 0$

$$1 + 4\sum_{k=1}^{\infty} (-1)^k k \operatorname{Erfc}\Big(\frac{kx}{\sqrt{2q(T)}}\Big) = 1 + 8\sum_{k=1}^{\infty} (-1)^k k \Phi\Big(-\frac{kx}{\sqrt{q(T)}}\Big) \tag{A.1}$$

and the desired extreme value convergence will be derived by a closer investigation of the distribution function $F_{\mathbb{M}}$. Observe, that $F_{\mathbb{M}}$ is twice differentiable with derivatives (for $x > 0$)

$$F_{\mathbb{M}}'(x) = \frac{4\sqrt{2}}{\sqrt{\pi q(T)}} \sum_{k=1}^{\infty} (-1)^{k+1} k^2 \exp\Big(-\frac{k^2 x^2}{2q(T)}\Big) \ ,$$

$$F''_{\mathbb{M}}(x) = \frac{4\sqrt{2}}{\sqrt{\pi q(T)}q(T)} \sum_{k=1}^{\infty} (-1)^k k^4 x \exp\Big(-\frac{k^2x^2}{2q(T)}\Big) ,$$

where we used that the series converge uniformly on all intervals $[\varepsilon, \infty)$ for $\varepsilon > 0$ and therefore term by term differentiation is allowed. Thus, by Theorem 1.1.8 from de Haan and Ferreira (2006) the distribution function $F_{\mathbb{M}}$ is in the domain of attraction of the Gumbel distribution if

$$\lim_{x\to\infty} \frac{(1 - F_{\mathbb{M}}(x))F''_{\mathbb{M}}(x)}{(F'_{\mathbb{M}}(x))^2} = -1 \tag{A.2}$$

and in that case the stated shape of the scaling sequences b_d and a_d follow from Remark 1.1.9 from de Haan and Ferreira (2006) with

$$b_d = U_{\mathbb{M}}(d) \quad \text{and} \quad a_d = dF'_{\mathbb{M}}(U_{\mathbb{M}}(d)) = \frac{F'_{\mathbb{M}}(U_{\mathbb{M}}(d))}{1 - F_{\mathbb{M}}(U_{\mathbb{M}}(d))} .$$

First note that by the definition of the complimentary error function, we obtain for $x > 0$

$$\begin{aligned} x\big(1 - F_{\mathbb{M}}(x)\big) \exp\Big(\frac{x^2}{2q(T)}\Big) &= 4\sum_{k=1}^{\infty}(-1)^{k+1}kx \operatorname{Erfc}\Big(\frac{kx}{\sqrt{2q(T)}}\Big) \exp\Big(\frac{x^2}{2q(T)}\Big) \\ &= A_1(x) + A_2(x) \end{aligned} \tag{A.3}$$

where the two summands on the right-hand side are given by

$$\begin{aligned} A_1(x) &= \frac{8}{\sqrt{\pi}} \int_{\frac{x}{\sqrt{2q(T)}}}^{\infty} \exp(-\tau^2)d\tau \cdot x \exp\Big(\frac{x^2}{2q(T)}\Big) , \\ A_2(x) &= \frac{8}{\sqrt{\pi}} \sum_{k=2}^{\infty}(-1)^{k+1}kx \int_{\frac{kx}{\sqrt{2q(T)}}}^{\infty} \exp(-\tau^2)d\tau \cdot \exp\Big(\frac{x^2}{2q(T)}\Big) . \end{aligned}$$

We will treat the two summands of the last display separately. For the first summand we obtain that

$$\begin{aligned} \lim_{x\to\infty} A_1(x) =& \frac{8}{\sqrt{\pi}} \lim_{x\to\infty} \frac{\int_{\frac{x}{\sqrt{2q(T)}}}^{\infty} \exp(-\tau^2)d\tau}{x^{-1}\exp\Big(\frac{-x^2}{2q(T)}\Big)} \\ =& \frac{8}{\sqrt{\pi}} \lim_{x\to\infty} \frac{\frac{-1}{\sqrt{2q(T)}}\exp\Big(\frac{-x^2}{2q(T)}\Big)}{-x^{-2}\exp\Big(\frac{-x^2}{2q(T)}\Big) - \frac{1}{q(T)}\exp\Big(\frac{-x^2}{2q(T)}\Big)} = 4\frac{\sqrt{2q(T)}}{\sqrt{\pi}} \end{aligned}$$

by L'Hôspital's rule. For the second summand of the last display in (A.3) note that

$$\Big|\frac{\sqrt{\pi}}{8}A_2(x)\Big| \le \sum_{k=2}^{\infty} kx \int_{\frac{kx}{\sqrt{2q(T)}}}^{\infty} \exp(-\tau^2)d\tau \cdot \exp\Big(\frac{x^2}{2q(T)}\Big)$$

$$\leq \sqrt{2q(T)} \sum_{k=2}^{\infty} \int_{\frac{kx}{\sqrt{2q(T)}}}^{\infty} \tau \exp(-\tau^2) d\tau \cdot \exp\Big(\frac{x^2}{2q(T)}\Big)$$

$$= \sqrt{\frac{q(T)}{2}} \exp\Big(\frac{-x^2}{2q(T)}\Big) \sum_{k=2}^{\infty} \exp\Big(-\frac{(k^2-2)x^2}{2q(T)}\Big)$$

$$\overset{(x\geq 1)}{\leq} \sqrt{\frac{q(T)}{2}} \exp\left(\frac{-x^2}{2q(T)}\right) \sum_{k=2}^{\infty} \exp\left(-\frac{k^2-2}{2q(T)}\right) = o(1) \ \text{ as } \ x \to \infty \ .$$

Combining the last statements with the decomposition in (A.3) yields

$$\lim_{x\to\infty} x\big(1 - F_{\mathbb{M}}(x)\big) \exp\Big(\frac{x^2}{2q(T)}\Big) = 4\frac{\sqrt{2q(T)}}{\sqrt{\pi}} \ . \tag{A.4}$$

For the denominator of (A.2) note that

$$\begin{aligned} \lim_{x\to\infty} F'_{\mathbb{M}}(x) \exp\Big(\frac{x^2}{2q(T)}\Big) &= \frac{4\sqrt{2}}{\sqrt{\pi q(T)}} + \lim_{x\to\infty} \frac{4\sqrt{2}}{\sqrt{\pi q(T)}} \sum_{k=2}^{\infty} (-1)^{k+1} k^2 \exp\Big(-\frac{(k^2-1)x^2}{2q(T)}\Big) \\ &= \frac{4\sqrt{2}}{\sqrt{\pi q(T)}} \ , \end{aligned} \tag{A.5}$$

where we used that for $x \geq 1$

$$\sum_{k=2}^{\infty} k^2 \exp\Big(-\frac{(k^2-1)x^2}{2q(T)}\Big) \leq \exp\Big(\frac{-x^2}{2q(T)}\Big) \sum_{k=2}^{\infty} k^2 \exp\Big(-\frac{(k^2-2)}{2q(T)}\Big) = o(1) \ \text{ as } \ x \to \infty \ .$$

Using similar arguments we obtain

(A.6)

$$\begin{aligned} &\lim_{x\to\infty} x^{-1} F''_{\mathbb{M}}(x) \exp\Big(\frac{x^2}{2q(T)}\Big) \\ &\quad = \frac{-4\sqrt{2}}{\sqrt{\pi q(T)} q(T)} + \lim_{x\to\infty} \frac{4\sqrt{2}}{\sqrt{\pi q(T)} q(T)} \sum_{k=2}^{\infty} (-1)^k k^4 \exp\Big(-\frac{x^2(k^2-1)}{2q(T)}\Big) = \frac{-4\sqrt{2}}{\sqrt{\pi q(T)} q(T)} \ . \end{aligned}$$

Combining (A.4), (A.5) and (A.6), it follows that

$$\begin{aligned} &\lim_{x\to\infty} \frac{(1 - F_{\mathbb{M}}(x)) F''_{\mathbb{M}}(x)}{(F'_{\mathbb{M}}(x))^2} = \\ &\qquad = \frac{\lim_{x\to\infty} x(1 - F_{\mathbb{M}}(x)) \exp\Big(\frac{x^2}{2q(T)}\Big) \cdot \lim_{x\to\infty} x^{-1} F''_{\mathbb{M}}(x) \exp\Big(\frac{x^2}{2q(T)}\Big)}{\lim_{x\to\infty} \Big(F'_{\mathbb{M}}(x) \exp\Big(\frac{x^2}{2q(T)}\Big)\Big)^2} = -1 \ , \end{aligned}$$

which completes the proof. □

Lemma A.2 *For the sequences a_d and b_d defined in (3.7), it holds that*

(i) $\lim_{d\to\infty} \frac{b_d}{\sqrt{\log(d)}} = \sqrt{2q(T)}$,

(ii) $\lim_{d\to\infty} \frac{a_d}{\sqrt{\log(d)}} = \sqrt{\frac{2}{q(T)}}$.

Proof of Lemma A.2. We will begin with the investigation of $b_d = U_{\mathbb{M}}(d)$, where $U_{\mathbb{M}}$ defined in Theorem 3.5. Let $c_b < \sqrt{2q(T)} < C_b$ be arbitrary positive constants. Observing the relation (A.4) from the proof of Lemma A.1 we obtain

$$\lim_{x\to\infty} \big(1 - F_{\mathbb{M}}(x)\big) \exp\Big(\frac{x^2}{c_b^2}\Big)$$
$$= \lim_{x\to\infty} x\big(1 - F_{\mathbb{M}}(x)\big) \exp\Big(\frac{x^2}{2q(T)}\Big) \cdot \lim_{x\to\infty} \frac{1}{x} \exp\Big(\Big(\frac{1}{c_b^2} - \frac{1}{2q(T)}\Big)x^2\Big) = \infty$$

and

$$\lim_{x\to\infty} \big(1 - F_{\mathbb{M}}(x)\big) \exp\Big(\frac{x^2}{C_b^2}\Big) = 0 \ .$$

Defining the function $H(x) := U_{\mathbb{M}}^{-1}(x) = \dfrac{1}{1 - F_{\mathbb{M}}(x)}$ yields

$$0 = \lim_{x\to\infty} \frac{\exp\big(-x^2/c_b^2\big)}{1 - F_{\mathbb{M}}(x)} = \lim_{x\to\infty} H(x) \exp\big(-x^2/c_b^2\big)$$

and

$$\infty = \lim_{x\to\infty} \frac{\exp\big(-x^2/C_b^2\big)}{1 - F_{\mathbb{M}}(x)} = \lim_{x\to\infty} H(x) \exp\big(-x^2/C_b^2\big)$$

and so there exists a sufficiently large constant x^* such that for all $x \geq x^*$

$$\exp\big(x^2/C_b^2\big) \leq H(x) \leq \exp\big(x^2/c_b^2\big),$$

or equivalently

$$U_{\mathbb{M}}\big(\exp\big(x^2/C_b^2\big)\big) \leq x \leq U_{\mathbb{M}}\big(\exp\big(x^2/c_b^2\big)\big) \ ,$$

where we used that the function $U_{\mathbb{M}}$ is monotone increasing. A further use of monotonicity gives that for all y with $c_b\sqrt{\log(y)} > x^*$

$$c_b\sqrt{\log(y)} \leq U_{\mathbb{M}}(y) \quad \text{and} \quad U_{\mathbb{M}}(y) \leq C_b\sqrt{\log(y)} \ .$$

Due to the choice of c_b, C_B we Therefore obtain that for all $\varepsilon > 0$ and y sufficiently large (depending on ε)

$$\sqrt{2q(T)} - \varepsilon \leq \frac{U_{\mathbb{M}}(y)}{\sqrt{\log(y)}} \leq \sqrt{2q(T)} + \varepsilon \ ,$$

which yields the claim for $b_d = U_{\mathbb{M}}(d)$ and so the proof of part *(i)* is completed. To obtain the claim for a_d in part *(ii)* we use (A.4) and (A.5), which yields

$$\lim_{d\to\infty} \frac{a_d}{\sqrt{\log d}} = \lim_{d\to\infty} \frac{b_d}{\sqrt{\log d}} \frac{F'_{\mathbb{M}}(b_d)}{b_d\big(1 - F_{\mathbb{M}}(b_d)\big)} = \sqrt{\frac{2}{q(T)}}$$

as we know that $b_d \to \infty$. □

A.2 Proof of Theorem 3.5

Recall the definition of M_h in (3.6). By Theorem 1 from Arratia et al. (1989) in the form as presented in Lemma A.4 in Jiang (2004) we obtain for any $x \in \mathbb{R}$ the inequality

(A.7) $$\left|\mathbb{P}\left(\max_{h=1}^{d} M_h \leq u_d(x)\right) - \exp(-\lambda)\right| \leq \big(1 \wedge \lambda^{-1}\big)(\Lambda_1 + \Lambda_2 + \Lambda_3\big) \ ,$$

where $\lambda = \sum_{h=1}^{d} \mathbb{P}\big(M_h > u_d(x)\big)$,

$$\begin{aligned}
\Lambda_1 &= \sum_{\substack{1\leq i,j\leq d \\ |i-j|\leq L_d}} \mathbb{P}\big(M_i > u_d(x)\big)\mathbb{P}\big(M_j > u_d(x)\big) \ , \\
\Lambda_2 &= \sum_{\substack{1\leq i,j\leq d \\ i\neq j, |i-j|\leq L_d}} \mathbb{P}\big(M_i > u_d(x), M_j > u_d(x)\big) \ , \\
\Lambda_3 &= \sum_{1\leq i\leq d} \mathbb{E}\left|\mathbb{P}\big(M_i > u_d(x) | M_j : |i-j| > L_d\big) - \mathbb{P}(M_i > u_d(x)\big)\right|
\end{aligned}$$

and $u_d(x) = x/a_d + b_d$. Let $M'_1, M'_2, \ldots$ be i.i.d. random variables with $M'_1 \stackrel{\mathcal{D}}{=} M_1$. With Lemma A.1 we have

(A.8) $$\lim_{d\to\infty} \mathbb{P}\Big(\max_{h=1}^{d} M'_h \leq u_d(x)\Big) = \lim_{d\to\infty} \Big(\mathbb{P}\big(M_1 \leq u_d(x)\big)\Big)^d = \exp\big(-\exp(-x)\big) \ .$$

As $b_d \to \infty$ and $\lim_{x\to 0} \frac{x}{\log(1-x)} = -1$, (A.8) yields

(A.9) $$\begin{aligned}
\lambda &= \sum_{h=1}^{d} \mathbb{P}\big(M_h > u_d(x)\big) = d\mathbb{P}\big(M_1 > u_d(x)\big) \\
&= -\, d \log\big(1 - \mathbb{P}\big(M_1 > u_d(x)\big)\big)\big(1 + o(1)\big) \\
&= -\,\big(\log\big(\mathbb{P}(M_1 \leq u_d(x))^d\big)\big)\big(1 + o(1)\big) \to \exp(-x) \quad \text{as } d \to \infty \ .
\end{aligned}$$

With Lemma A.2 it holds for $\sqrt{q(T)} < c_u < \sqrt{2q(T)}$ and sufficiently large d

$$u_d(x) \geq c_u \sqrt{\log d}\ . \tag{A.10}$$

Hence, it follows from (A.4) and Assumption 3.4 (SD1)

$$\begin{aligned} \Lambda_1 &= 2dL_d\big(1 - F_{\mathbb{M}}(u_d)\big)^2 = 2dL_d u_d(x)^{-2} \exp\big(-u_d(x)^2/q(T)\big)\, O(1) \\ &\leq \frac{2}{c_u^2} \frac{L_d}{\log d} d^{1-c_u^2/q(T)} O(1) = o(1) \quad \text{as } d \to \infty\ . \end{aligned} \tag{A.11}$$

Before we derive the asymptotic properties of Λ_2, first note that a comparison of the covariance structures of the two Gaussian processes yields

$$\Big\{\big(W_i(t), W_j(t)\big)\Big\}_{t\geq 0} \overset{\mathcal{D}}{=} \Big\{\ \Big(W_i(t), \sqrt{1-\rho_{i,j}^2}W_j'(t) + \rho_{i,j}W_i(t)\Big)\ \Big\}_{t\geq 0}\ ,$$

where W_j' is a standard Wiener process that is independent of W_i. Consequently, it also holds that

$$(M_i, M_j) \overset{\mathcal{D}}{=} \left(M_i, \sup_{0\leq t\leq q(T)} \sup_{0\leq s\leq t} \left|\sqrt{1-\rho_{i,j}^2}(W_j'(t) - W_j'(s)) + \rho_{i,j}(W_i(t) - W_i(s)\right|\right)$$

and by the triangle inequality

$$\begin{aligned} &\sup_{0\leq t\leq q(T)} \sup_{0\leq s\leq t} \left|\sqrt{1-\rho_{i,j}^2}(W_j'(t) - W_j'(s)) + \rho_{i,j}(W_i(t) - W_i(s)\right| \\ \leq &\sup_{0\leq t\leq q(T)} \sup_{0\leq s\leq t} \left|\sqrt{1-\rho_{i,j}^2}(W_j'(t) - W_j'(s))\right| + |\rho_{i,j}| \sup_{0\leq t\leq q(T)} \sup_{0\leq s\leq t} |(W_i(t) - W_i(s)| \\ = &\sqrt{1-\rho_{i,j}^2}M_j' + |\rho_{i,j}|M_i, \end{aligned}$$

where M_j' has the same distribution as M_j but is independent of M_i. Now, we obtain for $\eta > 1$ analogously to (B.35) in Jirak (2015)

$$\begin{aligned} &\mathbb{P}\big(M_i > u_d(x), M_j > u_d(x)\big) \leq \mathbb{P}\left(M_i > u_d(x), \sqrt{1-\rho_{i,j}^2}M_j' + |\rho_{i,j}|M_i > u_d(x)\right) \\ = &\int_{u_d(x)}^{\infty} \mathbb{P}\left(M_j' \geq \frac{u_d(x) - y|\rho_{i,j}|}{\sqrt{1-\rho_{i,j}^2}}\right) \mathbb{P}_{M_i}(dy) \\ \leq &\int_{u_d(x)}^{\eta u_d(x)} \mathbb{P}\left(M_j' \geq \frac{u_d(x) - y|\rho_{i,j}|}{\sqrt{1-\rho_{i,j}^2}}\right) \mathbb{P}_{M_i}(dy) + \mathbb{P}\big(M_i \geq \eta u_d(x)\big) \\ \leq &\,\mathbb{P}\left(M_j' > u_d(x)\frac{1-\eta|\rho_{i,j}|}{\sqrt{1-\rho_{i,j}^2}}\right) \mathbb{P}\big(M_i \geq u_d(x)\big) + \mathbb{P}\big(M_i \geq \eta u_d(x)\big) \end{aligned}$$

$$= \left(1 - F_{\mathbb{M}}\left(u_d(x)\frac{1-\eta|\rho_{i,j}|}{\sqrt{1-\rho_{i,j}^2}}\right)\right)\big(1 - F_{\mathbb{M}}(u_d(x)\big) + \big(1 - F_{\mathbb{M}}(\eta u_d(x)\big)$$

Applying the identity in (A.4), the last display equals

$$\left[u_d(x)^{-2}\frac{\sqrt{1-\rho_{i,j}^2}}{1-\eta|\rho_{i,j}|}\exp\left(-u_d(x)^2\left(\frac{1}{2q(T)} + \frac{(1-\eta|\rho_{i,j}|)^2}{2q(T)(1-\rho_{i,j}^2)}\right)\right)\right]O(1)$$

$$+\left[\eta^{-1}u_d(x)^{-1}\exp\left(-u_d(x)^2\frac{\eta^2}{2q(T)}\right)\right]O(1) \tag{A.12}$$

$$\le \left(\frac{1}{c_u^2\log d}\frac{\sqrt{1-\rho_{i,j}^2}}{1-\eta|\rho_{i,j}|}d^{-\frac{c_u^2}{2q(T)}\left(1+\frac{(1-\eta|\rho_{i,j}|)^2}{1-\rho_{i,j}^2}\right)} + \eta^{-1}\frac{1}{c_u\sqrt{\log d}}d^{-\frac{c_u^2\eta^2}{2q(T)}}\right)O(1)\ ,$$

where we used the estimate (A.10) in the last inequality. By Assumption 3.4 (SD3), we have $|\rho_{i,j}| \le \rho_+ < 1$. In the following we will show that we can choose the constants $c_u \in \big(\sqrt{q(T)}, \sqrt{2q(T)}\big)$ and $\eta > 1$ and such that

$$1 < \sqrt{\frac{2q(T)}{c_u^2}} < \eta < \frac{1-\sqrt{\left(\frac{2q(T)}{c_u^2}-1\right)(1-\rho_+^2)}}{\rho_+} < \frac{1}{\rho_+}\ . \tag{A.13}$$

As the function $g(y) = \frac{1-\sqrt{\left(\frac{2q(T)}{y^2}-1\right)(1-\rho_+^2)}}{(\rho_+)\cdot\sqrt{\frac{2q(T)}{y^2}}}$, is continuous on $\mathbb{R}_+$ and $g\left(\sqrt{2q(T)}\right) = \frac{1}{\rho_+} > 1$, $c_u \in \big(\sqrt{q(T)}, \sqrt{2q(T)}\big)$ can be chosen, such that $g(c_u) > 1$ and thus

$$\sqrt{\frac{2q(T)}{c_u^2}} < \frac{1-\sqrt{\left(\frac{2q(T)}{c_u^2}-1\right)(1-\rho_+^2)}}{\rho_+}\ .$$

Hence, $\eta > 1$ can be chosen as in (A.13). For such η consider the function $g_\eta(\rho) = \frac{(1-\eta\rho)^2}{1-\rho^2}$ for $\rho \in [0, \rho_+)$. As

$$\frac{\partial}{\partial\rho}g_\eta(\rho) = \frac{2(1-\eta\rho)(\rho-\eta)}{(1-\rho^2)^2} < 0$$

g is decreasing in ρ. Hence, it follows with (A.12), (A.13) and Assumption 3.4 (SD1)

$$\Lambda_2 \le\ 2dL_d\left(\frac{1}{c_u^2\log d}\frac{\sqrt{1-\rho_+^2}}{1-\eta\rho_+}d^{-\frac{c_u^2}{2q(T)}\left(1+\frac{(1-\eta\rho_+)^2}{1-\rho_+^2}\right)} + \eta^{-1}\frac{1}{c_u\sqrt{\log d}}d^{-\frac{c_u^2\eta^2}{2q(T)}}\right)O(1)$$

$$= \left(\frac{L_d}{c_u^2\log d}\frac{\sqrt{1-\rho_+^2}}{1-\eta\rho_+}d^{1-\frac{c_u^2}{2q(T)}\left(1+\frac{(1-\eta\rho_+)^2}{1-\rho_+^2}\right)} + \eta^{-1}\frac{L_d}{c_u\sqrt{\log d}}d^{1-\frac{c_u^2\eta^2}{2q(T)}}\right)O(1) = o(1)\ . \tag{A.14}$$

Due to $\rho_{i,j} = 0$ for $|i - j| > L_d$ we obtain that the Gaussian processes W_i and W_j are already independent whenever $|i - j| > L_d$ [see for instance Billingsley (1999)] and therefore we have that $\Lambda_3 = 0$. The assertion now follows by combining this fact with (A.7), (A.9), (A.11) and (A.14).

A.3 Proof of Theorem 3.6

Recall that the detector (2.10) is based on differences of component-wise mean estimators

$$\widehat{\mu}_{m+j+1}^{m+k}(h) - \widehat{\mu}_1^{m+j}(h)$$

and we may without loss of generality assume $\mathbb{E}[X_{t,h}] = 0$ throughout the proof. First, we introduce some necessary notations. Analogously to Theorem 3.5 let $\{W_h\}_{h\in\mathbb{N}}$ denote a sequence of Brownian motions on the interval $[0, q(T)]$ with correlations

$$\text{(A.15)} \qquad \operatorname{Corr}\big(W_h(t), W_i(t)\big) = \tilde{\rho}_{h,i} := \rho_{h,i} \cdot I\{|h - i| \le L_d\} ,$$

where $\rho_{h,i}$ denotes the long-run correlation defined in (3.5) and L_d is the sequence introduced in Assumption 3.4. Denote again by

$$\text{(A.16)} \qquad M_h := \max_{t\in[0,q(T)]} \max_{s\in[0,t]} \big|W_h(s) - W_h(t)\big| = \max_{t\in[0,q(T)]} W_h(t) - \min_{t\in[0,q(T)]} W_h(t)$$

the maximal increment of W_h. For $0 < c < q(T)$ define additionally the truncated version

$$\text{(A.17)} \qquad M_h(c) = \max_{t\in[q(c),q(T)]} \max_{s\in[0,\, q(q^{-1}(t)-c)]} \big|W_h(s) - W_h(t)\big|$$

where $q(x) = x/(x+1)$, $q^{-1}(x) = x/(1-x)$ and consider the overall maxima of these quantities by

$$\text{(A.18)} \qquad \mathcal{W}_d = \max_{h=1}^{d} M_h \ , \quad \mathcal{W}_d(c) = \max_{h=1}^{d} M_h(c) \ .$$

Recalling the definition of the Gaussian statistic $\widehat{\mathcal{T}}_{m,d}^{(Z)}$ in (3.21) based on the random variables $\{\widehat{Z}_{t,h}\}_{t=1,\dots,m+mT}^{h=1,\dots,d}$ we introduce two additional sets of Gaussian random variables $\{Z_{t,h}\}_{t=1,\dots,m+mT}^{h=1,\dots,d}$ and $\{\widetilde{Z}_{t,h}\}_{t=1,\dots,m+mT}^{h=1,\dots,d}$ with the properties

$$\text{(A.19)} \qquad \begin{gathered} \mathbb{E}[Z_{t,h}] = \mathbb{E}[\widetilde{Z}_{t,h}] = 0 \ , \\ \operatorname{Var}(Z_{t,h}) = \operatorname{Var}(\widetilde{Z}_{t,h}) = 1 \ , \\ \operatorname{Corr}\big(Z_{t,h},\, Z_{t,i}\big) = \rho_{h,i} \ \text{ and } \ \operatorname{Corr}\big(\widetilde{Z}_{t,h},\, \widetilde{Z}_{t,i}\big) = \tilde{\rho}_{h,i} \ , \end{gathered}$$

where $\rho_{h,i}$ and $\tilde{\rho}_{h,i}$ are the long-run correlations and truncated long-run correlations defined in (3.5) and (A.15), respectively. Further we assume that both random sets are independent with respect to the time index t, such that

$$\operatorname{Corr}\big(Z_{t_1,h}, Z_{t_2,h}\big) = 0 = \operatorname{Corr}\big(\widetilde{Z}_{t_1,h}, \widetilde{Z}_{t_2,h}\big)$$

whenever $t_1 \neq t_2$. Next, we define analogues of the statistic $\widehat{\mathcal{T}}^{(Z)}_{m,d}$ in (3.21) by

$$\text{(A.20)} \qquad \begin{aligned} \mathcal{T}^{(Z)}_{m,d} &:= \max_{h=1}^{d} \max_{k=1}^{Tm} \max_{j=0}^{k-1} \frac{(k-j)w(k/m)}{\sqrt{m}} \left| z^{m+k}_{m+j+1}(h) - z^{m+j}_{1}(h) \right| , \\ \widetilde{\mathcal{T}}^{(Z)}_{m,d} &:= \max_{h=1}^{d} \max_{k=1}^{Tm} \max_{j=0}^{k-1} \frac{(k-j)w(k/m)}{\sqrt{m}} \left| \widetilde{z}^{m+k}_{m+j+1}(h) - \widetilde{z}^{m+j}_{1}(h) \right| , \end{aligned}$$

where

$$\text{(A.21)} \qquad z^{j}_{i}(h) := \frac{1}{j-i+1} \sum_{t=i}^{j} Z_{t,h} \quad \text{and} \quad \widetilde{z}^{j}_{i}(h) := \frac{1}{j-i+1} \sum_{t=i}^{j} \widetilde{Z}_{t,h} \ .$$

For a constant $0 < c < T$ such that $cm \in \mathbb{N}$ we will now consider truncated versions of the statistics $\mathcal{T}_{m,d}, \mathcal{T}^{(Z)}_{m,d}, \widehat{\mathcal{T}}^{(Z)}_{m,d}$ and $\widetilde{\mathcal{T}}^{(Z)}_{m,d}$ defined by

$$\text{(A.22)} \qquad \begin{aligned} \mathcal{T}_{m,d}(c) &:= \max_{h=1}^{d} \max_{k=cm+1}^{Tm} \max_{j=0}^{k-cm-1} \frac{(k-j)w(k/m)}{\sqrt{m}\sigma_h} \left| \widehat{\mu}^{m+k}_{m+j+1}(h) - \widehat{\mu}^{m+j}_{1}(h) \right| , \\ \mathcal{T}^{(Z)}_{m,d}(c) &:= \max_{h=1}^{d} \max_{k=cm+1}^{Tm} \max_{j=0}^{k-cm-1} \frac{(k-j)w(k/m)}{\sqrt{m}\sigma_h} \left| z^{m+k}_{m+j+1}(h) - z^{m+j}_{1}(h) \right| , \\ \widehat{\mathcal{T}}^{(Z)}_{m,d}(c) &:= \max_{h=1}^{d} \max_{k=cm+1}^{Tm} \max_{j=0}^{k-cm-1} \frac{(k-j)w(k/m)}{\sqrt{m}\sigma_h} \left| \widehat{z}^{m+k}_{m+j+1}(h) - \widehat{z}^{m+j}_{1}(h) \right| , \\ \widetilde{\mathcal{T}}^{(Z)}_{m,d}(c) &:= \max_{h=1}^{d} \max_{k=cm+1}^{Tm} \max_{j=0}^{k-cm-1} \frac{(k-j)w(k/m)}{\sqrt{m}\sigma_h} \left| \widetilde{z}^{m+k}_{m+j+1}(h) - \widetilde{z}^{m+j}_{1}(h) \right| . \end{aligned}$$

Finally, recall the definition

$$\text{(A.23)} \qquad u_d(x) = x/a_d + b_d \ , \quad x \in \mathbb{R} \ ,$$

with the sequences a_d and b_d given by (3.7) and note that by Lemma A.2, there exists a constant $0 < c < \sqrt{2q(T)}$ such that

$$\text{(A.24)} \qquad u_d(x) \geq c\sqrt{\log(d)}$$

for sufficiently large d (depending on x).

The proof of Theorem 3.6 is now split into the following five Lemmata. If these are proven, then the claim is a consequence of Theorem 3.5.

Lemma A.3 (Truncation) *For $t_0 > 0$ sufficiently small*

$$\left| \mathbb{P}\Big(\mathcal{T}_{m,d} \leq u_d(x) \Big) - \mathbb{P}\Big(\mathcal{T}_{m,d}(t_0) \leq u_d(x) \Big) \right| = o(1) \quad \textit{as } m, d \to \infty.$$

Lemma A.4 (Gaussian approximation)

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\Big(\mathcal{T}_{m,d}(t_0) \leq x \Big) - \mathbb{P}\Big(\mathcal{T}^{(Z)}_{m,d}(t_0) \leq x \Big) \right| = o(1) \quad \textit{as } m, d \to \infty.$$

Lemma A.5 (Relaxation of correlation structure)

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\Big(\widetilde{\mathcal{T}}^{(Z)}_{m,d}(t_0)\le x\Big)-\mathbb{P}\Big(\mathcal{T}^{(Z)}_{m,d}(t_0)\le x\Big)\right| = o(1) \quad \textit{as } m,d\to\infty.$$

Lemma A.6 (Discretization of limit process)

$$\left|\mathbb{P}\Big(\widetilde{\mathcal{T}}^{(Z)}_{m,d}(t_0)\le u_d(x)\Big)-\mathbb{P}\big(\mathcal{W}_d(t_0)\le u_d(x)\big)\right| = o(1) \quad \textit{as } m,d\to\infty.$$

Lemma A.7 (Removing truncation)

$$\left|\mathbb{P}\Big(\mathcal{W}_d(t_0)\le u_d(x)\big)-\mathbb{P}\big(\mathcal{W}_d\le u_d(x)\big)\right| = o(1) \quad \textit{as } m,d\to\infty.$$

Proof of Lemma A.3. First note that

$$\begin{aligned}\mathcal{T}_{m,d} &= \max_{h=1}^{d}\max_{k=1}^{Tm}\max_{j=0}^{k-1}\frac{(k-j)w(k/m)}{\sqrt{m}\sigma_h}\left|\widehat{\mu}^{m+k}_{m+j+1}(h)-\widehat{\mu}^{m+j}_{1}(h)\right| \\ &= \max\Bigg\{\mathcal{T}_{m,d}(t_0)\ ,\max_{h=1}^{d}\max_{k=t_0m+1}^{Tm}\max_{j=k-t_0m}^{k-1}\frac{(k-j)w(k/m)}{\sqrt{m}\sigma_h}\left|\widehat{\mu}^{m+k}_{m+j+1}(h)-\widehat{\mu}^{m+j}_{1}(h)\right|\ , \\ &\qquad \max_{h=1}^{d}\max_{k=1}^{t_0m}\max_{j=0}^{k-1}\frac{(k-j)w(k/m)}{\sqrt{m}\sigma_h}\left|\widehat{\mu}^{m+k}_{m+j+1}(h)-\widehat{\mu}^{m+j}_{1}(h)\right|\Bigg\}\ .\end{aligned}$$

Hence, we obtain

$$\left|\mathbb{P}\Big(\mathcal{T}_{m,d}(t_0)\le u_d(x)\Big)-\mathbb{P}\Big(\mathcal{T}_{m,d}\le u_d(x)\Big)\right|\le P_1(x)+P_2(x)\ , \tag{A.25}$$

where

$$\begin{aligned}P_1(x)&=\mathbb{P}\Big(\max_{h=1}^{d}\max_{k=t_0m+1}^{Tm}\max_{j=k-t_0m}^{k-1}\frac{(k-j)}{\sqrt{m}\sigma_h}\left|\widehat{\mu}^{m+k}_{m+j+1}(h)-\widehat{\mu}^{m+j}_{1}(h)\right|\ge u_d(x)\Big)\ ,\\ P_2(x)&=\mathbb{P}\Big(\max_{h=1}^{d}\max_{k=1}^{t_0m}\max_{j=0}^{k-1}\frac{(k-j)}{\sqrt{m}\sigma_h}\left|\widehat{\mu}^{m+k}_{m+j+1}(h)-\widehat{\mu}^{m+j}_{1}(h)\right|\ge u_d(x)\Big)\ .\end{aligned}$$

and we additionally used that $w(k/m)\le 1$. We will treat the summands on the right-hand side of the last display separately. For the term $P_1(x)$ note that

$$\begin{aligned}&\mathbb{P}\left(\max_{k=t_0m+1}^{Tm}\max_{j=k-t_0m}^{k-1}\frac{(k-j)}{\sqrt{m}\sigma_h}\left|\widehat{\mu}^{m+k}_{m+j+1}(h)-\widehat{\mu}^{m+j}_{1}(h)\right|\ge u_d(x)\right)\\ &\qquad\le\ \mathbb{P}\left(\max_{k=t_0m+1}^{Tm}\max_{j=k-t_0m}^{k-1}\frac{(k-j)}{\sqrt{m}\sigma_h}\left|\widehat{\mu}^{m+k}_{m+j+1}(h)\right|\ge\frac{u_d(x)}{2}\right)\\ &\qquad\quad+\mathbb{P}\left(\max_{k=t_0m+1}^{Tm}\max_{j=k-t_0m}^{k-1}\frac{(k-j)}{\sqrt{m}\sigma_h}\left|\widehat{\mu}^{m+j}_{1}(h)\right|\ge\frac{u_d(x)}{2}\right)\ .\end{aligned}\tag{A.26}$$

Using stationarity and Assumption 3.3 (TD2), we have

$$\mathbb{P}\Big(\max_{k=t_0m+1}^{Tm}\max_{j=k-t_0m}^{k-1}\frac{(k-j)}{\sqrt{m}\sigma_h}\Big|\widehat{\mu}_{m+j+1}^{m+k}(h)\Big|\geq\frac{u_d(x)}{2}\Big)$$
$$\leq\sum_{k=t_0m+1}^{Tm}\mathbb{P}\left(\max_{j=k-t_0m}^{k-1}\Big|\sum_{i=m+j+1}^{m+k}X_{i,h}\Big|\geq\frac{\sqrt{m}c_\sigma u_d(x)}{2}\right).$$

Observing (A.24) and Lemma B.1, we obtain the following bound for the last display, which holds uniformly for $1\leq h\leq d$

$$C_p\frac{Tt_0m^{2-p/2}}{c_\sigma^p u_d(x)^p}+C_pTm\exp\left(-c_p\frac{c_\sigma^2u_d(x)^2}{4t_0}\right)\lesssim\frac{m^{2-p/2}}{(\log(d))^{p/2}}+md^{-\tilde{C}_p/t_0}\ ,$$

where $\tilde{C}_p>0$ is a sufficiently small constant. The second summand on the right-hand side of (A.26) can be estimated similarly, that is

$$\mathbb{P}\Big(\max_{k=t_0m+1}^{Tm}\max_{j=k-t_0m}^{k-1}\frac{(k-j)}{\sqrt{m}\sigma_h}\Big|\widehat{\mu}_{1}^{m+j}(h)\Big|\geq\frac{u_d(x)}{2}\Big)\leq\mathbb{P}\left(\max_{j=1}^{Tm+m}\Big|\sum_{i=1}^{j}X_{i,h}\Big|\geq\frac{c_\sigma\sqrt{m}}{t_0}\frac{u_d(x)}{2}\right)$$
$$\lesssim C_p\frac{t_0^p(T+1)m^{1-p/2}}{c_\sigma^p u_d(x)^p}+C_p\exp\left(-c_p\frac{c_\sigma^2u_d(x)^2}{4t_0^2(T+1)}\right)\lesssim\frac{m^{1-p/2}}{(\log(d))^{p/2}}+d^{-\tilde{C}_p/t_0^2}\ ,$$

where $\tilde{C}_p>0$ is again a sufficiently small constant. Hence, we obtain by Assumption 3.1, (A.26) (observing $p>2D+4$) that

$$P_1(x)\lesssim\frac{dm^{2-p/2}}{(\log(d))^{p/2}}+md^{1-\frac{\tilde{C}_p}{t_0}}=\frac{m^{D+2-p/2}}{(\log(d))^{p/2}}+m^{1+D\left(1-\tilde{C}_p/t_0\right)}=o(1)\tag{A.27}$$

if $t_0>0$ is chosen sufficiently small. Analogously, we obtain for the second summand on the right-hand side of (A.25) with a possibly smaller constant $t_0>0$, that

$$P_2(x)=o(1)\ ,\tag{A.28}$$

where we have used the following two inequalities which are a consequence of Lemma B.1

$$\mathbb{P}\left(\max_{k=1}^{t_0m}\max_{j=0}^{k-1}\frac{(k-j)}{\sqrt{m}\sigma_h}\Big|\widehat{\mu}_{m+j+1}^{m+k}(h)\Big|\geq\frac{u_d(x)}{2}\right)\leq\sum_{k=1}^{t_0m}\mathbb{P}\left(\max_{j=0}^{k-1}\Big|\sum_{i=m+j+1}^{m+k}X_{i,h}\Big|\geq\sqrt{m}c_\sigma\frac{u_d(x)}{2}\right)$$
$$\leq\sum_{k=1}^{t_0m}\mathbb{P}\left(\max_{j=0}^{t_0m-1}\Big|\sum_{i=m+j+1}^{m+t_0m}X_{i,h}\Big|\geq\sqrt{m}c_\sigma\frac{u_d(x)}{2}\right)$$
$$\lesssim C_p\frac{t_0^2m^{2-p/2}}{c_\sigma^p(\log(d))^{p/2}}+C_pt_0m\exp\left(-c_p\frac{c_\sigma^2u_d(x)^2}{4t_0}\right)$$
$$\lesssim\frac{m^{2-p/2}}{(\log(d))^{p/2}}+md^{-\tilde{C}_p/t_0}\ ,$$

$$\mathbb{P}\Big(\max_{k=1}^{t_0m}\max_{j=0}^{k-1}\frac{(k-j)}{\sqrt{m}\sigma_h}\Big|\widehat{\mu}_1^{m+j}(h)\Big|\geq\frac{u_d(x)}{2}\Big)\leq\ \mathbb{P}\Big(\max_{j=1}^{t_0m+m}\Big|\sum_{i=1}^{j}X_{i,h}\Big|\geq\frac{\sqrt{m}c_\sigma}{t_0}\frac{u_d(x)}{4}\Big)$$
$$\lesssim\ C_p\frac{(t_0+1)t_0^pm^{1-p/2}}{c_\sigma^p(\log(d))^{p/2}}+C_p\exp\left(-c_p\frac{c_\sigma^2u_d(x)^2}{4(t_0+1)t_0^2(T+1)}\right)$$
$$\lesssim\ \frac{m^{1-p/2}}{(\log(d))^{p/2}}+d^{-\frac{\tilde{C}_p}{(t_0+1)t_0^2}}\ .$$

Combining (A.27) and (A.28) the assertion of Lemma A.3 now follows from (A.25).

Proof of Lemma A.4. We will use a Gaussian Approximation provided in Corollary 2.2 of Zhang and Cheng (2018). For this purpose we introduce the notation

$$v_{m,k,j,h}:=\frac{(k-j)w(k/m)}{\sigma_h\sqrt{m}}\Big(\widehat{\mu}_{m+j+1}^{m+k}(h)-\widehat{\mu}_1^{m+j}(h)\Big)\ ,$$

with $k=t_0m+1,\dots,Tm$; $j=0\dots,k-t_0m-1$ and $h=1,\dots,d$. We stack all these quantities together in one vector

$\boldsymbol{V}_+:=$

$$(v_{m,t_0m+1,0,1},v_{m,t_0m+2,0,1},v_{m,t_0m+2,1,1},\dots,v_{m,Tm,Tm-t_0m-1,1},v_{m,t_0m+1,0,2},\dots,v_{m,Tm,Tm-t_0m-1,d})^\top\ .$$

Next define the vector

$$\boldsymbol{V}=(V_1,V_2,\dots,V_{d_{\boldsymbol{V}}})^\top:=\big(\boldsymbol{V}_+^\top,-\boldsymbol{V}_+^\top\big)^\top$$

and denote its dimension by $d_{\boldsymbol{V}}$. Observe that by construction the identity

$$\max_{i=1}^{d_{\boldsymbol{V}}}V_i=\mathcal{T}_{m,d}(t_0)=\max_{h=1}^{d}\max_{k=mt_0+1}^{Tm}\max_{j=0}^{k-mt_0-1}\frac{(k-j)w(k/m)}{\sigma_h\sqrt{m}}\left|\widehat{\mu}_{m+j+1}^{m+k}(h)-\widehat{\mu}_1^{m+j}(h)\right|$$

holds, where we use the fact that $\boldsymbol{V}$ contains both, the positive and negative version of all random variables which appear in the maximum in the definition of the statistic $\mathcal{T}_{m,d}(t_0)$. Further note that the dimension of $\boldsymbol{V}$ is bounded by

$$d_{\boldsymbol{V}}\leq 2d(Tm)^2. \tag{A.29}$$

By the construction above each component V_i corresponds either to $v_{m,k,j,h}$ or to $-v_{m,k,j,h}$ for some combination k,j,h. Hence, it can be represented by

$$V_i=\frac{1}{\sqrt{m}}\sum_{t=1}^{m(T+1)}X_{t,i}^*$$

with

$$X_{t,i}^*=\begin{cases}\dfrac{a_{t,m,k,j}}{\sigma_h}X_{t,h} & \text{for }\ 1\leq i\leq d_{\boldsymbol{V}}/2\ ,\\[2ex] -\dfrac{a_{t,m,k,j}}{\sigma_h}X_{t,h} & \text{for }\ d_{\boldsymbol{V}}/2+1\leq i\leq d_{\boldsymbol{V}}\ ,\end{cases} \tag{A.30}$$

where the indices k, j, h correspond to i according to the construction of the vector $\boldsymbol{V}$ and the coefficients $a_{t,m,k,j}$ are given by

$$a_{t,m,k,j} = \begin{cases} 0 & \text{if} \quad t > m+k \ , \\ a^{(1)}_{m,k} := w(k/m) & \text{if} \quad m+j < t \leq m+k \ , \\ a^{(2)}_{m,k,j} := -\dfrac{(k-j)w(k/m)}{(m+j)} & \text{if} \quad t \leq m+j \ . \end{cases} \tag{A.31}$$

Using the fact $w(k/m) = 1/(1+k/m)$ and $1 \leq k \leq mT$, we obtain

$$\frac{1}{T+1} \leq w(k/m) = a^{(1)}_{m,k} \leq 1 \tag{A.32}$$

and as $t_0 m \leq k - j$ and $j \leq k \leq Tm$ it follows that

$$\frac{t_0}{(T+1)^2} \leq |a^{(2)}_{m,k,j}| = \left| \frac{(k-j)w(k/m)}{m+j} \right| \leq T \ , \tag{A.33}$$

which yields by definition of $a_{t,m,k,j}$ in (A.31) the upper bound

$$|a_{t,m,k,j}| \leq T_+ := \max\{T, 1\} \ . \tag{A.34}$$

Moreover, the temporal dependence structure of the $d_{\boldsymbol{V}}$-dimensional time series

$$\big\{(X^*_{t,1}, \ldots, X^*_{1,d_{\boldsymbol{V}}})^\top\big\}_{t\in\mathbb{Z}}$$

still satisfies the concept of physical dependence as

$$X^*_{t,i} = g^*_{t,i}(\varepsilon_t, \varepsilon_{t-1}, \ldots)$$

with

$$g^*_{m,t,i}(\varepsilon_t, \varepsilon_{t-1}, \ldots) := \begin{cases} \dfrac{a_{t,m,k,j}}{\sigma_h} g_h(\varepsilon_t, \varepsilon_{t-1}, \ldots) & \text{for} \ \ 1 \leq i \leq d_{\boldsymbol{V}}/2 \ , \\ -\dfrac{a_{t,m,k,j}}{\sigma_h} g_h(\varepsilon_t, \varepsilon_{t-1}, \ldots) & \text{for} \ \ d_{\boldsymbol{V}}/2 + 1 \leq i \leq d_{\boldsymbol{V}} \ , \end{cases} \tag{A.35}$$

where the indices k, j, h correspond to i according to the construction of the vector $\boldsymbol{V}$.
In the following let $(\dot{V}^{(z)}_1, \ldots, \dot{V}^{(z)}_{d_{\boldsymbol{V}}})^\top$ denote a Gaussian distributed vector, which has (exactly) the same covariance structure as $\boldsymbol{V}$. Next, recall the definition of the Gaussian random variables $\{Z_{t,h}\}^{h=1,\ldots,d}_{t=1,\ldots,m+mT}$ in (A.19) and let

$$Z^*_{t,i} = \begin{cases} a_{t,m,k,j} Z_{t,h} & \text{for} \ \ 1 \leq i \leq d_{\boldsymbol{V}}/2 \ , \\ -a_{t,m,k,j} Z_{t,h} & \text{for} \ \ d_{\boldsymbol{V}}/2 + 1 \leq i \leq d_{\boldsymbol{V}} \ . \end{cases}$$

Further define the vector $\boldsymbol{V}^{(z)} = (V_1^{(z)}, \dots, V_{d_{\boldsymbol{V}}}^{(z)})^\top$ by

$$V_i^{(z)} := \frac{1}{\sqrt{m}} \sum_{t=1}^{m(T+1)} Z_{t,i}^* \qquad i = 1, \dots, d_{\boldsymbol{V}} \ .$$

We now proceed as follows:

Step 1: Show that for a constant $C > 0$

$$\text{(A.36)} \qquad \sup_{x \in \mathbb{R}} \Big| \mathbb{P}\Big(\max_{i=1}^{d_{\boldsymbol{V}}} \dot{V}_i^{(z)} \le x \Big) - \mathbb{P}\Big(\max_{i=1}^{d_{\boldsymbol{V}}} V_i^{(z)} \le x \Big) \Big| \lesssim m^{-C} \ .$$

Step 2: Establish that for a constant $C > 0$

$$\text{(A.37)} \qquad \sup_{x \in \mathbb{R}} \Big| \mathbb{P}\Big(\max_{i=1}^{d_{\boldsymbol{V}}} V_i \le x \Big) - \mathbb{P}\Big(\max_{i=1}^{d_{\boldsymbol{V}}} \dot{V}_i^{(z)} \le x \Big) \Big| \lesssim m^{-C} \ .$$

If both steps have been proven, the claim of Lemma A.4 follows from the identity

$$\max_{i=1}^{d_{\boldsymbol{V}}} V_i^{(z)} = \mathcal{T}_{m,d}^{(Z)}(t_0) \ .$$

Proof of Step 1: As we aim to compare the maxima of the two Gaussian distributed vectors $\boldsymbol{V}^{(z)}$ and $\dot{\boldsymbol{V}}^{(z)}$ we will apply Lemma B.3. Therefore, we analyze the covariance structures of $\boldsymbol{V}^{(z)}$ and $\dot{\boldsymbol{V}}^{(z)}$ [or equivalently $\boldsymbol{V}$]. Let $i_1, i_2 \in \{1, \dots, d_{\boldsymbol{V}}/2\}$ with corresponding indices h_1, j_1, k_1 and h_2, j_2, k_2 according to equation (A.30). For the calculation we assume without loss of generality that $j_1 \le j_2$ and use the notation $k_{min} = \min\{k_1, k_2\}$, $k_{max} = \max\{k_1, k_2\}$ and $j_2 \wedge k_1 = \min\{j_2, k_1\}$. Further we use the convention $\sum_{i=j}^k \beta_i = 0$, whenever $k < j$. For the covariance of the components of the vector $\boldsymbol{V}^{(z)}$ note that temporal independence yields

$$\begin{aligned}
\operatorname{Cov}\Big(V_{i_1}^{(z)}, V_{i_2}^{(z)}\Big) &= \frac{1}{m} \sum_{t=1}^{m+k_{min}} \operatorname{Cov}\Big(a_{t,m,k_1,j_1} Z_{t,h_1}, a_{t,m,k_2,j_2} Z_{t,h_2}\Big) \\
&= \frac{1}{m} \sum_{t=1}^{m+j_1} \operatorname{Cov}\Big(a_{t,m,k_1,j_1} Z_{t,h_1}, a_{t,m,k_2,j_2} Z_{t,h_2}\Big) \\
&\qquad + \frac{1}{m} \sum_{t=m+j_1+1}^{m+(j_2 \wedge k_1)} \operatorname{Cov}\Big(a_{t,m,k_1,j_1} Z_{t,h_1}, a_{t,m,k_2,j_2} Z_{t,h_2}\Big) \\
&\qquad + \frac{1}{m} \sum_{t=m+(j_2 \wedge k_1)+1}^{m+k_{min}} \operatorname{Cov}\Big(a_{t,m,k_1,j_1} Z_{t,h_1}, a_{t,m,k_2,j_2} Z_{t,h_2}\Big) \ .
\end{aligned}$$

Using the definition in (A.31) and (A.19) we obtain

$$\text{(A.38)} \qquad \begin{aligned}
\operatorname{Cov}\Big(V_{i_1}^{(z)}, V_{i_2}^{(z)}\Big) &= \frac{a_{m,k_1,j_1}^{(2)} a_{m,k_2,j_2}^{(2)}}{m \sigma_{h_1} \sigma_{h_2}} (m + j_1) \gamma_{h_1,h_2} + \frac{a_{m,k_1}^{(1)} a_{m,k_2,j_2}^{(2)}}{m \sigma_{h_1} \sigma_{h_2}} \big((j_2 \wedge k_1) - j_1\big) \gamma_{h_1,h_2} \\
&\qquad + \frac{a_{m,k_1}^{(1)} a_{m,k_2}^{(1)}}{m \sigma_{h_1} \sigma_{h_2}} (k_{min} - j_2) \gamma_{h_1,h_2} I\{j_2 < k_{min}\} \ .
\end{aligned}$$

Similar calculations also yield

$$\mathrm{Var}\left(V_{i_1}^{(z)}\right) = \mathrm{Cov}\left(V_{i_1}^{(z)}, V_{i_1}^{(z)}\right) = \left(a_{m,k_1,j_1}^{(2)}\right)^2 \frac{(m+j_1)}{m} + \left(a_{m,k_1}^{(1)}\right)^2 \frac{k_1 - j_1}{m}$$

and from (A.32) and (A.33) it follows that

$$\frac{t_0^2}{(T+1)^4} + \frac{t_0}{(T+1)^2} \le \mathrm{Var}\left(V_{i_1}^{(z)}\right) \le T^3 + T \ . \tag{A.39}$$

By the same arguments we obtain for the covariance structure of the components of the vector $\dot{\boldsymbol{V}}^{(z)}$ [note that we cannot use temporal independence here]:

$$\begin{aligned} \mathrm{Cov}\left(\dot{V}_{i_1}^{(z)}, \dot{V}_{i_2}^{(z)}\right) &= \mathrm{Cov}\left(\frac{1}{\sqrt{m}}\sum_{t=1}^{m+k_1} X_{t,i_1}^*, \frac{1}{\sqrt{m}}\sum_{s=1}^{m+k_2} X_{t,i_2}^*\right) \\ &= \sum_{\ell=1}^{4} \mathrm{Cov}\left(S_\ell^{(1)}, S_i^{(2)}\right) + \sum_{\substack{\ell,j=1 \\ i\neq j}}^{4} \mathrm{Cov}\left(S_\ell^{(1)}, S_j^{(2)}\right) \ , \end{aligned} \tag{A.40}$$

where the terms involved in the right-hand side are defined for $u = 1, 2$ by

$$\begin{aligned} S_1^{(u)} &= \frac{1}{\sqrt{m}} \sum_{t=1}^{m+j_1} X_{t,i_u}^* \ , \quad S_2^{(u)} = \frac{1}{\sqrt{m}} \sum_{t=m+j_1+1}^{m+(j_2\wedge k_1)} X_{t,i_u}^* \ , \\ S_3^{(u)} &= \frac{1}{\sqrt{m}} \sum_{t=m+(j_2\wedge k_1)+1}^{m+k_{min}} X_{t,i_u}^* \ , \quad S_4^{(u)} = \frac{1}{\sqrt{m}} \sum_{t=m+k_{min}+1}^{m+k_{max}} X_{t,i_u}^* \ . \end{aligned}$$

We will now treat the two summands on the right-hand side of (A.40) separately and show that the first summand is close to $\mathrm{Cov}\left(V_{i_1}^{(z)}, V_{i_2}^{(z)}\right)$, while the second vanishes sufficiently fast. Using that by construction, either $S_4^{(1)} = 0$ or $S_4^{(2)} = 0$, we obtain that

$$\begin{aligned} \sum_{\ell=1}^{4} \mathrm{Cov}\left(S_\ell^{(1)}, S_\ell^{(2)}\right) &= \frac{a_{m,k_1,j_1}^{(2)} a_{m,k_2,j_2}^{(2)}}{m\sigma_{h_1}\sigma_{h_2}} \sum_{t=1}^{m+j_1}\sum_{s=1}^{m+j_1} \mathrm{Cov}\left(X_{t,h_1}, X_{s,h_2}\right) \\ &\quad + \frac{a_{m,k_2}^{(1)} a_{m,k_2,j_2}^{(2)}}{m\sigma_{h_1}\sigma_{h_2}} \sum_{t=m+j_1+1}^{m+(j_2\wedge k_1)}\sum_{s=m+j_1+1}^{m+(j_2\wedge k_1)} \mathrm{Cov}\left(X_{t,h_1}, X_{s,h_2}\right) \\ &\quad + \frac{a_{m,k_1}^{(1)} a_{m,k_2}^{(1)}}{m\sigma_{h_1}\sigma_{h_2}} \sum_{t=m+(j_2\wedge k_1)+1}^{m+k_{min}}\sum_{s=m+(j_2\wedge k_1)+1}^{m+k_{min}} \mathrm{Cov}\left(X_{t,h_1}, X_{s,h_2}\right) \\ &= \frac{a_{m,k_1,j_1}^{(2)} a_{m,k_2,j_2}^{(2)}}{m\sigma_{h_1}\sigma_{h_2}} \sum_{t=-m-j_1}^{m+j_1} \left(m + j_1 - |t|\right)\phi_{t,h_1,h_2} \\ &\quad + \frac{a_{m,k_2}^{(1)} a_{m,k_2,j_2}^{(2)}}{m\sigma_{h_1}\sigma_{h_2}} \sum_{t=-(j_2\wedge k_1)+j_1}^{(j_2\wedge k_1)-j_1} \left((j_2\wedge k_1) - j_1 - |t|\right)\phi_{t,h_1,h_2} \end{aligned}$$

$$+\frac{a^{(1)}_{m,k_1}a^{(1)}_{m,k_2}}{m\sigma_{h_1}\sigma_{h_2}} I\big\{k_{min} > j_2 \wedge k_1\big\} \sum_{t=-k_{min}+(j_2\wedge k_1)}^{k_{min}-(j_2\wedge k_1)} \big(k_{min} - (j_2 \wedge k_1) - |t|\big)\phi_{t,h_1,h_2} \ .$$

Using the bounds in (A.32), (A.33), (A.38) and Assumption 3.3 (TD2) it follows that

$$\begin{aligned}
&\left|\operatorname{Cov}\left(V^{(z)}_{i_1},\, V^{(z)}_{i_2}\right) - \sum_{\ell=1}^{4} \operatorname{Cov}\left(S^{(1)}_\ell, S^{(2)}_\ell\right)\right| \\
&\quad \le \frac{C_{T,t_0}}{c_\sigma^2 m}\Bigg[\sum_{t\in\mathbb{Z}} \min\{|t|, m+j_1\}|\phi_{t,h_1,h_2}| + \sum_{t\in\mathbb{Z}} \min\{|t|,\, (j_2\wedge k_1) - j_1\}|\phi_{t,h_1,h_2}| \\
&\qquad\qquad + \sum_{t\in\mathbb{Z}} \min\{|t|,\, k_{min} - (j_2\wedge k_1)\}|\phi_{t,h_1,h_2}|\Bigg] \\
&\quad \le \frac{3C_{T,t_0}}{c_\sigma^2 m} \sum_{t\in\mathbb{Z}} |t||\phi_{t,h_1,h_2}| \ ,
\end{aligned} \tag{A.41}$$

where the constant C_{T,t_0} depends on T and t_0 only and we used the definition of γ_{h_1,h_2} in (3.4). Using Assumption 3.3 (TD1) and Lemma E4 from Jirak (2015) it follows that

$$\sup_{h_1,h_2\in\mathbb{N}} \sum_{t\in\mathbb{Z}} |t||\phi_{t,h_1,h_2}| < \infty \ , \tag{A.42}$$

which yields

$$\left|\operatorname{Cov}\left(V^{(z)}_{i_1},\, V^{(z)}_{i_2}\right) - \sum_{\ell=1}^{4} \operatorname{Cov}\left(S^{(1)}_\ell, S^{(2)}_\ell\right)\right| \lesssim \frac{1}{m} \ , \tag{A.43}$$

where the involved constant is independent of i_1 and i_2 [or equivalently j_1, j_2, k_1, k_2, h_1 and h_2]. Next, we treat the second sum on the right-hand of (A.40). For that purpose, note that for arbitrary points in time $p_1 < p_2 < p_3 < p_4$, it holds that

$$\begin{aligned}
\left|\operatorname{Cov}\Big(\sum_{t=p_1}^{p_2} X_{t,h_1},\, \sum_{s=p_3}^{p_4} X_{s,h_2}\Big)\right| &\le \sum_{t=1}^{p_2}\sum_{s=p_2+1}^{p_4} \left|\operatorname{Cov}\left(X_{t,h_1},\, X_{s,h_2}\right)\right| \\
&= \sum_{t=1}^{p_2}\sum_{s=p_2+1}^{p_4} |\phi_{s-t,h_1,h_2}| = \sum_{t=1}^{p_2}\sum_{s=p_2-t+1}^{p_4-t} |\phi_{s,h_1,h_2}| \\
&\le \sum_{t=1}^{p_2}\sum_{s=p_2-t+1}^{p_4} |\phi_{s,h_1,h_2}| = \sum_{s=1}^{p_4}\sum_{t=p_2-s+1}^{p_2} |\phi_{s,h_1,h_2}| = \sum_{s=1}^{p_4} s|\phi_{s,h_1,h_2}| \ .
\end{aligned} \tag{A.44}$$

Using the upper bound for the coefficients $a_{t,m,k,j}$ in (A.34), the uniform bound in (A.42) and that all the pairs of the sums under consideration are non-overlapping as treated above in (A.44), we obtain directly that

$$\sum_{\substack{i,j=1\\ i\neq j}}^{4} \operatorname{Cov}\left(S^{(1)}_i, S^{(2)}_j\right) \lesssim \frac{1}{m} \ , \tag{A.45}$$

where the constant is again independent of i_1, i_2. Combining the estimates (A.43) and (A.45), we conclude

$$\Delta_m := \max_{i_1,i_2=1}^{d_{\boldsymbol{V}}} \left| \operatorname{Cov}\left(V_{i_1}^{(z)}, V_{i_2}^{(z)}\right) - \operatorname{Cov}\left(\dot{V}_{i_1}^{(z)}, \dot{V}_{i_2}^{(z)}\right) \right| \lesssim \frac{1}{m} \,. \tag{A.46}$$

Due to (A.39) we can now apply Lemma B.3, which gives

$$\begin{aligned}\sup_{x\in\mathbb{R}} \Big| \mathbb{P}\Big(\max_{i=1}^{d_{\boldsymbol{V}}} V_i^{(z)} \le x \Big) - \mathbb{P}\Big(\max_{i=1}^{d_{\boldsymbol{V}}} \dot{V}_i^{(z)} \le x \Big) \Big| & \\ \lesssim \Delta_m^{1/3} \cdot \max\Big\{1,\, \log\big(d_{\boldsymbol{V}}/\Delta_m\big)\Big\}^{2/3} & \\ \lesssim \max\Big\{\Delta_m^{1/2},\, \Delta_m^{1/2}\big|\log d_{\boldsymbol{V}}\big| + \Delta_m^{1/2}\big|\log \Delta_m\big|\Big\}^{2/3} & \,.\end{aligned}$$

Using (A.29), Assumption 3.1 the assertion of Step 1 follows.

Proof of Step 2: Corollary 2.2 of Zhang and Cheng (2018) yields the Gaussian approximation in (A.37) if there exist positive constants c_1, c_2 and C and a deterministic sequence $B_m^* \ge 1$, such that the following four inequalities hold uniformly in t and i (or equivalently in t, k, j, h).

(i) $d_{\boldsymbol{V}} \lesssim \exp((Tm)^b)$ for some $0 \le b < 1/11$,

(ii) With β as in Assumption 3.3 (TD1) it holds

$$\sum_{\ell=u}^{\infty} \sup_{t\in\mathbb{Z}} \big\| g_{t,i}^*(\varepsilon_t, \varepsilon_{t-1}, \dots) - g_{t,i}^*(\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-\ell+1}, \varepsilon_{t-\ell}', \varepsilon_{t-\ell-1}, \dots) \big\|_p \lesssim \beta^u \ ,$$

where ε_{t-l}' is an independent copy of $\varepsilon_{t-\ell}$,

(iii) $c_1 \le \operatorname{Var}\big(V_i\big) \le c_2$,

(iv) $\mathbb{E}\Big[\exp\big(|X_{t,i}^*|/B_m^*\big)\Big] \le C, \quad B_m^* \lesssim m^{(3-17b)/8}$.

Therefore the proof of Lemma A.4 is completed by establishing these conditions.

Proof of (i): By (A.29) and Assumption 3.1 (D1) the inequality $d_{\boldsymbol{V}} \lesssim \exp((Tm)^b)$ holds for any $b > 0$.

Proof of (ii): With (A.35), (A.34), Assumption 3.3 (TD1) and the stationarity of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ we obtain

$$\begin{aligned}&\sum_{\ell=u}^{\infty} \sup_{t\in\mathbb{Z}} \big\| g_{m,t,i}^*(\varepsilon_t, \varepsilon_{t-1}, \dots) - g_{m,t,i}^*(\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-\ell+1}, \varepsilon_{t-\ell}', \varepsilon_{t-\ell-1}, \dots) \big\|_p \\ &\quad = \sum_{\ell=u}^{\infty} \sup_{t\in\mathbb{Z}} \frac{|a_{t,m,k,j}|}{\sigma_h} \big\| g_h(\varepsilon_\ell, \varepsilon_{\ell-1}, \dots) - g_h(\varepsilon_\ell, \varepsilon_{\ell-1}, \dots, \varepsilon_1, \varepsilon_0', \varepsilon_{-1}, \dots) \big\|_p\end{aligned}$$

$$\leq \frac{T_+}{c_\sigma} \sum_{\ell=u}^{\infty} \vartheta_{\ell,h,p} \leq \frac{T_+ C_\vartheta}{c_\sigma} \sum_{\ell=u}^{\infty} \beta^\ell \lesssim \beta^u$$

Proof of (iii): The assertion follows by combining (A.39) and (A.46).

Proof of (iv): Due to Assumption 3.2 and the upper bound on $|a_{t,m,k,j}|$ in (A.34) we obtain that

$$|X_{t,i}^*| \overset{\mathcal{D}}{=} \frac{|a_{t,m,k,j}|}{\sigma_h} |X_{1,h}| \leq \frac{T_+}{c_\sigma} |X_{1,h}|. \tag{A.47}$$

Hence, it follows from Assumption 3.2 (S1) observing the inequalities $B_m^* = \dfrac{T_+}{c_\sigma} B_m \lesssim m^B$ and $B < 3/8$ that

$$\begin{aligned} \mathbb{E}\Big[\exp\big(|X_{t,i}^*|/B_m^*\big)\Big] &= \mathbb{E}\Big[\exp\big(|a_{t,m,k,j}||X_{t,h}|/(\sigma_h B_m^*)\big)\Big] \\ &\leq \mathbb{E}\Big[\exp\big(|X_{1,h}|/B_m\big)\Big] \leq C \ . \end{aligned}$$

As (i) holds for any $b > 0$, we can choose b to be sufficiently small such that
$B < (3 - 17b)/8 < 3/8$. □

Proof of Lemma A.5. Recall the definition of the Gaussian vector $\boldsymbol{V}^{(z)} = \big(V_1^{(z)}, \dots, V_{d_{\boldsymbol{V}}}^{(z)}\big)^\top$ from the proof of Lemma A.4, which fulfills the identity

$$\mathcal{T}_{m,d}^{(Z)}(t_0) = \max_{i=1}^{d_{\boldsymbol{V}}} V_i^{(z)} \ .$$

Applying again the vectorization technique as introduced in the proof of Lemma A.4, we can define analogously a Gaussian vector for the statistic $\widetilde{\mathcal{T}}_{m,d}^{(Z)}$ in (A.22). Recall the definition of $\tilde{z}_i^j$ in (A.21) and introduce the notation

$$\tilde{v}_{m,k,j,h} := \frac{(k-j)w(k/m)}{\sqrt{m}} \Big(\tilde{z}_{m+j+1}^{m+k}(h) - \tilde{z}_1^{m+j}(h) \Big) \ , \tag{A.48}$$

with $k = t_0 m + 1, \dots, Tm$ and $j = 0 \dots, k - t_0 m - 1$ and $h = 1, \dots, d$. We stack all these quantities together in one vector, this is

$$\widetilde{\boldsymbol{V}}_+^{(z)} := (\tilde{v}_{m,t_0m+1,0,1}, \tilde{v}_{m,t_0m+2,0,1}, \tilde{v}_{m,t_0m+2,1,1}, \dots, \tilde{v}_{m,Tm,Tm-t_0m-1,1}, \tilde{v}_{m,t_0m+1,0,2}, \dots, \tilde{v}_{m,Tm,Tm-t_0m-1,d})^\top \ .$$

Next let $\widetilde{\boldsymbol{V}}^{(z)} = \Big(\big(\widetilde{\boldsymbol{V}}_+^{(z)}\big)^\top, -\big(\widetilde{\boldsymbol{V}}_+^{(z)}\big)^\top\Big)^\top$ with dimension $d_{\boldsymbol{V}}$ and denote its components by

$$\widetilde{\boldsymbol{V}}^{(z)} = (\widetilde{V}_1^{(z)}, \widetilde{V}_2^{(z)}, \dots, \widetilde{V}_{d_{\boldsymbol{V}}}^{(z)})^\top \ .$$

By construction of $\widetilde{\boldsymbol{V}}^{(z)}$ we have

$$\widetilde{\mathcal{T}}_{m,d}^{(Z)} = \max_{i=1}^{d_{\boldsymbol{V}}} \widetilde{V}_i^{(z)} .$$

The covariance structure of $\boldsymbol{V}^{(z)}$ was already calculated in (A.38) and is given by

$$\begin{aligned}\text{(A.49)}\qquad \operatorname{Cov}\left(V_{i_1}^{(z)}, V_{i_2}^{(z)}\right) =& \frac{a_{m,k_1,j_1}^{(2)} a_{m,k_2,j_2}^{(2)}}{m}(m+j_1)\rho_{h_1,h_2} + \frac{a_{m,k_1}^{(1)} a_{m,k_2,j_2}^{(2)}}{m}\big((j_2 \wedge k_1) - j_1\big)\rho_{h_1,h_2} \\ &+ \frac{a_{m,k_1}^{(1)} a_{m,k_2}^{(1)}}{m}\rho_{h_1,h_2}(k_{min} - j_2) I\{j_2 < k_{min}\} ,\end{aligned}$$

where k_1, j_1, h_1 and k_2, j_2, h_2 are the corresponding indices to i_1 and i_2, respectively and we use the notation $k_{min} = \min\{k_1, k_2\}$. A similar calculation for the vector $\widetilde{\boldsymbol{V}}^{(z)}$ gives

$$\begin{aligned}\text{(A.50)}\qquad \operatorname{Cov}\left(\widetilde{V}_{i_1}^{(z)}, \widetilde{V}_{i_2}^{(z)}\right) =& \frac{a_{m,k_1,j_1}^{(2)} a_{m,k_2,j_2}^{(2)}}{m}(m+j_1)\tilde{\rho}_{h_1,h_2} + \frac{a_{m,k_1}^{(1)} a_{m,k_2,j_2}^{(2)}}{m}\big((j_2 \wedge k_1) - j_1\big)\tilde{\rho}_{h_1,h_2} \\ &+ \frac{a_{m,k_1}^{(1)} a_{m,k_2}^{(1)}}{m}\tilde{\rho}_{h_1,h_2}(k_{min} - j_2) I\{j_2 < k_{min}\} .\end{aligned}$$

Note that by definition of the truncated correlations in (A.15) the quantities in (A.49) and (A.50) coincide, whenever $|h_1 - h_2| \leq L_d$. Therefore we obtain for the maximum distance of the covariance structures

$$\Delta_m := \max_{i_1,i_2=1}^{d_{\boldsymbol{V}}} \left| \operatorname{Cov}\left(V_{i_1}^{(z)}, V_{i_2}^{(z)}\right) - \operatorname{Cov}\left(\widetilde{V}_{i_1}^{(z)}, \widetilde{V}_{i_2}^{(z)}\right) \right| \leq C_T \sup_{h_1,h_2:|h_1-h_2|>L_d} |\rho_{h_1,h_2}| ,$$

where C_T is a constant depending on T only, as we used that $j_1, j_2, k_1, k_2 \leq mT$ and the upper bound in (A.34). Using Assumption 3.4 (SD2), it follows that $\Delta_m = o\big(\log^{-2}(d)\big)$. Due to (A.39), we can apply Lemma B.3, which gives

$$\begin{aligned}\sup_{x\in\mathbb{R}} \left| \mathbb{P}\Big(\max_{i=1}^{d_{\boldsymbol{V}}} V^{(z)} \leq x\Big) - \mathbb{P}\Big(\max_{i=1}^{d_{\boldsymbol{V}}} \widetilde{V}_i^{(z)} \leq x\Big) \right| &\lesssim \Delta_m^{1/3} \cdot \max\Big\{1, \log\big(d_{\boldsymbol{V}}/\Delta_m\big)\Big\}^{2/3} \\ &\leq \max\Big\{\Delta_m^{1/2}, \Delta_m^{1/2}\big|\log d_{\boldsymbol{V}}\big| + \Delta_m^{1/2}\big|\log \Delta_m\big|\Big\}^{2/3} .\end{aligned}$$

Observing (A.29), Assumption 3.1 (D1) the proof of Lemma A.5 is completed. □

Proof of Lemma A.6. We use similar arguments as given in the proof of Lemma B.7 of Jirak (2015). Let $\{W_h'\}_{h\in\mathbb{N}}$ denote an independent copy of the sequence of Brownian motions $\{W_h\}_{h\in\mathbb{N}}$ defined in (A.15). Recalling the notation (A.21) we obtain the representation

$$\text{(A.51)}\qquad \widetilde{\mathcal{T}}_{m,d}^{(Z)}(t_0) = \max_{h=1}^{d} \max_{k=t_0m+1}^{Tm} \max_{j=0}^{k-mt_0-1} \frac{1}{\sqrt{m}(1+k/m)} \left| \sum_{t=m+j+1}^{m+k} \widetilde{Z}_{t,h} - \frac{k-j}{m+j} \sum_{t=1}^{m+j} \widetilde{Z}_{t,h} \right| .$$

To investigate the quantities in the maximum we note that

(A.52)

$$\begin{aligned}
&\frac{1}{\sqrt{m}(1+k/m)}\left|\sum_{t=m+j+1}^{m+k}\widetilde{Z}_{t,h}-\frac{k-j}{m+j}\sum_{t=1}^{m+j}\widetilde{Z}_{t,h}\right|\\
&=\frac{1}{\sqrt{m}(1+k/m)}\left|\sum_{t=1}^{m+k}\widetilde{Z}_{t,h}-\frac{m+k}{m+j}\sum_{t=1}^{m+j}\widetilde{Z}_{t,h}\right|\\
&\overset{\mathcal{D}}{=}\frac{1}{1+k/m}\left|W_h(k/m+1)-\frac{m+k}{m+j}W_h(j/m+1)\right|\\
&=\frac{1}{1+k/m}\left|W_h(k/m+1)-W_h(1)-\frac{m+k}{m+j}\Big(W_h(j/m+1)-W_h(1)\Big)-\frac{k-j}{m+j}W_h(1)\right|\\
&\overset{\mathcal{D}}{=}\frac{1}{1+k/m}\left|W_h(k/m)-\frac{m+k}{m+j}W_h(j/m)-\frac{k-j}{m+j}W_h'(1)\right|\\
&=\frac{1}{(1+k/m)(1+j/m)}\bigg|(1+j/m)\Big\{W_h(k/m)-k/mW_h'(1)\Big\}\\
&\qquad\qquad-(1+k/m)\Big\{W_h(j/m)-j/mW_h'(1)\Big\}\bigg|
\end{aligned}$$

where in all steps the correlation structure of $\{W_h\}_{h\in\mathbb{N}}$ is preserved. A calculation of the covariance kernel implies the identity (in distribution)

$$\Big\{W_h(t)-tW_h'(1)\Big\}_{t\geq 0,\,h\in\mathbb{N}}\overset{\mathcal{D}}{=}\Big\{(1+t)W_h\Big(\frac{t}{t+1}\Big)\Big\}_{t\geq 0,\,h\in\mathbb{N}}.$$

Applying this to (A.52) yields

$$\begin{aligned}
&\frac{1}{(1+k/m)(1+j/m)}\bigg|(1+j/m)\Big\{W_h(k/m)-k/mW_h'(1)\Big\}\\
&\qquad\qquad-(1+k/m)\Big\{W_h(j/m)-j/mW_h'(1)\Big\}\bigg|\\
&\overset{\mathcal{D}}{=}\left|W_h\Big(\frac{k}{m+k}\Big)-W_h\Big(\frac{j}{m+j}\Big)\right|
\end{aligned}$$

This now gives

$$\begin{aligned}
\max_{k=t_0m+1}^{Tm}\max_{j=0}^{k-mt_0-1}&\frac{1}{\sqrt{m}(1+k/m)}\left|\sum_{t=m+j+1}^{m+k}\widetilde{Z}_{t,h}-\frac{k-j}{m+j}\sum_{t=1}^{m+j}\widetilde{Z}_{t,h}\right|\\
&\overset{\mathcal{D}}{=}\max_{k=t_0m+1}^{Tm}\max_{j=0}^{k-mt_0-1}\left|W_h\Big(\frac{k}{m+k}\Big)-W_h\Big(\frac{j}{m+j}\Big)\right|\\
&=M_{h,m}(t_0):=\max_{\substack{j,k\in\{1,\dots,Tm\}\\ k-j>mt_0}}\left|W_h\Big(\frac{k}{m+k}\Big)-W_h\Big(\frac{j}{m+j}\Big)\right|,
\end{aligned}$$

which is the discrete counterpart of the random variable $M_h(t_0)$ defined in (A.17). Observing the identity

$$\begin{aligned} M_h(t_0) &= \max_{t\in[q(t_0),\ q(T)]}\ \max_{s\in[0,\ q(q^{-1}(t)-t_0)]} \big|W_h(t)-W_h(s)\big| \\ &= \max_{t\in[q(t_0),\ q(T)]}\ \max_{s\in[0,\ q^{-1}(t)-t_0]} \big|W_h(t)-W_h(q(s))\big| \\ &= \max_{t\in[t_0,T]}\ \max_{s\in[0,t-t_0]} \Big|W_h\Big(\frac{t}{t+1}\Big)-W_h\Big(\frac{s}{s+1}\Big)\Big| \end{aligned}$$

the inequality $M_{h,m}(t_0)\le M_h(t_0)$ already yields

$$\begin{aligned} \mathbb{P}\Big(\widetilde{\mathcal{T}}^{(Z)}_{m,d}(t_0)\le u_d(x)\Big) &= \mathbb{P}\Big(\max_{h=1}^{d} M_{h,m}(t_0)\le u_d(x)\Big) \\ &\ge \mathbb{P}\Big(\max_{h=1}^{d} M_h(t_0)\le u_d(x)\Big) = \mathbb{P}\Big(\mathcal{W}_d(t_0)\le u_d(x)\Big) \end{aligned}$$

for all $x\in\mathbb{R}$. So it remains to find a suitable upper bound for

$$\mathbb{P}\Big(\widetilde{\mathcal{T}}^{(Z)}_{m,d}(t_0)\le u_d(x)\Big)-\mathbb{P}\Big(\mathcal{W}_d(t_0)\le u_d(x)\Big)\ge 0\ . \tag{A.53}$$

Observing the inequality (which holds for all $y\in\mathbb{R}$)

$$\begin{aligned} \mathbb{P}\Big(W_d(t_0)\le u_d(x)\Big) &\ge \mathbb{P}\Big(\widetilde{\mathcal{T}}^{(Z)}_{m,d}(t_0)\le u_d(x)-y, \big|\widetilde{\mathcal{T}}^{(Z)}_{m,d}(t_0)-\mathcal{W}_d(t_0)\big|<y\Big) \\ &\ge \mathbb{P}\Big(\widetilde{\mathcal{T}}^{(Z)}_{m,d}(t_0)\le u_d(x)-y\Big)-\mathbb{P}\Big(\big|\widetilde{\mathcal{T}}^{(Z)}_{m,d}(t_0)-\mathcal{W}_d(t_0)\big|>y\Big) \end{aligned}$$

the left-hand side in (A.53) is bounded by

$$\mathbb{P}\Big(\widetilde{\mathcal{T}}^{(Z)}_{m,d}(t_0)\le u_d(x)\Big)-\mathbb{P}\Big(\widetilde{\mathcal{T}}^{(Z)}_{m,d}(t_0)\le u_d(x)-y\Big)+\mathbb{P}\Big(\big|\widetilde{\mathcal{T}}^{(Z)}_{m,d}(t_0)-\mathcal{W}_d(t_0)\big|>y\Big)\ .$$

We now choose $y_d=d^{-1/(4D)}$, where D is the constant in Assumption 3.1. Then the claim is a consequence of the following two assertions:

(i) $\mathbb{P}\Big(\big|\widetilde{\mathcal{T}}^{(Z)}_{m,d}(t_0)-\mathcal{W}_d(t_0)\big|>y_d\Big)=o(1)$,

(ii) $\mathbb{P}\Big(\widetilde{\mathcal{T}}^{(Z)}_{m,d}(t_0)\le u_d(x)\Big)-\mathbb{P}\Big(\widetilde{\mathcal{T}}^{(Z)}_{m,d}(t_0)\le u_d(x)-y_d\Big)=o(1)$,

which will be proven below to complete the proof of Lemma A.6. To show (i), note that due to the time reversal properties and scaling properties of Brownian motions, it holds for all $k\le Tm$, $1\le h\le d$

$$\max_{t\in[(k-1)/m,\ k/m]}\Big|W_h\Big(\frac{t}{t+1}\Big)-W_h\Big(\frac{k/m}{1+k/m}\Big)\Big|$$

$$\stackrel{\mathcal{D}}{=} \max_{\lambda\in\left[0,\ \frac{k}{m+k}-\frac{k-1}{m+k-1}\right]} \left|W_h\Big(\frac{k/m}{1+k/m}-\lambda\Big) - W_h\Big(\frac{k/m}{1+k/m}\Big)\right|$$
$$\stackrel{\mathcal{D}}{=} \max_{\lambda\in[0,\ \frac{k}{m+k}-\frac{k-1}{m+k-1}]} \left|W_h(\lambda)\right| \leq \max_{0\leq\lambda\leq 1/m} \left|W_h(\lambda)\right| \stackrel{\mathcal{D}}{=} \max_{0\leq\lambda\leq 1} \left|W_h(\lambda)\right|/\sqrt{m}\ ,$$

which yields

$$\mathbb{P}\Big(\big|\widetilde{\mathcal{T}}^{(Z)}_{m,d}(t_0) - \mathcal{W}_d(t_0)\big| > y_d\Big) = \mathbb{P}\bigg(\max_{h=1}^{d} M_h(t_0) - \max_{h=1}^{d} M_{h,m}(t_0) > y_d\bigg)$$
$$=\mathbb{P}\bigg(\max_{h=1}^{d}\max_{t\in[t_0,T]}\max_{s\in[0,t-t_0]}\left|W_h\Big(\frac{t}{t+1}\Big) - W_h\Big(\frac{s}{s+1}\Big)\right|$$
$$- \max_{h=1}^{d}\max_{k=t_0m+1}^{Tm}\max_{j=0}^{k-mt_0-1}\left|W_h\Big(\frac{k}{m+k}\Big) - W_h\Big(\frac{j}{m+j}\Big)\right| > y_d\bigg)$$
$$\leq\mathbb{P}\bigg(\max_{h=1}^{d}\max_{k=t_0m+1}^{Tm}\max_{j=0}^{k-mt_0-1}\max_{t\in[(k-1)/m,\,k/m]}\max_{s\in[j/m,\,(j+1)/m]}\left|W_h\Big(\frac{t}{t+1}\Big) - W_h\Big(\frac{s}{s+1}\Big)\right|$$
$$- \max_{h=1}^{d}\max_{k=t_0m+1}^{Tm}\max_{j=0}^{k-mt_0-1}\left|W_h\Big(\frac{k/m}{1+k/m}\Big) - W_h\Big(\frac{j/m}{1+j/m}\Big)\right| > y_d\bigg)$$
$$\leq\mathbb{P}\bigg(\max_{h=1}^{d}\max_{k=t_0m+1}^{Tm}\max_{j=0}^{k-mt_0-1}\max_{t\in[(k-1)/m,\ k/m]}\max_{s\in[j/m,\ (j+1)/m]}\left|W_h\Big(\frac{t}{t+1}\Big) - W_h\Big(\frac{s}{s+1}\Big)\right.$$
$$\left. - W_h\Big(\frac{k/m}{1+k/m}\Big) - W_h\Big(\frac{j/m}{1+j/m}\Big)\right| > y_d\bigg)$$
$$\leq\mathbb{P}\bigg(2\max_{h=1}^{d}\max_{k=1}^{Tm}\max_{t\in[(k-1)/m,\ k/m]}\left|W_h\Big(\frac{t}{t+1}\Big) - W_h\Big(\frac{k/m}{1+k/m}\Big)\right| > y_d\bigg)$$
$$\leq\sum_{h=1}^{d}\sum_{k=1}^{Tm}\mathbb{P}\bigg(\sup_{0\leq\lambda\leq 1}|W_h(\lambda)| > y_d\sqrt{m}/2\bigg) \leq dTm\frac{4}{\sqrt{2\pi}y_d}e^{-y_d^2m/8}\ ,$$

where we have used the elementary bound [see for instance Karatzas and Shreve (1991)]

$$\mathbb{P}\bigg(\sup_{0\leq\lambda\leq 1}|W_h(\lambda)| \geq z\bigg) \leq \frac{4}{\sqrt{2\pi}z}e^{-z^2/2}\ .$$

This yields (i) since by Assumption 3.1 (D1) the choice of y_d gives $my_d^2 = C_D^{-1/(2D)}m^{1/2}$.

To obtain the estimate (ii), recall the definition of the Gaussian vector $\widetilde{\boldsymbol{V}}^{(z)} = \big(\widetilde{V}_1^{(z)}, \ldots, \widetilde{V}_{d_{\boldsymbol{V}}}^{(z)}\big)^\top$ in the proof of Lemma A.5, which yields the identity

$$\max_{i=1}^{d_{\boldsymbol{V}}}\widetilde{V}_i^{(z)} = \widetilde{\mathcal{T}}^{(Z)}_{m,d}(t_0)\ .$$

For each component $\widetilde{V}_i^{(z)}$ of $\widetilde{\boldsymbol{V}}^{(z)}$ there are indices k, j, h such that

$$\widetilde{V}_i^{(z)} = \tilde{v}_{m,k,j,h}\ ,$$

where $\tilde{v}_{m,k,j,h}$ is s defined in (A.48). Thus, we obtain the following bounds for the variance of the components of $\widetilde{\boldsymbol{V}}^{(z)}$:

$$\begin{aligned}\operatorname{Var}\Big(\tilde{v}_{m,k,j,h}\Big) &= \frac{\Big((k-j)w(k/m)\Big)^2}{m}\operatorname{Var}\Big(\tilde{z}_{m+j+1}^{m+k}(h)-\tilde{z}_1^{m+j}(h)\Big)\\ &= \frac{\Big((k-j)w(k/m)\Big)^2}{m}\Big[\operatorname{Var}\Big(\tilde{z}_{m+j+1}^{m+k}(h)\Big)+\operatorname{Var}\Big(\tilde{z}_1^{m+j}(h)\Big)\Big]\\ &\geq \frac{mt_0^2}{(1+T)^2}\operatorname{Var}\Big(\tilde{z}_1^{m+j}(h)\Big)=\frac{mt_0^2}{(m+j)(1+T)^2}\geq\frac{t_0^2}{(1+T)^3}\end{aligned}$$

and

$$\operatorname{Var}\Big(\tilde{v}_{m,k,j,h}\Big)\leq\frac{mT^2}{(1+T)^2}\Big(\frac{1}{k-j}+\frac{1}{m}\Big)\leq\frac{T^2}{(1+T)^2}\frac{1+t_0}{t_0}\ .$$

Using these bounds, we can apply Lemma B.2 which yields

$$\begin{aligned}\mathbb{P}\Big(\widetilde{\mathcal{T}}_{m,d}^{(Z)}(t_0)\leq u_d(x)\Big)-\mathbb{P}\Big(\widetilde{\mathcal{T}}_{m,d}^{(Z)}(t_0)\leq u_d(x)-y_d\Big) &= \mathbb{P}\Big(-y_d\leq\widetilde{\mathcal{T}}_{m,d}^{(Z)}(t_0)-u_d(x)\leq 0\Big)\\ &\leq\sup_{z\in\mathbb{R}}\mathbb{P}\Big(\Big|\widetilde{\mathcal{T}}_{m,d}^{(Z)}(t_0)-z\Big|\leq y_d\Big)\\ &\leq C_{T,t_0}\cdot y_d\Big(\sqrt{2\log(d)}+\sqrt{\max\{1,\log(\sigma_\ell/y_d)\}}\Big)=o(1)\ ,\end{aligned}$$

such that the assertion of Lemma A.6 follows by the choice of y_d. □

Proof of Lemma A.7. First, recall the definition of $\mathcal{W}_d$ and $\mathcal{W}_d(t_0)$ in (A.18) and note that

$$\begin{aligned}\mathcal{W}_d &= \max_{h=1}^{d}\max_{t\in[0,q(T)]}\max_{s\in[0,t]}\big|W_h(s)-W_h(t)\big|\\ &= \max\Big\{\mathcal{W}_d(t_0)\ ,\max_{h=1}^{d}\max_{t\in[0,q(t_0)]}\max_{s\in[0,t]}\big|W_h(s)-W_h(t)\big|\ ,\\ &\qquad\max_{h=1}^{d}\max_{t\in[q(t_0),q(T)]}\max_{s\in[q(q^{-1}(t)-t_0),t]}\big|W_h(s)-W_h(t)\big|\Big\}\\ &\leq\max\Big\{\mathcal{W}_d(t_0)\ ,\max_{h=1}^{d}\max_{\substack{|t-s|\leq t_0\\ s,t\in[0,q(T)]}}|W(t)-W(s)|\Big\}\end{aligned}$$

as $q(t_0)\leq t_0$ and $t-q(q^{-1}(t)-t_0)\leq t_0$. Hence, we obtain

$$\begin{aligned}\mathbb{P}\big(\mathcal{W}_d(t_0)\leq u_d(x)\big)-\mathbb{P}\big(\mathcal{W}_d\leq u_d(x)\big) &\leq\mathbb{P}\Big(\max_{h=1}^{d}\max_{\substack{|t-s|\leq t_0\\ s,t\in[0,q(T)]}}|W_h(t)-W_h(s)|>u_d(x)\Big)\\ &\leq d\,\mathbb{P}\Big(\max_{\substack{|t-s|\leq t_0\\ s,t\in[0,q(T)]}}|W_1(t)-W_1(s)|>u_d(x)\Big)\ .\end{aligned}$$

To control this probability we define an overlapping decomposition of the interval $[0, q(T)]$ by

$$I_j := [jt_0, (j+2)t_0] \ , \quad j = 0, 1, 2, \ldots, \lceil q(T)/t_0 \rceil - 2 \ .$$

Observing that the length of I_j is $2t_0$ we obtain

$$\begin{aligned}
&\mathbb{P}\Big(\max_{\substack{|t-s|\le t_0 \\ s,t\in[0,q(T)]}} |W_1(t) - W_1(s)| > u_d(x) \Big) \\
&\le \sum_{j=1}^{\lceil q(T)/t_0 \rceil - 2} \mathbb{P}\Big(\max_{\substack{|t-s|\le t_0 \\ s,t\in I_j}} |W_1(t) - W_1(s)| > u_d(x) \Big) \\
&\le \frac{q(T)}{t_0} \mathbb{P}\Big(\max_{\substack{|t-s|\le t_0 \\ s,t\in[0,2t_0]}} |W_1(t) - W_1(s)| > u_d(x) \Big) \\
&\le \frac{q(T)}{t_0} \mathbb{P}\Big(\max_{s,t\in[0,2t_0]} |W_1(t) - W_1(s)| > u_d(x) \Big) \\
&= \frac{q(T)}{t_0} \mathbb{P}\Big(\max_{s,t\in[0,q(T)]} |W_1(t) - W_1(s)| > u_d(x) \cdot \sqrt{q(T)/(2t_0)} \Big) \\
&\le \frac{q(T)}{t_0} \mathbb{P}\Big(\max_{s,t\in[0,q(T)]} |W_1(t) - W_1(s)| > c\sqrt{\log(d) q(T)/(2t_0)} \Big)
\end{aligned}$$

as for fixed x Lemma A.2 yields, that there exists a constant $c < \sqrt{2q(T)}$, such that

$$u_d(x) \ge c \cdot \sqrt{\log(d)}$$

for d sufficiently large. Using the representation of the distribution function $F_{\mathbb{M}}$ in (3.8) we obtain

$$\mathbb{P}\Big(\max_{h=1}^{d} \max_{\substack{|t-s|\le t_0 \\ s,t\in[0,q(T)]}} |W_h(t) - W_h(s)| > u_d(x) \Big), \le d\frac{q(T)}{t_0} \Big[1 - F_{\mathbb{M}}\Big(c\sqrt{\log(d) q(T)/(2t_0)} \Big) \Big] .$$

and L'Hôspital's rule gives

$$\begin{aligned}
\lim_{d\to\infty} d \Big[1 - F_{\mathbb{M}}\Big(c\sqrt{\log(d) q(T)/(2t_0)} \Big) \Big] &= c\sqrt{q(T)/(2t_0)} \lim_{d\to\infty} d \frac{F'_{\mathbb{M}}\Big(c\sqrt{\log(d) q(T)/(2t_0)} \Big)}{2\sqrt{\log(d)}} \\
&\le c\sqrt{q(T)/(2t_0)} \lim_{d\to\infty} dF'_{\mathbb{M}}\Big(c\sqrt{\log(d) q(T)/(2t_0)} \Big) \ .
\end{aligned}$$

Now substituting $d = \exp\Big(\frac{2y^2 t_0}{c^2 q(T)} \Big)$ yields that the last display can be written as

$$c\sqrt{q(T)/(2t_0)} \lim_{y\to\infty} \exp\Big(\frac{2y^2 t_0}{c^2 q(T)} \Big) F'_{\mathbb{M}}(y) \ ,$$

which by assertion (A.5) tends to zero for sufficiently small $t_0 > 0$ and thus completes the proof of Lemma A.7. □

A.4 Proof of Theorem 3.10

Denote the size of the change by $\Delta\mu_m = |\mu_{m+k^*-1,h^*} - \mu_{m+k^*,h*}|$ and the centered observations in component h^* by

$$X^{(c)}_{t,h^*} := X_{t,h^*} - \mathbb{E}[X_{t,h^*}] \ .$$

Observe the following lower bound

$$\begin{aligned}
\widehat{\mathcal{T}}_{m,d} &= \max_{h=1}^{d} \max_{k=1}^{Tm} w(k/m)\hat{E}_{m,h}(k) \\
&\overset{h=h^*}{\geq} \max_{k=1}^{Tm} w(k/m)\hat{E}_{m,h^*}(k) \overset{k=mT}{\geq} w(T)\hat{E}_{m,h^*}(mT) \\
&= \frac{1}{1+T} \max_{j=0}^{Tm-1} \frac{mT-j}{\sqrt{m}\hat{\sigma}_h} \left| \widehat{\mu}_{m+j+1}^{m+mT}(h^*) - \widehat{\mu}_1^{m+j}(h^*) \right| \\
&\overset{j=k^*-1}{\geq} \frac{1}{1+T} \frac{mT-k^*+1}{\sqrt{m}\hat{\sigma}_h} \left| \widehat{\mu}_{m+k^*}^{m+mT}(h^*) - \widehat{\mu}_1^{m+k^*-1}(h^*) \right| \\
&\geq \frac{1}{1+T} \left\{ \frac{mT-k^*+1}{\sqrt{m}\hat{\sigma}_h} \Delta\mu_m - \left| \frac{1}{\sqrt{m}\hat{\sigma}_h} \sum_{t=m+k^*}^{m+mT} X^{(c)}_{t,h^*} - \frac{mT-k^*+1}{\sqrt{m}\hat{\sigma}_h(m+k^*-1)} \sum_{t=1}^{m+k^*-1} X^{(c)}_{t,h^*} \right| \right\} .
\end{aligned} \tag{A.54}$$

The consistency of the long-run variance estimator $\hat{\sigma}_h$, Assumption 3.3, the FCLT in Theorem 3 of Wu (2005) and the Continuous Mapping Theorem show that

$$\begin{aligned}
&\left| \frac{1}{\sqrt{m}\hat{\sigma}_h} \sum_{t=m+k^*}^{m+mT} X^{(c)}_{t,h^*} - \frac{mT-k^*+1}{\sqrt{m}\hat{\sigma}_h(m+k^*-1)} \sum_{t=1}^{m+k^*-1} X^{(c)}_{t,h^*} \right| \\
&\leq \max_{s\in[0,T]} \left| \frac{1}{\sqrt{m}\hat{\sigma}_h} \sum_{t=m+\lfloor ms \rfloor+1}^{m+mT} X^{(c)}_{t,h^*} - \frac{mT-\lfloor ms \rfloor}{\sqrt{m}\hat{\sigma}_h(m+\lfloor ms \rfloor)} \sum_{t=1}^{m+\lfloor ms \rfloor} X^{(c)}_{t,h^*} \right| \\
&\overset{\mathcal{D}}{\Longrightarrow} \max_{s\in[0,T]} \left| W(1+T) - W(1+s) - \frac{T-s}{1+s} W(1+s) \right| ,
\end{aligned} \tag{A.55}$$

where W is a standard one-dimensional Brownian motion. Next note that (A.54) gives that

$$\begin{aligned}
\mathbb{P}\Big(a_d\big(\widehat{\mathcal{T}}_{m,d} - b_d\big) > g_{1-\alpha} \Big) &= \mathbb{P}\Big(\widehat{\mathcal{T}}_{m,d} > \frac{g_{1-\alpha}}{a_d} + b_d \Big) \\
&\geq \mathbb{P}\Big(- \left| \frac{1}{\sqrt{m}\hat{\sigma}_h} \sum_{t=m+k^*}^{m+mT} X^{(c)}_{t,h} - \frac{mT-k^*+1}{\sqrt{m}\hat{\sigma}_h(m+k^*-1)} \sum_{t=1}^{m+k^*-1} X^{(c)}_{t,h} \right| \\
&\qquad > \frac{g_{1-\alpha}}{a_d} + b_d - \frac{mT-k^*+1}{\sqrt{m}\hat{\sigma}_h} \Delta\mu_m \Big) .
\end{aligned}$$

By Assumption 3.1 (D1) and Lemma A.2 we obtain $b_d \sim \sqrt{\log(m)}$. Applying now (3.13)

$$\frac{g_{1-\alpha}}{a_d} + b_d - \frac{mT-k^*+1}{\sqrt{m}\hat{\sigma}_h} \Delta\mu_m \overset{\mathbb{P}}{\Longrightarrow} -\infty \ . \tag{A.56}$$

Now the proof is completed combining (A.54), (A.55), (A.56) with an application of Slutsky's Theorem.

A.5 Proof of Corollary 3.9

The result is obtained analogously to the corresponding parts of Theorem 2.5 in Jirak (2015) or Theorem 3.11 in Dette and Gösmann (2018). Therefore the proofs are omitted.

A.6 Proof of Theorem 3.13

Recall the definition of $u_d(x) = x/a_d + b_d$, $Z_{t,h}$, $\widehat{Z}_{t,h}$, $\mathcal{T}_{m,d}^{(Z)}$, $\mathcal{T}_{m,d}^{(Z)}(c)$ and $\widehat{\mathcal{T}}_{m,d}^{(Z)}(c)$ in (A.23), (A.19), (3.19), (A.20) and (A.22), respectively. The proof of Theorem 3.13 is based on the following three Lemmata.

Lemma A.8 *For the constant C_γ from Assumption 3.11 it holds that*

$$m^{C_\gamma} \max_{h,i=1}^{d} \big|\hat{\rho}_{h,i} - \rho_{h,i}\big| = o_{\mathbb{P}}(1) \ . \tag{A.57}$$

Proof. First, note that Assumption 3.3 and the Cauchy-Schwarz inequality imply that

(i) $\max_{h,i=1}^{d} \gamma_{h,i} \le \sigma_h \sigma_i \le C_\sigma^2$,

(ii) $\min_{h=1}^{d} \hat{\sigma}_h \ge \min_{h=1}^{d} \sigma_h - \max_{h=1}^{d} |\hat{\sigma}_h - \sigma_h| \ge c_\sigma - \max_{h=1}^{d} |\hat{\sigma}_h - \sigma_h|$,

(iii)
$$\begin{aligned}\max_{h,i=1}^{d} |\hat{\sigma}_h\hat{\sigma}_i - \sigma_h\sigma_i| &\le \max_{h,i=1}^{d} \hat{\sigma}_i|\hat{\sigma}_h - \sigma_h| + C_\sigma \max_{h=1}^{d} |\hat{\sigma}_h - \sigma_h| \\ &\le C_\sigma \max_{h=1}^{d} |\hat{\sigma}_h - \sigma_h|^2 + 2C_\sigma \max_{h=1}^{d} |\hat{\sigma}_h - \sigma_h| \ .\end{aligned}$$

Combining (i), (ii) and using again Assumption 3.3 gives

$$\max_{h,i=1}^{d} \left|\frac{\hat{\gamma}_{h,i}}{\hat{\sigma}_h\hat{\sigma}_i}\right| \le \frac{1}{\big(c_\sigma - \max_{h=1}^{d} |\hat{\sigma}_h - \sigma_h|\big)^2} \cdot \Big(C_\sigma^2 + \max_{h,i=1}^{d} \big|\hat{\gamma}_{h,i} - \gamma_{h,i}\big|\Big) = O_{\mathbb{P}}(1) \ .$$

Thus we obtain the upper bound

$$\begin{aligned}\max_{h,i=1}^{d} \big|\hat{\rho}_{h,i} - \rho_{h,i}\big| &\le \max_{h,i=1}^{d} \left|\frac{\hat{\gamma}_{h,i}}{\hat{\sigma}_h\hat{\sigma}_i} - \frac{\hat{\gamma}_{h,i}}{\sigma_h\sigma_i}\right| + \max_{h,i=1}^{d} \left|\frac{\hat{\gamma}_{h,i} - \gamma_{h,i}}{\sigma_h\sigma_i}\right| \\ &\le \frac{1}{c_\sigma^2} \max_{h,i=1}^{d} \left|\frac{\hat{\gamma}_{h,i}}{\hat{\sigma}_h\hat{\sigma}_i}\right| \big|\hat{\sigma}_h\hat{\sigma}_i - \sigma_h\sigma_i\big| + \frac{1}{c_\sigma^2} \max_{h,i=1}^{d} \big|\hat{\gamma}_{h,i} - \gamma_{h,i}\big| \\ &\lesssim O_{\mathbb{P}}(1) \max_{h=1}^{d} |\hat{\sigma}_h - \sigma_h|^2 + O_{\mathbb{P}}(1) \max_{h=1}^{d} |\hat{\sigma}_h - \sigma_h| + \max_{h,i=1}^{d} \big|\hat{\gamma}_{h,i} - \gamma_{h,i}\big| \ ,\end{aligned}$$

The assertion of Lemma A.8 now follows from Assumption 3.11. □

Lemma A.9 *There exists a sufficiently small constant $t_0 > 0$, such that for $x \in \mathbb{R}$*

$$\left|\mathbb{P}_{|\mathcal{X}}\Big(\widehat{\mathcal{T}}_{m,d}^{(Z)}(t_0) \le u_d(x)\Big) - \mathbb{P}_{|\mathcal{X}}\Big(\widehat{\mathcal{T}}_{m,d}^{(Z)} \le u_d(x)\Big)\right| = o_{\mathbb{P}}(1) \ ,$$

Proof. We provide a (stochastic) version of the proof of Lemma A.3. First note that

$$\begin{aligned}
\widehat{\mathcal{T}}_{m,d}^{(Z)} &= \max_{h=1}^{d} \max_{k=1}^{Tm} \max_{j=0}^{k-1} \frac{(k-j)w(k/m)}{\sqrt{m}} \left|\hat{\mu}_{m+j+1}^{m+k}(h) - \hat{\mu}_{1}^{m+j}(h)\right| \\
&= \max \Bigg\{ \widehat{\mathcal{T}}_{m,d}^{(Z)}(t_0) \ , \max_{h=1}^{d} \max_{k=t_0m+1}^{Tm} \max_{j=k-t_0m}^{k-1} \frac{(k-j)w(k/m)}{\sqrt{m}} \left|\hat{z}_{m+j+1}^{m+k}(h) - \hat{z}_{1}^{m+j}(h)\right| \ , \\
&\qquad \max_{h=1}^{d} \max_{k=1}^{t_0m} \max_{j=0}^{k-1} \frac{(k-j)w(k/m)}{\sqrt{m}} \left|\hat{z}_{m+j+1}^{m+k}(h) - \hat{z}_{1}^{m+j}(h)\right| \Bigg\} \ .
\end{aligned}$$

Hence, we obtain

$$\left|\mathbb{P}_{|\mathcal{X}}\Big(\widehat{\mathcal{T}}_{m,d}^{(Z)}(t_0) \le u_d(x)\Big) - \mathbb{P}_{|\mathcal{X}}\Big(\widehat{\mathcal{T}}_{m,d}^{(Z)} \le u_d(x)\Big)\right| \le P_1(x) + P_2(x) \ , \tag{A.58}$$

where the random variables $P_1(x)$ and $P_2(x)$ are defined by

$$\begin{aligned}
P_1(x) &= \mathbb{P}_{|\mathcal{X}}\Big(\max_{h=1}^{d} \max_{k=t_0m+1}^{Tm} \max_{j=k-t_0m}^{k-1} \frac{(k-j)}{\sqrt{m}} \left|\hat{z}_{m+j+1}^{m+k}(h) - \hat{z}_{1}^{m+j}(h)\right| \ge u_d(x)\Big) \ , \\
P_2(x) &= \mathbb{P}_{|\mathcal{X}}\Big(\max_{h=1}^{d} \max_{k=1}^{t_0m} \max_{j=0}^{k-1} \frac{(k-j)}{\sqrt{m}} \left|\hat{z}_{m+j+1}^{m+k}(h) - \hat{z}_{1}^{m+j}(h)\right| \ge u_d(x)\Big)
\end{aligned}$$

and we additionally used that $w(k/m) \le 1$. To complete the proof, it suffices by Markow's inequality to establish that

$$\mathbb{E}[P_1(x)] = o(1) \quad \text{and} \quad \mathbb{E}[P_2(x)] = o(1) \ .$$

To prove that assertions, observe the bounds

$$\begin{aligned}
\mathbb{E}[P_1(x)] &= \mathbb{P}\Big(\max_{h=1}^{d} \max_{k=t_0m+1}^{Tm} \max_{j=k-t_0m}^{k-1} \frac{(k-j)}{\sqrt{m}} \left|\hat{z}_{m+j+1}^{m+k}(h) - \hat{z}_{1}^{m+j}(h)\right| \ge u_d(x)\Big) \\
&\le \sum_{h=1}^{d} \mathbb{P}\Big(\max_{k=t_0m+1}^{Tm} \max_{j=k-t_0m}^{k-1} \frac{(k-j)}{\sqrt{m}} \left|\hat{z}_{m+j+1}^{m+k}(h) - \hat{z}_{1}^{m+j}(h)\right| \ge u_d(x)\Big)
\end{aligned} \tag{A.59}$$

and

$$\begin{aligned}
\mathbb{E}[P_2(x)] &= \mathbb{P}\Big(\max_{h=1}^{d} \max_{k=1}^{t_0m} \max_{j=0}^{k-1} \frac{(k-j)}{\sqrt{m}} \left|\hat{z}_{m+j+1}^{m+k}(h) - \hat{z}_{1}^{m+j}(h)\right| \ge u_d(x)\Big) \\
&\le \sum_{h=1}^{d} \mathbb{P}\Big(\max_{k=1}^{t_0m} \max_{j=0}^{k-1} \frac{(k-j)}{\sqrt{m}} \left|\hat{z}_{m+j+1}^{m+k}(h) - \hat{z}_{1}^{m+j}(h)\right| \ge u_d(x)\Big) \ .
\end{aligned} \tag{A.60}$$

The terms in (A.59) and (A.60) can now be controlled by the same arguments as given in the proof of Lemma A.3. □

Lemma A.10 *It holds that*

$$\left|\mathbb{P}_{|\mathcal{X}}\Big(\widehat{\mathcal{T}}^{(Z)}_{m,d} \le u_d(x)\Big) - \mathbb{P}_{H_0}\Big(\mathcal{T}_{m,d} \le u_d(x)\Big)\right| = o_{\mathbb{P}}(1) \ . \tag{A.61}$$

Proof. Observing Lemmas A.3, A.4, A.9, the assertion of Lemma A.10 follows, if we can establish that

$$\left|\mathbb{P}_{|\mathcal{X}}\Big(\widehat{\mathcal{T}}^{(Z)}_{m,d}(t_0) \le u_d(x)\Big) - \mathbb{P}\Big(\mathcal{T}^{(Z)}_{m,d}(t_0) \le u_d(x)\Big)\right| = o_{\mathbb{P}}(1) \ , \tag{A.62}$$

To obtain this, we will reuse the vector technique applied in the proof of Lemma A.4. From the proof of this Lemma recall the definition and construction of the Gaussian vector $\boldsymbol{V}^{(z)} = \Big(V_1^{(z)}, \dots, V_{d_{\boldsymbol{V}}}^{(z)}\Big)^\top$ which fulfilled the identity

$$\max_{i=1}^{d_{\boldsymbol{V}}} V_i^{(z)} = \mathcal{T}^{(Z)}_{m,d}(t_0) \ .$$

In exactly the same manner we can construct a vector $\widehat{\boldsymbol{V}}^{(z)} = \Big(\widehat{V}_1^{(z)}, \dots, \widehat{V}_{d_{\boldsymbol{V}}}^{(z)}\Big)^\top$ from $\big\{\widehat{Z}_{t,h}\big\}$, such that

$$\max_{i=1}^{d_{\boldsymbol{V}}} \widehat{V}_i^{(z)} = \widehat{\mathcal{T}}^{(Z)}_{m,d}(t_0) \ .$$

The covariance structure of $\boldsymbol{V}^{(z)}$ was already calculated in Lemma A.4. Repeating these steps for the conditional covariance structure of $\widehat{\boldsymbol{V}}^{(z)}$ with respect to $\mathcal{X}$, we directly obtain that

$$\max_{i_1,i_2=1}^{d_{\boldsymbol{V}}} \left|\operatorname{Cov}\left(V_{i,1}^{(z)}, V_{i,2}^{(z)}\right) - \operatorname{Cov}_{|\mathcal{X}}\left(\widehat{V}_{i,1}^{(z)}, \widehat{V}_{i,2}^{(z)}\right)\right| \lesssim \max_{h,i=1}^{d} \left|\hat{\rho}_{h,i} - \rho_{h,i}\right| \ . \tag{A.63}$$

In the remainder of the proof we use the notation

$$\Delta_\rho = \max_{h,i=1}^{d} \left|\hat{\rho}_{h,i} - \rho_{h,i}\right| \ .$$

By assertion (A.63) we are able to apply the Gaussian comparison inequality from Lemma B.3, which gives

$$\begin{aligned}
&\sup_{x\in\mathbb{R}} \left|\mathbb{P}_{|\mathcal{X}}\Big(\widehat{\mathcal{T}}^{(Z)}_{m,d}(t_0) \le x\Big) - \mathbb{P}\Big(\mathcal{T}^{(Z)}_{m,d} \le x\Big)\right| \\
=&\sup_{x\in\mathbb{R}} \left|\mathbb{P}_{|\mathcal{X}}\Big(\max_{i=1}^{d_{\boldsymbol{V}}} \widehat{V}_i^{(z)} \le x\Big) - \mathbb{P}\Big(\max_{i=1}^{d_{\boldsymbol{V}}} V_i^{(z)} \le x\Big)\right| \le C\Delta_\rho^{1/3} \cdot \max\Big\{1, \log\big(d_{\boldsymbol{V}}/\Delta_\rho\big)\Big\}^{2/3}
\end{aligned}$$

due to Lemma A.8 and Assumption 3.1 the upper bound in the last display is of order $o_{\mathbb{P}}(1)$, which proves (A.62). □

Actual proof of Theorem 3.13. To obtain the theorem's assertions, note that by Corollary 3.9 we already know that

$$a_d(\widehat{\mathcal{T}}_{m,d} - b_d) \stackrel{\mathcal{D}}{\Longrightarrow} G \ , \tag{A.64}$$

and as the Gumbel distribution has a continuous c.d.f., Polya's theorem [see Serfling (2009), p. 18] directly implies convergence in Kolmogorov-metric, that is

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\Big(a_d(\widehat{\mathcal{T}}_{m,d} - b_d) \le u_d(x)\Big) - \mathbb{P}\big(G \le x\big) \right| = o(1) \ . \tag{A.65}$$

On the other hand, combining (A.61) with Theorem 3.6 implies that

$$a_d(\widehat{\mathcal{T}}^{(Z)}_{m,d} - b_d) \stackrel{\mathcal{D}}{\Longrightarrow} G \ , \tag{A.66}$$

conditional on $\mathcal{X}$ in probability. So a conditional version of Polya's theorem gives

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}_{|\mathcal{X}}\Big(a_d(\widehat{\mathcal{T}}^{(Z)}_{m,d} - b_d) \le u_d(x)\Big) - \mathbb{P}\big(G \le x\big) \right| = o_{\mathbb{P}}(1) \ .$$

By (A.65) and (A.66) the proof of Theorem 3.13 is complete. □

A.7 Proof of Theorem 3.15

Denote the centered observations by

$$X^{(c)}_{t,h} = X_{t,h} - \mathbb{E}[X_{t,h}] \ .$$

We first prove assertions (3.25) and (3.27) for the Gumbel quantile, that is $q = g_{1-\alpha}/a_d + b_d$.

Proof of (3.25): It holds that

$$\begin{aligned}
\mathbb{P}\Big(\widehat{\mathcal{S}}_{d,\alpha} \subset \mathcal{S}_d\Big) &= \mathbb{P}\Big(\max_{h\in\mathcal{S}_d}\max_{k=1}^{Tm} w(k/m)\hat{E}_{m,h}(k) \le g_{1-\alpha}/a_d + b_d\Big) \\
&= \mathbb{P}_{H_0}\Big(\max_{h\in\mathcal{S}_d}\max_{k=1}^{Tm} w(k/m)\hat{E}_{m,h}(k) \le g_{1-\alpha}/a_d + b_d\Big) \\
&\ge \mathbb{P}_{H_0}\Big(\max_{h=1}^{d}\max_{k=1}^{Tm} w(k/m)\hat{E}_{m,h}(k) \le g_{1-\alpha}/a_d + b_d\Big) \longrightarrow 1-\alpha \ ,
\end{aligned}$$

where we applied Corollary 3.9 for the last convergence.

Proof of (3.27): First, note that:

$$\mathbb{P}\Big(\mathcal{S}^c_d \subset \widehat{\mathcal{S}}^c_{d,\alpha}\Big) = \mathbb{P}\Big(\min_{h\in\mathcal{S}^c_d}\max_{k=1}^{Tm} w(k/m)\hat{E}_{m,h}(k) > g_{1-\alpha}/a_d + b_d\Big) \ . \tag{A.67}$$

We have the lower bound

$$\min_{h\in\mathcal{S}^c_d}\max_{k=1}^{Tm} w(k/m)\hat{E}_{m,h}(k) \ge \min_{h\in\mathcal{S}^c_d} w(T)\hat{E}_{m,h}(Tm)$$

$$\geq \min_{h\in\mathcal{S}_d^c} \frac{Tm - k_h^* - 1}{\sqrt{m}\hat{\sigma}_h(T+1)} \big|\widehat{\mu}_{m+k_h^*}^{m+Tm} - \widehat{\mu}_1^{m+k_h^*-1}\big| \geq A_1 - A_2 \ ,$$

where the terms A_1 and A_2 are given by

$$A_1 = \min_{h\in\mathcal{S}_d^c} \frac{Tm - k_h^* + 1}{\sqrt{m}\hat{\sigma}_h(T+1)} \big|\mu_{m+k_h^*-1} - \mu_{m+k_h^*}\big| \ ,$$

$$A_2 = \max_{h\in\mathcal{S}_d^c} \frac{1}{\sqrt{m}\hat{\sigma}_h(T+1)} \Bigg| \sum_{t=m+k_h^*}^{m+Tm} X_{t,h}^{(c)} - \frac{mT - k_h^* + 1}{m + k_h^* - 1} \sum_{t=1}^{m+k_h^*-1} X_{t,h}^{(c)} \Bigg| \ .$$

Therefore the probability given in (A.67) has the lower bound

$$\mathbb{P}\Big(a_d(A_1 - 2b_d) - a_d(A_2 - b_d) > g_{1-\alpha}\Big) \ .$$

Using Corollary 3.9 we obtain that

(A.68)

$$a_d\big(A_2 - b_d\big) \leq a_d\Bigg(\max_{h=1}^{d} \max_{k=1}^{Tm} w(k/m) \max_{j=0}^{k-1} \Bigg| \sum_{t=m+j+1}^{m+k} X_{t,h}^{(c)} - \frac{k-j}{m+j} \sum_{t=1}^{m+j} X_{t,h}^{(c)} \Bigg| - b_d \Bigg) = O_{\mathbb{P}}(1) \ .$$

Further it holds by Assumption (3.26) that for m sufficiently large

$$Tm - \max_{h\in\mathcal{S}_d^c} k_h^* > c \ ,$$

where $c > 0$ is a sufficiently small constant. By Assumptions 3.3 and 3.11 we have

$$\frac{1}{\max_{h=1}^{d} \hat{\sigma}_h} \geq \frac{1}{C_\sigma + \max_{h=1}^{d} |\hat{\sigma}_h - \sigma_h|} \overset{\mathbb{P}}{\Longrightarrow} \frac{1}{C_\sigma}$$

and Lemma A.2 shows that $b_d \sim \log(d)$ and $a_d \to \infty$. Combining this with the assertions above yields

$$\text{(A.69)} \qquad a_d\big(A_1 - 2b_d\big) \gtrsim a_d\Bigg(\sqrt{m}\frac{1}{\max_{h=1}^{d} \hat{\sigma}_h} \min_{h\in\mathcal{S}_d^c} \big|\mu_{m+k_h^*-1} - \mu_{m+k_h^*}\big| - 2b_d\Bigg) \overset{\mathbb{P}}{\Longrightarrow} \infty$$

A combination of (A.68) and (A.69) now proves (3.27).

To complete the proof of Theorem 3.15 it remains to discuss the case, where Bootstrap quantiles

$$\hat{q}_{m,1-\alpha} := \inf\Big\{ x \in \mathbb{R} \ \Big| \ \mathbb{P}_{|\mathcal{X}}\Big(a_d\big(\widehat{\mathcal{T}}_{m,d}^{(Z)} - b_d\big) \leq x\Big) \geq \alpha \Big\} \ ,$$

are used in the algorithm. However, it follows from Theorem 3.13 combined with Lemma 21.2 and (the arguments from) Lemma 23.2 in van der Vaart (1998) that

$$\hat{q}_{m,1-\alpha} \overset{\mathbb{P}}{\Longrightarrow} g_{1-\alpha} \ .$$

An application of Slutsky's Lemma to the statements above then completes the proof of Theorem 3.15.

B Technical auxiliary results

We require the following Nagaev-type inequality as given in the online supplement of Jirak (2015) which is a version of Theorem 2 in Liu et al. (2013). In particular the reader should note that the second bound is independent of h.

Lemma B.1 *Under Assumption 3.3 it holds for $x \gtrsim \sqrt{n}$*

$$\mathbb{P}\Big(\max_{k=1}^{n}\Big|\sum_{t=1}^{k} X_{t,h} - \mathbb{E}[X_{t,h}]\Big| > x\Big) \le C_p \frac{n}{x^p} + C_p \exp\Big(-c_p \frac{x^2}{n}\Big) ,$$

where the constants $c_p, C_p > 0$ depend on p and the sequence $\Big\{\sup_{h\in\mathbb{N}} \vartheta_{t,h,p}\Big\}_{t\in\mathbb{N}}$ only.

As an immediate consequence of the bound

$$\max_{k=1}^{n}\Big|\sum_{t=k}^{n} X_{t,h} - \mathbb{E}[X_{t,h}]\Big| \le 2\max_{k=1}^{n}\Big|\sum_{t=1}^{k} X_{t,h} - \mathbb{E}[X_{t,h}]\Big|$$

Lemma B.1 holds with adjusted constants also for the reversed partial sum maximum.

The following inequality is Lemma 2.1 in Chernozhukov et al. (2013).

Lemma B.2 *Let $Z = (Z_1, \ldots, Z_d)^\top$ be a zero mean Gaussian vector with covariance matrix Σ^Z whose diagonal entries are bounded by two constants σ_ℓ and σ_u, that is*

$$\sigma_\ell \le \Sigma^Z_{j,j} \le \sigma_u$$

for $j = 1, \ldots, d$. Then for $\delta > 0$ it holds that

$$\sup_{z\in\mathbb{R}} \mathbb{P}\Big(\Big|\max_{h=1}^{d} Z_h - z\Big| \le \delta\Big) \le C_\sigma \delta\Big(\sqrt{2\log(d)} + \sqrt{\max\{1, \log(\sigma_\ell/\delta)\}}\Big) ,$$

where the constant $C_\sigma > 0$ depends on σ_ℓ, σ_u.

The next tool is Lemma 3.1 from Chernozhukov et al. (2013).

Lemma B.3 *Let $\boldsymbol{U} = (U_1, \ldots, U_d)^\top$ and $\boldsymbol{V} = (V_1, \ldots, V_d)^\top$ denote two d-dimensional Gaussian vectors with covariance matrices Σ^U and Σ^V, respectively. Further assume that there are two constants $c_1, C_1 > 0$, such that for all $j = 1, \ldots, d$*

$$c_1 \le \big|\Sigma^U_{j,j}\big| \le C_1 .$$

Denote the maximum entry-wise distance of both covariance matrices by

$$\Delta := \max_{i,j=1}^{d} \big|\Sigma^U_{i,j} - \Sigma^V_{i,j}\big| .$$

Then it holds that

$$\sup_{x\in\mathbb{R}} \Big|\mathbb{P}\Big(\max_{i=1}^{d} U_i \le x\Big) - \mathbb{P}\Big(\max_{i=1}^{d} V_i \le x\Big)\Big| \le C\Delta^{1/3} \cdot \max\Big\{1, \log\big(d/\Delta\big)\Big\}^{2/3} ,$$

where the constant $C > 0$ depends on c_1 and C_1 only.