SEQUENTIAL CONTROL OF TIME SERIES BY FUNCTIONALS OF KERNEL-WEIGHTED EMPIRICAL PROCESSES UNDER LOCAL ALTERNATIVES

(Revision)

Ansgar Steland[1]
Fakultät für Mathematik
Ruhr-Universität Bochum, Germany

Date: 07/2003.

[1]Address of correspondence: Ansgar Steland, Ruhr-Universität Bochum, Fakultät für Mathematik, Mathematik 3 NA 3/71, Universitätsstr. 150, D-44780 Bochum, Germany.

Abstract. Motivated in part by applications in model selection in statistical genetics and sequential monitoring of financial data, we study an empirical process framework for a class of stopping rules which rely on kernel-weighted averages of past data. We are interested in the asymptotic distribution for time series data and an analysis of the joint influence of the smoothing policy and the alternative defining the deviation from the null model (in-control state). We employ a certain type of local alternative which provides meaningful insights. Our results hold true for short memory processes which satisfy a weak mixing condition. By relying on an empirical process framework we obtain both asymptotic laws for the classical fixed sample design and the sequential monitoring design. As a by-product we establish the asymptotic distribution of the Nadaraya-Watson kernel smoother when the regressors do not get dense as the sample size increases.

Keywords: Control chart, finance, microarrays, sequential test, smoothing, statistical genetics.

Introduction

In many applications one is interested in sequential statistical procedures in order to detect the first time point where a sequence of observations $\{Y_n\}$, a time series, is no longer homogeneous (stationary). For instance, in statistical genetics the problem arises to select appropriate models explaining the genetic component of a (complex) disease. Due to the large number of genes, checking all possible models is infeasible in many cases and one has to rely on (heuristic) search algorithms which analyze a certain subset. If we define an appropriate statistic to measure the explanatory power of a model, we obtain a sequence of observations where changes in the mean may indicate reasonable statistical models for the data at hand. A further potential field of application is the analysis of microarray data where time series of gene expression levels of genes are obtained. Changes in the gene expression levels may reflect certain biological processes. Finally, an important field of application is the analysis of financial time series. Capital markets produce huge sequential streams of financial data as returns, prices, or interest rates. Hence, sequential methods are an appropriate tool to detect departures from stationarity which may give rise to portfolio adjustments or other actions. Although unexpected structural changes give rise to level shifts (jumps) of economic processes, we often expect gradual structural changes, since in general markets process information in a continuous fashion.

Many (truncated) detection rules to detect changes in the distribution of a sequence of observations can be written as stopping times

$$S_N = \inf\{1 \le n \le N : T_n = T(Y_1, \dots, Y_n) > c\}$$

where $T_n = T(Y_1, \dots, Y_n)$ is a control statistic attaining large values if there is evidence that the process is no longer homogenous. Note that detection rules of this type are also of particular interest if we want to get a *sequential* answer to the following *a posteori* question: Given data $Y_1, \dots, Y_N$, when was it possible for the first time to detect a change in the time series *without* using data after the hypothesized change-point?

Often these stopping times can be represented as functionals of certain sequential empirical processes. We use this approach here, because it has several merits. First, we obtain asymptotic laws for both sampling designs, the sequential design and the fixed sample design. Second, it will turn out that the asymptotic distribution of the sequential procedure depends only on the second moments of the underlying process. Third, the construction of both control charts and sequential tests is quite straightforward. Finally, one may study which (stochastic) properties of a procedure are in fact properties of the underlying sequential empirical process and not due to the definition of the functional yielding the statistic of interest.

For an a posteriori approach to detect multiple change points which is based on similar kernel-weighted statistics we refer to Huskova and Slaby (2001). The application of U-statistics for a posteriori detection has been studied by several authors, we refer to Ferger (1994, 1997), Gombay and Horvath (1995), and the references given in these papers.

To motivate our approach let us briefly recall some basic results for the classical i.i.d. case. If $\{Y_n\}$ are i.i.d.(F_Y) with $E(Y_n) = 0$ and $EY_n^2 = 1$, one may use $T_n = N^{-1/2} \sum_{i=1}^n Y_i$, a CUSUM-type statistic. Then, the process $T_{[Ns]}$, $s \in [0,1]$, converges weakly to Brownian motion $B(s)$, $s \in [0,1]$,

$$T_{[N\circ]} \Rightarrow B, \tag{1}$$

as $N \to \infty$, and therefore, since $S_N/N = \inf\{s \in [0,1] : T_{[Ns]} > c\}$, we have

$$S_N/N \to \inf\{s \in [0,1] : B(s) > c\} \tag{2}$$

in distribution, as $N \to \infty$. Observing that

$$S_N/N > x \;\Leftrightarrow\; \sup_{s \in [0,x]} T_{[Ns]} \leq c \tag{3}$$

and $P(\sup_{s \in [0,x]} B(s) > b) = 2P(N(0,x) > b)$, we can further conclude that the distribution function (d.f.) of S_N/N satisfies

$$P(S_N/N \leq x) \to \int_0^x \frac{c}{\sqrt{2\pi s^3}} \exp(-c^2/(2s))\, ds, \tag{4}$$

as $N \to \infty$, for each $x \geq 0$ and by continuity of the right side also uniformly in $x \geq 0$ (cf. Shorack and Wellner (1986), p.33.) Whereas (2) still holds true for weakly dependent time series under mixing conditions, explicit formulas as (4) are hard to obtain under general conditions.

Motivated by these considerations and previous work (Brodsky and Darkhovsky (1993, 2000), Schmid and Steland (2000), Steland (2002a, 2002b, 2003a, 2003b)), the contribution of this paper is to establish invariance principles as (1) and (4) for detection rules based on certain sequential kernel smoothers. The results can be applied when the correlation structure of the underlying time series is known or can be estimated consistently. This means, we obtain approximate solutions without fitting a parametric times series model, i.e., estimating the full distribution. Whereas Steland (2003b) provides a law of large number for the normed delay and studies the optimal kernel choice, the results presented here provide the asymptotic distribution of the underlying empirical process and therefore also about the asymptotic distribution of the detection rule.

We allow for dependent mixing data and study certain local alternatives. From a statistical point of view it is interesting to model the deviations from stationarity in order to analyze how the components of the model affect the asymptotic distribution and thus the statistical properties of the procedures. Therefore we shall work with a semiparametric model. An essential component is a generic alternative m_0 which is translated and scaled to define the mean of the process under the alternative. The scaling parameter h will be related to the effective sample size of the stopping rule and will tend to ∞. The procedure can be interpreted as a sequential test of the one-sided testing problem

$$H_0 : m_0 = \mathbf{0} \qquad \text{versus} \qquad H_1 : m_0 \geq^* \mathbf{0}.$$

Here we use the notation $f \geq^* g$ for two functions $f, g : D \to \mathbb{R}$ if $f(s) \geq g(s)$ for all $s \in D$ with strict inequality for at least one $s \in D$. Note that using a different terminology we may say that the process is in a state of statistical control if H_0 holds true and is out-of-control if H_1 is true. Concerning the choice of T_n we will study a class of weighted averages of past data where the weights are defined by a smoothing kernel K.

The structure of the paper is as follows. Section 1 introduces the statistical model and the statistical detection procedure in detail. Our basic assumptions are stated in Section 2. The weak convergence of the underlying sequential empirical process is derived in Section 3 for a large class of dependent time series. We provide both results under the null hypothesis and under the alternative as specified above. It turns out that the weak limit is a nonstationary

Gaussian process. A weak sufficient condition for a.s. continuous sample paths is given. In Section 4 we apply the results to establish corresponding results for several sequential detection procedures which can be defined in terms of the sequential empirical process.

1. Model and detection procedure

Assume the observations $Y_1, \dots, Y_N$, $N \in \mathbb{N}$, arrive sequentially and satisfy

$$Y_n = m_n + \epsilon_n, \qquad n = 1, \dots, N,\ N \in \mathbb{N},$$

where $\{\epsilon_n\}$ is a mean zero, stationary, and α-mixing process with covariance function

$$r_0(k) = E(\epsilon_1 \epsilon_{1+k}), \qquad k \geq 0.$$

We parameterize the drift m_n as

$$m_n = m_0((n - t_q)/h_N)\mathbf{1}(n \geq t_q), \qquad n = 1, \dots, N,\ N \in \mathbb{N},$$

where $h = h_N,\ N \in \mathbb{N}$, is a sequence of positive constants with

$$N/h_N \to \zeta \in (0, +\infty),$$

as $N \to \infty$. t_q is a fixed but unknown change-point. $m_0 : [0, \infty) \to [0, \infty)$ is called *generic alternative* function and is assumed to be continuous in $t = 0$ with $m_0(0) = 0$. Precise conditions on m_0 will be given below, but they cover the important case that m_0 is a piecewise smooth function. Note that for each fixed $n \in \mathbb{N}$ we have $m_n \to m_0(0)$, as $N \to \infty$. In this sense m_0 defines a sequence of local alternatives. Note that we assume equidistant time points $n \in \mathbb{N}$ which will be denoted by $t_n,\ n \in \mathbb{N}$, to make calculations more transparent. The generalization to non-equidistant designs is straightforward.

The stopping rule used to detect deviations from stationarity is based on a weighted average of past data. We consider this type of detection rule, since it might be the most popular device in analyzing sequential streams of data. For example, financial analysts look at such weighted averages sequentially to derive signals to buy or sell financial instruments. Define

$$\widehat{m}_n = \sum_{i=1}^{n} K_h(t_i - t_n) Y_i.$$

Here K is a smoothing kernel and $K_h(z) = h^{-1}K(z/h)$ its rescaled version. By definition of $\widehat{m}_n$ the kernel K is evaluated for arguments $z \le 0$. Hence, h is the effective number of observations used by the procedure, if the support of K equals $[-1, 1]$. Consider the sequential decision rule which gives a signal at the random (stopping) time

$$S_N = \inf\{1 \le n \le N : \widehat{m}_n > c\}$$

where c is a prespecified threshold and N is a (large) integer. The properties of S_N and other related stopping rules, and the choice of the critical value c will be discussed in greater detail in Section 4. Let us now put the procedure in an empirical process framework. Introduce the stochastic sequential kernel-weighted partial sum process

$$\mathbb{M}_N(s) = \frac{h}{\sqrt{N}} \sum_{i=1}^{[Ns]} K_h(t_i - t_{[Ns]})Y_i, \qquad s \in [0,1]. \tag{5}$$

Now we can represent S_N as

$$S_N = N \inf\{s \in [0,1] : N^{1/2}h^{-1}\mathbb{M}_N(s) > c\}.$$

In view of $\mathbb{M}_N(0) = 0$ it can be assumed w.l.o.g. that $c \ge 0$.

This representation motivates to study the weak convergence of the process $\{\mathbb{M}_N\}$. We will show that $\mathbb{M}_N$ converges weakly for a rich class of dependent time series and therefore governs the asymptotic distributional properties of any stopping time which can be defined as a functional of $\mathbb{M}_N$.

2. Assumptions

Concerning the error terms (innovations) we require the following assumptions.

(E1) $\{\epsilon_n\}$ is a strictly stationary process with $E|\epsilon_1|^{r+\delta} < \infty$ for some $r \ge 4$ and $\delta > 0$.

(E2) $\{\epsilon_n\}$ is strongly mixing with

$$\alpha(k) \sim ak^{-\beta}$$

for some $\beta > \frac{r(r+\delta)}{2\delta}$.

Recall that such mixing conditions, which are standard in nonparametric statistics for weakly dependent data, are satisfied by many parametric time series models.

We restrict attention to kernels from the following class.

(K) Concerning the kernel K we assume that K is non-negative, bounded, i.e., $\|K\|_\infty < \infty$, $K \in L_1(\mathbb{R}_0^+)$, and Lipschitz continuous, i.e., there exists a constant L such that

$$|K(z_1) - K(z_2)| \leq L|z_1 - z_2|$$

for all $z_1, z_2 \in \mathbb{R}$. W.l.o.g. we can and shall assume that K is symmetric.

For results under the alternative, we have to assume the following condition (M) concerning m_0, and a condition on both K and m_0.

(M) m_0 is assumed to be a piecewise continuous function.

For $x \geq 0$ define

$$I(x) = \int_0^x K(s - x) m_0(s)\, ds.$$

(KM) We assume $|I(x)| < \infty$ for all $x \geq 0$, $I \in C(\mathbb{R}_0^+)$, $K \cdot m_0$ has bounded variation, i.e., $\int |d(K m_0)| < \infty$, and that there exists some $x^* > 0$ such that $I(x^*) > c$.

Notice that a sufficient condition for $\int |d(K m_0)| < \infty$ and $I \in C(\mathbb{R}_0^+)$ is to require $K, m_0 \in L_1(\mathbb{R}_0^+)$ with $\|K\|_\infty$, $\|m_0\|_\infty$, $\int |dK|$, $\int |dm_0| < \infty$, and K Lipschitz continuous. Then $\int |d(K m_0)| < \infty$ and I is Lipschitz continuous,

$$|I(x_1) - I(x_2)| \leq \zeta \|m_0\|_\infty \max\{LU\zeta, \|K\|_\infty\} |x_1 - x_2|,$$

for all $0 \leq x_1, x_2 \leq U$, $U > 0$ fixed.

3. The kernel-weighted sequential empirical process

Observe that $\mathbb{M}_N$ is a random element of the space $D[0,1]$ of right-continuous functions on $[0,1]$ with left-hand limits. When equipped with the Borel-σ-algebra and the Skorohod metric, $D[0,1]$ is a separable space, and empirical processes are measurable. Recall that a sequence $\{X_n\} \subset D[0,1]$ converges weakly in $D[0,1]$ to some $X \in D[0,1]$, denoted by $X_n \Rightarrow X$, as $n \to \infty$, if

$$\int h(X_n)dP \to \int h(X)dP,$$

as $n \to \infty$, holds true for all continuous and bounded functions $h : D[0,1] \to \mathbb{R}$. If we interpret $h(X_n)$, $h \in C_b(D([0,1]);\mathbb{R})$, as a characteristic or an aspect of the random function X_n, weak convergence to X means that all aspects $h(X_n)$ converge to the aspects $h(X)$ of X. Recall that $X_n \Rightarrow X$, as $n \to \infty$, holds true if and only if the finite-dimensional distributions converge, denoted by $X_n \stackrel{fidis}{\to} X$, $n \to \infty$, and the process $\{X_n\}$ is tight, i.e., for each $\varepsilon > 0$ there exists a compact set $K \subset D[0,1]$ such that $P(X_n \in K) \geq 1 - \varepsilon$. We refer to Billingsley (1968), Pollard (1984), and to Vaart and Wellner (1996) for treatments of the theory in general metric spaces.

3.1. **Weak convergence under stationarity.** The following Theorem formulates an invariance principle which asserts that the process $\mathbb{M}_N$ converges weakly in $D[0,1]$ to some random element $\mathbb{M}_\zeta$, as $N \to \infty$. Recall that $N/h_N \to \zeta \in (0,+\infty)$, as $N \to \infty$. We will not mention this fact in the sequel. The result will imply that we may approximate the distribution of interesting functionals of $\mathbb{M}_N$, e.g., stopping times, by the distribution of the functional of the Gaussian process $\mathbb{M}_\zeta$.

Our first Theorem provides weak convergence of $\mathbb{M}_N$ under the (global) hypothesis $H_0 : m_0 = \mathbf{0}$.

Theorem 3.1. *Assume (E1), (E2), and (K). For all* $0 \leq s,t \leq 1$ *the limit*

$$C_\zeta(s,t) = \lim_{N\to\infty} C_N(s,t)$$

exists, where

$$C_N(s,t) = \frac{h^2}{N} \sum_{i=1}^{[Ns]} \sum_{j=1}^{[Nt]} K_h(t_i - t_{[Ns]}) K_h(t_j - t_{[Nt]}) r_0(|t_i - t_j|).$$

Under the hypothesis $H_0 : m_0 = \mathbf{0}$ *the process* $\mathbb{M}_N(t)$ *defined by (5) converges weakly to a Gaussian process* $\mathbb{M}_\zeta$*,*

$$\mathbb{M}_N \Rightarrow \mathbb{M}_\zeta \qquad \text{in } D[0,1],$$

as $N \to \infty$. The Gaussian process $\mathbb{M}_\zeta$ is determined by

$$E\mathbb{M}_\zeta(t) = 0 \quad \textit{and} \quad Cov(\mathbb{M}_\zeta(t), \mathbb{M}_\zeta(s)) = C_\zeta(t,s)$$

for all $0 \le s, t \le 1$.

Proof. Let $0 \le s, t \le 1$. Then we have

$$\begin{aligned} C_N(t,s) &= \frac{h^2}{N}\left|\sum_{i=1}^{[Nt]}\sum_{j=1}^{[Ns]} K_h(t_i - t_n)K_h(t_j - t_n)r_0(|i-j|)\right| \\ &\le \frac{\|K\|_\infty^2}{N}\sum_{i=1}^{[Nt]}\sum_{j=1}^{[Ns]} |r_0(|i-j|)| \\ &\le \frac{\|K\|_\infty^2}{N}\sum_{i=1}^{N}\sum_{i=1}^{N} |r_0(|i-j|)| \\ &= \|K\|_\infty^2 \left\{ r_0(0) + 2\sum_{k=1}^{N-1}(1-k/N)|r_0(k)| \right\}. \end{aligned}$$

The right side converges absolutely by assumptions (A) and (B)(cf. Bosq (1996), Th. 1.5), since $\beta > \gamma/(\gamma-2)$ holds true if we define $\gamma = r + \delta$.

We will now verify asymptotic normality of the fidis. Fix a dimension $l \in \mathbb{N}$ and let $\mathbf{t} = (t_1, \ldots, t_l) \in \mathbb{R}^l$ be a vector of time points. W.l.o.g. we assume $t_1 \le \cdots \le t_l$. We shall employ the Cramer-Wold device to establish convergence of the fidis $P_{(\mathbb{M}_N(t_1),\ldots,\mathbb{M}_N(t_l))}$. Let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_l)' \in \mathbb{R}^l - \{\mathbf{0}\}$. Define

$$\sqrt{N}T_N = T_{N\mathbf{t}\boldsymbol{\lambda}} = \sum_{k=1}^{l} \lambda_k \mathbb{M}_N(t_k)$$

and $\sigma_N^2 = \mathrm{Var}\,(\sqrt{N}T_N)$. Obviously,

$$\sigma^2 = \lim_{N\to\infty} \sigma_N^2 = \sum_{k,k'=1}^{l} \lambda_k \lambda_{k'} C_\zeta(t_k, t_{k'}) < \infty.$$

We have

$$T_N = \sum_{i=1}^{[Nt_l]} w_i Y_i$$

with weights

$$w_i = w_{N\mathbf{t}\boldsymbol{\lambda}i} = \sum_{k=1}^{l} \lambda_k K((i - [Nt_k])/h)\mathbf{1}(i \leq [Nt_k]),$$

$i = 1, \dots, [Nt_l]$. For simplicity of notation put $n = [Nt_l]$. We use the standard large-block-small-block argument. The mixing condition ensures that the dependence between sums of large-block summands vanishes fast enough, whereas the small-block sums have only q summands where q is chosen such that their contribution is asymptotically negligible. Further, we apply Bradley's lemma as in (Bosq 1996, Th. 1.7) to approximate dependent r.v.s by independent ones. More precisely, the large blocks will have block lengths

$$p \sim n/\log n - n^{1/4},$$

and the small ones $q \sim n^{1/4}$ yielding $b \sim \log n$ blocks of each kind. Define for $j = 0, \dots, b-1$

$$\begin{aligned} L_j &= \sum_{i=j(p+q)+1}^{j(p+q)+p} w_i Y_i, \\ S_j &= \sum_{i=j(p+q)+p+1}^{(j+1)(p+q)-1} w_i Y_i, \\ R_N &= \sum_{i=(b-1)(p+q)+1}^{n} w_i Y_i. \end{aligned}$$

Then we can decompose the statistic T_N as

$$T_N = \sum_{j=0}^{b-1} L_j + \sum_{j=0}^{b-1} S_j + R_N.$$

Bradley's lemma (Bradley, 1983) yields the existence of b independent random variables $\widetilde{L}_0, \dots, \widetilde{L}_{b-1}$ with $L_j \stackrel{d}{=} \widetilde{L}_j$ and

$$P\left(|L_j - \widetilde{L}_j| > \frac{\varepsilon\sigma\sqrt{n}}{b} \right) \leq 11 \sup_j \left(\frac{\|L_j + c\|_\gamma}{\varepsilon\sigma\sqrt{n}} b \right)^{\gamma/(2\gamma+1)} [\alpha(q)]^{2\gamma/(2\gamma+1)},$$

if we put $c = p\eta \sup_j \|w_j Y_j\|_\gamma$ for some $\eta > 1$. Noting that

$$\sum_{i=1}^{N} w_i^2 \to \sum_{k=1}^{l} \lambda_k \int_0^{\zeta t_k} K(s - \zeta t_k)\, ds,$$

as $N \to \infty$, which yields $\sum_{i=(p+q)j+1}^{(p+q)j+p} w_i^2 \to w^*$ for some w^*, and using Yokoyama (1980, Th. 1) and the Cauchy-Schwarz inequality, one may show that $E|L_j|^{\gamma'} = O(p^{\gamma'/2})$ and $\|L_j + c\|_{\gamma'} = O(p^{1/2})$ for any $2 < \gamma' < \gamma$. Thus, we obtain

$$\begin{aligned} P\left(\left|\frac{1}{\sigma\sqrt{n}}\sum_{j=0}^{b-1}\widetilde{L}_j - \frac{1}{\sigma\sqrt{n}}\sum_{j=0}^{b-1}L_j\right| > \varepsilon\right) &= P\left(|\sum_j(\widetilde{L}_j - L_j)| > \varepsilon\sigma/\sqrt{n}\right) \\ &\leq \sum_j P(|\widetilde{L}_j - L_j| > \varepsilon\sigma/(b\sqrt{n})) \\ &= O\left(b \cdot \left(\frac{p^{1/2}}{\sqrt{n}/b}\right)^{\gamma/(2\gamma+1)} \alpha([n^{1/4}])^{2\gamma/(2\gamma+1)}\right) \\ &= o(1), \end{aligned}$$

since $\alpha(k) \sim ak^{-\beta}$ with $\beta > \gamma/(\gamma-2)$.

We shall now verify asymptotic normality of $(\sqrt{n}\sigma)^{-1}\sum_{j=0}^{b-1}\widetilde{L}_j$ by using a truncation argument. Define

$$\widetilde{L}_j^M = \widetilde{L}_j\mathbf{1}(|\widetilde{L}_j| \leq M), \qquad j = 0, \ldots, b-1,$$

where $M > 0$ is an arbitrary constant. Now the r.v.s $\{\widetilde{L}_j : j = 0, \ldots, b-1\}$ are bounded and therefore satisfy the Lindeberg condition. Further, independence yields

$$\operatorname{Var}\left(\frac{1}{\sqrt{n}\sigma}\sum_{j=0}^{b-1}\widetilde{L}_j - \frac{1}{\sqrt{n}\sigma}\sum_{j=0}^{b-1}\widetilde{L}_j^M\right) = \frac{1}{n\sigma^2}\sum_{j=0}^{b-1}E\widetilde{L}_j^2\mathbf{1}(|\widetilde{L}_j| > M).$$

By dominated convergence, $E\widetilde{L}_j^2\mathbf{1}(|L_j| > M) \to 0$, as $M \to \infty$. Hence

$$\frac{1}{\sqrt{n}\sigma}\sum_{j=0}^{b-1}\widetilde{L}_j^M - \frac{1}{\sqrt{n}\sigma}\sum_{j=0}^{b-1}\widetilde{L}_j \stackrel{P,L_2}{\to} 0,$$

as $n \to \infty$ and then $M \to \infty$, which verifies

(6) $$\frac{1}{\sqrt{n}\sigma}\sum_{j=0}^{b-1}L_j \stackrel{d}{\to} N(0,1),$$

as $n \to \infty$.

Analogously, since S_j has q summands,

(7) $$\frac{1}{\sigma\sqrt{qr}}\sum_{j=0}^{b-1}S_j \stackrel{d}{\to} N(0,1),$$

and therefore

$$\text{(8)} \qquad \frac{1}{\sqrt{n}\sigma}\sum_{j=0}^{b-1} S_j = \sqrt{\frac{qb}{n}}\frac{1}{\sigma\sqrt{qb}}\sum_{j=0}^{b-1} S_j \stackrel{L_2,P}{\to} 0,$$

as $n \to \infty$, since $qb/n = o(1)$. Finally, since R_N has less than $p+q \sim n/\log n$ L_2-summands, we obtain

$$\frac{R_n}{\sqrt{n}\sigma} \stackrel{L_2,P}{\to} 0,$$

as $n \to \infty$, which gives

$$\text{(9)} \qquad \frac{1}{\sqrt{n}\sigma}\sum_{j=0}^{b-1} L_j - \sqrt{N}T_N \stackrel{L_2}{\to} 0,$$

as $n \to \infty$. Hence, the asymptotic distribution of $(\sqrt{n}\sigma)^{-1}T_N$ coincides with the asymptotic distribution of $(\sqrt{n}\sigma)^{-1}\sum_{j=0}^{b-1} L_j$. Finally, (6), (9) and $N/n \to t_l^{-1}$ also imply

$$\operatorname{Var}\left(\frac{1}{\sqrt{N}\sigma}\sum_{j=0}^{b-1} L_j\right) \to \sigma^2 = \lim_{N\to\infty} \operatorname{Var}(\sqrt{N}T_N),$$

as $n = [Nt_l] \to \infty$. Thus, we may conclude

$$\mathbb{M}_N \stackrel{fidis}{\to} \mathbb{M}_\zeta,$$

as $N \to \infty$, by definition of $\mathbb{M}_\zeta$.

It remains to verify tightness of the process $\{\mathbb{M}_N(t)\}$. Recall that by assumption (A) $E|\epsilon_1|^{r+\delta} < \infty$ for some $r \geq 4$ and $\delta > 0$. We show that

$$\text{(10)} \qquad \limsup_N E(\mathbb{M}_N(t) - \mathbb{M}_N(s))^4 = O(|t-s|^4).$$

for all $0 \leq s \leq t \leq 1$. Then the Cauchy-Schwarz inequality yields for all $0 \leq t_1 \leq t \leq t_2 \leq 1$

$$\begin{aligned} &E|\mathbb{M}_N(t) - \mathbb{M}_N(t_1)|^2|\mathbb{M}_N(t_2) - \mathbb{M}_N(t)|^2 \\ &\quad \leq \sqrt{E(\mathbb{M}_N(t) - \mathbb{M}_N(t_1))^4}\sqrt{E(\mathbb{M}_N(t_2) - \mathbb{M}_N(t))^4} \\ &\quad = O(|t-s|^4) \end{aligned}$$

and tightness follows from (Billingsley 1968, Th. 15.6). Fix $0 \leq s \leq t \leq 1$ and define

$$\begin{aligned} a_{Ni}(t,s) &= K((t_i - t_{[Nt]})/h) - K((t_i - t_{[Ns]})/h), \\ b_{Ni}(t) &= K((t_i - t_{[Nt]})/h). \end{aligned}$$

Clearly, $\max_i |b_{Ni}(t)| \leq \|K\|_\infty$. Since K is Lipschitz-continuous and $\zeta = \lim N/h$, we have

$$|a_{Ni}(t,s)| = O(\zeta|t-s|)$$

where the O does not depend on i. Note that

$$\mathbb{M}_N(t) - \mathbb{M}_N(s) = \frac{1}{\sqrt{N}} \sum_{i=1}^{[Ns]} a_{Ni}(t,s) Y_i + \frac{1}{\sqrt{N}} \sum_{i=[Ns]+1}^{[Nt]} b_{Ni}(t) Y_i.$$

We shall estimate both terms separately. First recall that assumption (B) immediately implies

$$\sum_{j=0}^{\infty} (j+1)^{r/2-1} (\alpha(j))^{\delta/(\delta+r)} < \infty.$$

We apply Th. 1 of Yokohama (1980). In particular, there it is shown (see p. 47 eq. (4.1), p. 47 last estimate and p. 48) that

$$\sum_{i,j,k,l=1}^{n} |E\eta_i\eta_j\eta_k\eta_l| = O(n^2).$$

Therefore, we have

$$\begin{aligned} E\left(\frac{1}{\sqrt{N}} \sum_{i=1}^{[Ns]} a_{Ni}(t,s) Y_i \right)^4 &\leq \frac{1}{N^2} \sum_{i,j,k,l=1}^{[Ns]} a_{Ni}a_{Nj}a_{Nk}a_{Nl} |EY_iY_jY_kY_l| \\ &\leq \sup_i |a_{Ni}(t,s)|^4 \frac{1}{N^2} \sum_{i,j,k,l=1}^{[Ns]} |EY_iY_jY_kY_l| \\ &= O(\zeta|t-s|^4([Ns]/N)^2) \\ &= O(\zeta|t-s|^4). \end{aligned}$$

since $[Ns]/N \leq 1$. Using the same arguments we also obtain

$$\begin{aligned} E\left(\frac{1}{\sqrt{N}}\sum_{i=[Ns]+1}^{[Nt]} b_{Ni}(t)Y_i\right)^4 &\leq \sup_i |b_{Ni}(t)|^4 \frac{1}{N^2}\sum_{i,j,k,l=[Ns]+1}^{[Nt]} |EY_iY_jY_kY_l| \\ &= O(\zeta|t-s|^4(([Nt]-[Ns])/N)^2) \\ &= O(\zeta|t-s|^4). \end{aligned}$$

Thus we may conclude (10). □

For applications we need the asymptotic law of a finite approximation of $\mathbb{M}_N$ at arbitrary points $s_1, \dots, s_L$, as summarized in the following Corollary.

Corollary 3.1. *Assume (E1), (E2), and (K). Let $0 \leq s_1 < \cdots < s_L \leq 1$ be L ordered time points. Then, under the hypothesis $H_0 : m_0 = 0$,*

$$(\mathbb{M}_N(s_1), \dots, \mathbb{M}_N(s_L))$$

converges in distribution to the distribution of the random vector

$$(\mathbb{M}_\zeta(s_1), \dots, \mathbb{M}_\zeta(s_L)),$$

which is given by a multivariate normal distribution with mean $\mathbf{0} \in \mathbb{R}^L$ and covarianve matrix $\mathbf{S}_\zeta = (s_{\zeta,ij})$ with elements $s_{\zeta,ij} = C_\zeta(s_i, s_j)$, $1 \leq i, j, \leq L$, provided $N \to \infty$.

Let us briefly discuss convergence of the covariance function $C_N(s,t)$.

Remark 3.1. *In the proof of Theorem 3.1 it was shown that $|C_N(s,t)| = O(\|K\|_\infty \sum_k |r_0(k)|)$. Also note that for unbounded L_2-kernels one may use the bound*

$$|C_N(s,t)| = O\bigg(\zeta^{-1}(\int_0^{\zeta s}\int_0^{\zeta t} K^2(z_1-\zeta s)K^2(z_2-\zeta t)\,dz_2 dz_1)^{1/2}(\sum_k r_0(k)^2)^{1/2}\bigg)$$

to check convergence. However, for bounded kernels summability of the covariances suffices.

Remark 3.2. *If $\{Y_i\}$ are i.i.d. with common variance $0 < \sigma^2 < \infty$, it is straightforward to show*

$$C_\zeta(s,t) = \frac{\sigma^2 \int_0^{\zeta s} K(z-\zeta s)K(z-\zeta t)\,dz}{\zeta^2 \int_0^{\zeta s} K(z-\zeta s)\,dz \int_0^{\zeta t} K(z-\zeta t)\,dz}$$

for $0 \leq s \leq t \leq 1$.

The conditions on the error terms are satisfied by many time series models. Let us briefly discuss the following important case.

Remark 3.3. *Assume* $\{\epsilon_n\}$ *is a causal ARMA(p, q) process.* $\phi(B)\epsilon_n = \theta(B)Z_n$*, where* $\{Z_n\}$ *is white noise and*

$$\begin{aligned}\phi(z) &= 1 - \phi_1 z - \cdots - \phi_p z^p \\ \theta(z) &= 1 + \theta_1 z + \cdots + \theta_q z^q.\end{aligned}$$

Then, for $h \geq \max(p, q+1) - p$ *there exists* p *constants* β_{ij} *such that*

$$r_0(k) = \sum_{i=1}^{l} \sum_{j=0}^{r_i} \beta_{ij} k^j \xi_i^{-k}$$

where $\xi_1, \ldots, \xi_l$ *are the distinct (possibly complex) zeroes of* $\phi(z)$ *with multiplicities* r_i*. By causality,* $|\xi_i| > 1$, $i = 1, \ldots, l$*, is ensured. Thus, the convergence rate of* $r_0(k)$ *to* 0*, as* $k \to \infty$*, depends on the zeros* ξ_i *which are closest to the unit circle. Simple real zeroes contribute geometrically decreasing terms, whereas a pair of complex conjugate zeroes together contribute a geometrically damped sinusoidal term (cf. Brockwell and Davies (1991), Sec. 3.3).*

Remark 3.4. *The asymptotic covariance matrix* $\mathbf{S}_\zeta$ *can be estimated by using consistent estimates* $\widehat{r}_0(k)$ *for the covariance function* $r_0(k)$*,* $k \geq 1$*, of the underlying time series* $\{Y_n\}$*. A detailed discussion of appropriate estimators is beyond the scope of the paper, but if* $s_i = t_i, i = 1, \ldots, n$*, a candidate could be the estimator*

$$\widehat{s}_{\zeta,ij} = \frac{h^2}{N^2} \sum_{i'=1}^{[Nt_i]} \sum_{j'=1}^{[Nt_j]} K_h(t_i - t_n) K_h(t_j - t_n) \widehat{r}_0(|t_i - t_j|).$$

3.2. **Sample path properties of the process** $\mathbb{M}_\zeta$**.** It is of interest to discuss sufficient conditions which ensure that the process $\mathbb{M}_\zeta$ has a.s. continuous sample paths. In the proof of Theorem 3.1 we verified (10) for tightness using the mixing condition of $\{Y_N\}$. This condition is sufficient to ensure that the limit process is an element of $C[0,1]$ w.p. 1 (c.f. Vaart and Wellner (1986), 2.2.3). However, the Gaussian process $\mathbb{M}_\zeta$ can be defined as long as the covariances $C_N(\circ_1, \circ_2)$ converge, and that convergence does not require the mixing condition (B). Therefore, in this subsection we provide a sufficient condition for a.s. bounded and continuous sample paths which does not use the mixing condition.

For a centered $D[0,1]$-valued process $\mathbb{X}$ define the semi-metric

$$d_{\mathbb{X}}(s,t) = \{E|\mathbb{X}(t) - \mathbb{X}(s)|\}^{1/2}, \qquad s,t \in [0,1].$$

If $T = [0,1]$ is compact w.r.t. the $d_{\mathbb{X}}$-topology, define the covering number $N(d_{\mathbb{X}}, \epsilon)$ as the smallest number of $d_{\mathbb{X}}$-balls centered at points $t \in T$ with radius $\epsilon > 0$ that cover T. The packing number $D(d_{\mathbb{X}}, \epsilon)$ is the maximum number of ϵ-separated points in T. Covering and packing numbers are related by the fact that $N(d_{\mathbb{X}}, \epsilon) \leq D(d_{\mathbb{X}}, \epsilon) \leq N(d_{\mathbb{X}}, \epsilon/2)$. The entropy is given by $H(d_{\mathbb{X}}, \epsilon) = \log N(d_{\mathbb{X}}, \epsilon)$. For stationary (Gaussian) processes convergence of the related entropy integral, $\int_0^\eta H(d_{\mathbb{X}}, \epsilon)\, d\epsilon$, $\eta > 0$, is necessary and sufficient for a.s. continuous and bounded sample paths, whereas for nonstationary processes conditions on the entropy integral provide sufficient criteria (Adler (2003), ch. 2).

Lemma 3.1. *Assume $\{Y_n\}$ is a stationary process with $\sum_k |r_0(k)| < \infty$ where $r_0(k) = EY_1Y_{1+k}$. If K is Lipschitz continuous, we have*

$$d^2_{\mathbb{M}_N}(t,s) = O(|t-s|^2).$$

uniformly in $0 \leq s,t \leq 1$.

Proof. For $0 \leq s,t \leq 1$ define

$$\Delta K_h(t_i; t_{[Ns]}, t_{[Nt]}) = K_h(t_i - t_{[Ns]}) - K_h(t_i - t_{[Nt]})$$

and note that by Lipschitz continuity of K

$$|\Delta K_h(t_i; t_{[Nt]}, t_{[Ns]})| = O(h^{-1}L|([Nt] - [Ns])/h|)$$

W.l.o.g. we now assume $0 \le s \le t \le 1$. Observing that

$$\begin{aligned}
C_N(s,s) - C_N(s,t) &= \frac{h^2}{N} \sum_{i=1}^{[Ns]} \sum_{j=1}^{[Ns]} K_h(t_i - t_{[Ns]}) \Delta K_h(t_j; t_{[Ns]}, t_{[Nt]}) r_0(|i-j|) \\
&\quad - \frac{h^2}{N} \sum_{i=1}^{[Ns]} \sum_{j=[Ns]+1}^{[Nt]} K_h(t_i - t_{[Ns]}) K_h(t_j - t_{[Nt]}) r_0(|i-j|) \\
C_N(t,t) - C_N(t,s) &= \frac{h^2}{N} \sum_{i=1}^{[Nt]} \sum_{j=1}^{[Nt]} K_h(t_i - t_{[Nt]}) \Delta K_h(t_j; t_{[Nt]}, t_{[Ns]}) r_0(|i-j|) \\
&\quad + \frac{h^2}{N} \sum_{i=1}^{[Nt]} \sum_{j=[Ns]+1}^{[Nt]} K_h(t_i - t_{[Nt]}) K_h(t_j - t_{[Ns]}) r_0(|i-j|),
\end{aligned}$$

and re-arraging terms we see that

$$d^2_{\mathbb{M}_N}(s,t) = C_N(s,s) - C_N(s,t) + C_N(t,t) - C_N(t,s) \tag{11}$$

can be written as

$$d^2_{\mathbb{M}_N}(s,t) = U_N(s,t) + V_N(s,t) + W_N(s,t)$$

where

$$\begin{aligned}
U_N(s,t) &= \frac{h^2}{N} \sum_{i=1}^{[Ns]} \sum_{j=1}^{[Ns]} \Delta K_h(t_i; t_{[Ns]}, t_{[Nt]}) \Delta K_h(t_j; t_{[Ns]}, t_{[Nt]}) r_0(|i-j|) \\
V_N(s,t) &= \frac{h^2}{N} \sum_{i=[Ns]+1}^{[Nt]} \sum_{j=[Ns]+1}^{[Nt]} K_h(t_i - t_{[Nt]}) \Delta K_h(t_j; t_{[Nt]}, t_{[Ns]}) r_0(|i-j|) \\
W_N(s,t) &= \frac{h^2}{N} \sum_{i=[Ns]+1}^{[Nt]} \sum_{j=[Ns]+1}^{[Nt]} K_h(t_j - t_{[Ns]}) \Delta K_h(t_i; t_{[Nt]}, t_{[Ns]}) r_0(|i-j|)
\end{aligned}$$

First, we have

$$\begin{aligned}
U_N(s,t) &\le L^2 \left| \frac{[Nt] - [Ns]}{h} \right|^2 \frac{1}{N} \sum_{i,j=1}^{[Ns]} r_0(|i-j|) \\
&= O\left(2L^2 |t-s|^2 \zeta s \sum_k |r_0(k)| \right),
\end{aligned}$$

where the O does not depend on (s,t), since $|[Nt]/h - \zeta t| = O(|N/h - \zeta| + 1/h)$ if $|t| \leq 1$. Further, by stationarity, $V_N(s,t)$ can be estimated as follows.

$$\begin{aligned} V_N(s,t) &\leq h^{-1}\|K\|_\infty L \left| \frac{[Nt]-[Ns]}{h} \right| \frac{[Nt]-[Ns]}{h} \frac{1}{[Nt]-[Ns]} \sum_{i=1}^{[Nt]-[Ns]} r_0(|i-j|) \\ &= O\left(2\|K\|_\infty L |t-s|^2 \zeta \sum_k |r_0(k)| \right), \end{aligned}$$

uniformly in $0 \leq s,t \leq 1$. $W_N(s,t)$ is estimated analogously. Thus, we may conclude that the pseudo-metric satisfies

$$d^2_{\mathbb{M}_N}(t,s) = O(|t-s|^2),$$

uniformly in $0 \leq s,t \leq 1$. □

We are now in a position to formulate our sufficient criterion.

Theorem 3.2. *Assume $\{Y_N\}$ is stationary with $\sum_k |r_0(k)| < \infty$ where $r_0(k) = EY_1Y_{1+k}$. If K is bounded, in $L_1(\mathbb{R}_0^+)$, and Lipschitz continuous, then the Gaussian process $\mathbb{M}_\zeta$ is continuous and bounded on $[0,1]$ with probability 1.*

Proof. Lemma 3.1 immediately implies

$$d_{\mathbb{M}_\zeta}(s,t) = \lim_{N\to\infty} d_{\mathbb{M}_N}(s,t) = O(|t-s|),$$

uniformly in $0 \leq s,t \leq 1$. This in particular yields that $[0,1]$ is compact w.r.t. the $d_{\mathbb{M}_\zeta}$-topology. Put

$$p^2(u) = \sup_{|s-t|<u} d^2_{\mathbb{M}_\zeta}(s,t), \qquad u \geq 0.$$

Now there exists a constant $C > 0$ such that for any $\delta > 0$

$$\int_\delta^\infty p(e^{-u^2})\,du \leq C \int_0^\infty e^{-u^2}\,du = C\,\sqrt{\pi/2}.$$

Thus, Adler (2003, Th. 2.2.1) yields a.s. continuity and boundedness of $\mathbb{M}_\zeta$. □

3.3. **Behavior under the alternative.** Under the alternative $\mathbb{M}_N$ diverges at the rate $\sqrt{N}$ and converges to a finite constant when rescaled, as stated in the next result.

Theorem 3.3. *Assume (E1), (E2), and (K).*

(i) *Under the alternative $H_1 : m_0 \geq^* 0$ we have for fixed $s \in [0,1]$*

$$\frac{1}{\sqrt{N}}\mathbb{M}_N(s) \xrightarrow{P} \frac{1}{\zeta}\int_0^{\zeta s} K(z-\zeta s)m_0(z)dz,$$

as $N \to \infty$.

(ii) *If $K \cdot m_0$ has bounded variation, for any $0 < a \leq 1$*

$$\sup_{a\leq s\leq 1}\left|\frac{1}{\sqrt{N}}\mathbb{M}_N(s) - \int_0^{\zeta s} K(z-\zeta s)m_0(z)dz\right| \xrightarrow{P} 0,$$

as $N \to \infty$.

Proof. We may assume that m_0 has no jumps, otherwise one may argue on subintervals. Since $Y_n = m_n + \epsilon_n$ with $m_n = m_0(n/h)$, we have

$$\mathbb{M}_N(s) \;=\; \frac{h}{\sqrt{N}}\sum_{i=1}^{[Ns]} K_h(t_i - t_{[Ns]})m_0([t_i - t_q]/h) + \mathbb{M}'_N(s)$$

where $\mathbb{M}'_N$ stands for the process $\mathbb{M}_N$ where the Y_i's are substituted by the ϵ_i's. Now the assertion follows, since

$$\frac{1}{h}\sum_{i=1}^{[Ns]} K((t_i - t_{[Ns]})/h)m_0((t_i - t_q)/h) \to \frac{1}{\zeta}\int_0^{\zeta s} K(z-\zeta s)m_0(z)dz$$

and by Theorem 3.1

$$N^{-1/2}\mathbb{M}'_N = o_P(1),$$

as $N \to \infty$. To show (ii) we verify

$$\left\|\frac{1}{h}\sum_{i=1}^{[Ns]} K([t_i - t_{[Ns]}]/h)m_0((t_i - t_q)/h) - \int_0^{\zeta\circ} K(z-\zeta\circ)m_0(z)\,dz\right\|_{\infty,[a,1]} = O(1/N).$$

where the integral equals $\mu_\zeta(s) = \zeta s\int_0^1 K(\zeta s(y-1))m_0(\zeta sy)\,dy$. Define $t_i^* = \frac{i}{\zeta sh}$ for $i = 1,\ldots,[Ns]$. Since $N/h \to \zeta$, there exists a $N_0 \in \mathbb{N}$ such that for all $N \geq N_0$ we have

$0 \le t_i^* \le 1$ for all $i = 1, \ldots, [Ns] - 1$. Note that

$$\frac{1}{[Ns]} \sum_{i=1}^{[Ns]} \mu_\zeta(t_i^*) = \frac{1}{h} \sum_{i=1}^{[Ns]} K((t_i - t_{[Ns]})/h) m_0(t_i/h).$$

Hence, Koksma (1942/43) yields

$$\left| \frac{1}{h} \sum_{i=1}^{[Ns]} K((t_i - t_{[Ns]})/h) m_0((t_i - t_q)/h) - \mu_\zeta(s) \right| \le 2 \int |dK(\circ - \zeta s) m_0(\circ)| D_N(t_1^*, \ldots, t_{[Ns]}^*),$$

where

$$D_N(t_1^*, \ldots, t_{[Ns]}^*) = \sup_{[a,b] \subset [0,1]} \left| [Ns]^{-1} \sum_i \mathbf{1}(t_i^* \in [a,b]) - (b-a) \right|.$$

For our choice $\{t_i^*\}$ $D_N(t_1^*, \ldots, t_{[Ns]}^*) = 1/[Ns] = O(1/N)$, since $s \ge a$, (cf. Niederreiter (1992)). □

4. Sequential stopping rules

Let us now apply the results of the previous Section to stopping rules. Our starting point was the stopping time

$$S_N = \inf\{1 \le n \le N : \widehat{m}_n > c\},$$

for which we will derive the asymptotic behavior under the null hypothesis. Since in Theorem 3.3 the convergence is not uniform over $[0,1]$, let us consider the follwowing modified stopping rule to obtain a meaningful result under the alternative. For $0 < a \le 1$ define

$$\widetilde{S}_N^{(a)} = \inf\{[aN] \le n \le N : \widehat{m}_n > c\}.$$

and note that

$$\widetilde{S}_N^{(a)} = N \inf\{a \le s \le 1 : N^{1/2} h^{-1} \mathbb{M}_N(s) > c\}.$$

However, in view of the weak convergence of $\mathbb{M}_N$ to a Gaussian process, it is also interesting to consider the stopping rule

$$S_N^* = \inf\{s \in [0,1] : \mathbb{M}_N > c\}.$$

We discuss one-sided stopping rules, but the results carry over to two-sided procedures where a signal is given if the absolute value of the control statistic exceeds a positive threshold.

We will show that S_N/N and $\widetilde{S}_N^{(a)}/N$ converge to deterministic quantities. Under the alternative, that quantity is a function of the generic alternative m_0 and the smoothing kernel K which summarizes the influence of these components on the statistical properties of the stopping rule. In contrast to these results, we show that S_N^* converges weakly to the random variable $S_\zeta^* = \inf\{s \in [0,1] : \mathbb{M}_\zeta(s) > c\}$ which has a non-degenerate weak limit. The critical value can be obtained from the asymptotic distribution, e.g., to ensure certain average run lengths or type I error rates under the null hypothesis, asymptotically. In the former case, we choose c such that $E_0 S_\zeta^* \geq \xi$ for some given in-control average run length ξ. To control asymptotically the type I error rate, c is chosen to satisfy

$$P_0(S_\zeta^* < 1) = P_0\left(\sup_{0 \leq s < 1} \mathbb{M}_\zeta(s) > c \right) = \alpha \tag{12}$$

for some given $\alpha \in (0,1)$. Here P_0 and E_0 indicate reference to the null hypothesis.

Note that we will be concerned with convergence of infimums of the type $\inf\{s \in [a,b] : T_n(s) \in A\}$ with $0 < a < b < \infty$ and therefore use the convention $\inf \emptyset = b$.

In view of (3), we start with the following result about a sup-functional of $\mathbb{M}_N$.

Theorem 4.1. *Assume (E1), (E2), and (K). Under the null hypothesis $H_0 : m_0 = 0$ (in-control model) the following assertions hold true.*

(i) *We have*

$$\sup_{0 \leq s \leq \circ} \mathbb{M}_N(s) \Rightarrow \sup_{0 \leq s \leq \circ} \mathbb{M}_\zeta(s) \quad \textit{in } (D[0,1], d),$$

as $N \to \infty$.

(ii) *If $c > 0$ we have for all $x \in [0,1]$*

$$P\left(\sup_{0 \leq s \leq x} \mathbb{M}_N(s) \leq c \right) \to P\left(\sup_{0 \leq s \leq x} \mathbb{M}_\zeta(s) \leq c \right), \tag{13}$$

as $N \to \infty$. If $Var \mathbb{M}_\zeta(t) > 0$ for all $t \in (0,1]$, (13) holds true for all $c \in \mathbb{R}$.

Proof. We first verify assertion (i). Define the functional $\varphi : D[0,1] \to D[0,1]$,

$$\varphi(f)(x) = \sup_{0 \leq s \leq x} f(s), \qquad x \in \mathbb{R}.$$

We have to show $\varphi(\mathbb{M}_N) \Rightarrow \varphi(\mathbb{M}_\zeta)$ in $(D[0,1], d)$, as $N \to \infty$, which easily follows when working with equivalent versions. By the Dudley/Skorohod/Wichura representation theorem in general metric spaces (e.g. Shorack and Wellner (1986), Th. 4, p.47, and Remark 2, p. 49) there exists a probability space with equivalent versions $\widetilde{\mathbb{M}}_N$ (of $\mathbb{M}_N$) and $\widetilde{\mathbb{M}}_\zeta$ (of $\mathbb{M}_\zeta$) with a.s. convergent sample paths w.r.t. the d-topology, i.e.,

$$d(\widetilde{\mathbb{M}}_N, \widetilde{\mathbb{M}}_\zeta) \stackrel{a.s.}{\to} 0,$$

as $N \to \infty$. Since Theorem 3.2 ensures that $\mathbb{M}_\zeta \in C[0,1]$ w.p. 1, we even have

$$\|\widetilde{\mathbb{M}}_N - \widetilde{\mathbb{M}}_\zeta\|_\infty \stackrel{a.s.}{\to} 0,$$

as $N \to \infty$. Clearly, the latter implies that the right-hand side of the inequality

$$\left| \sup_{0 \le s \le x} \widetilde{\mathbb{M}}_N(s) - \sup_{0 \le s \le x} \widetilde{\mathbb{M}}_\zeta(s) \right| \le \sup_{0 \le s \le x} |\widetilde{\mathbb{M}}_N(s) - \widetilde{\mathbb{M}}_\zeta(s)|$$

converges to 0, as $N \to \infty$, for any $0 \le x \le 1$. Thus, we obtain

$$\begin{aligned} \|\varphi(\widetilde{\mathbb{M}}_N) - \varphi(\mathbb{M}_\zeta)\|_\infty &= \sup_{0 \le x \le 1} \left| \sup_{0 \le s \le x} \widetilde{\mathbb{M}}_N(s) - \sup_{0 \le s \le x} \widetilde{\mathbb{M}}_\zeta(s) \right| \\ &\le \sup_{0 \le s \le 1} |\widetilde{\mathbb{M}}_N(s) - \widetilde{\mathbb{M}}_\zeta(s)| \\ &\stackrel{a.s.}{\to} 0, \end{aligned}$$

yielding

$$d(\varphi(\widetilde{\mathbb{M}}_N), \varphi(\widetilde{\mathbb{M}}_\zeta)) \stackrel{a.s.}{\to} 0,$$

as $N \to \infty$. This implies weak convergence of the related functionals of the orginal processes,

$$\varphi(\mathbb{M}_N) \Rightarrow \varphi(\mathbb{M}_\zeta) \qquad \text{in } (D[0,1], d), \tag{14}$$

as $N \to \infty$ (cf. Shorack and Wellner (1986), Corollary 1, p. 48.) Of course, the latter fact yields convergence of the d.f.s in all continuity points of the limit distribution. The question arises whether the distribution of $\varphi(\mathbb{M}_\zeta) = \sup_{0 \le s \le x} \mathbb{M}_\zeta(s)$ may have atoms. Since $\mathbb{M}_\zeta \in C[0,1]$ w.p. 1, it is sufficient to consider $\varphi | C[0,1]$. Clealy, $(C[0,1], \| \circ \|_\infty)$ is a separable Banach function space. Thus, we may apply Lifshits (1982, Th. 2) which asserts that $\nu_x = \mathcal{L}(\sup_{0\le s\le x} \mathbb{M}_\zeta(s))$ can have an atom only at the point

$$\gamma_x = \sup_{0\le t\le x:\operatorname{Var}\mathbb{M}_\zeta(t)=0} E\mathbb{M}_\zeta(t),$$

vanishes on the ray $(-\infty, \gamma_x)$, since $\sup_{0\le s\le x} \mathbb{M}_\zeta(s) \ge \gamma_x$ w.p. 1, and is absolutely continuous with respect to Lebesgue measure on $(\gamma_x, +\infty)$. Since $E\mathbb{M}_\zeta(s) = 0$ for all $s \in [0,1]$ provided H_0 is true, we have $\gamma_x = 0$ for all x. Hence, all $c > 0$ are continuity points of the distribution of $\sup_{0\le s\le x} \mathbb{M}_\zeta(s)$. Therefore, assertion (ii) follows. □

It is clear that expectation and d.f. of S_N^* are given by

$$ES_N^* = \int_0^\infty P\left(\sup_{s\in[0,x]} \mathbb{M}_N(s) \le c\right) dx,$$

and

$$F_{S_N^*}(x) = 1 - P\left(\sup_{s\in[0,x]} \mathbb{M}_N(s) \le c\right).$$

Theorem 4.1 now yields the following Corollary about the convergence of the latter, which justifies (12).

Corollary 4.1. *Under the assumptions of Theorem 4.1 (ii) for each* $x \in [0,1]$,

$$P(S_N^* \le x) \to P(S_\zeta^* \le x),$$

as $N \to \infty$, *where* S_ζ^* *has d.f.*

$$F_\zeta^*(x) = 1 - P\left(\sup_{0\le s\le x} \mathbb{M}_\zeta(s) \le c\right).$$

Let us now study the asymptotic behavior of the stopping rules S_N and $S_N^{(a)}$.

Theorem 4.2. *For each critical value* $c > 0$ *the following assertions hold true.*

(i) *Assume (E1), (E2), and (K). Under the null hypothesis* $H_0 : m_0 = 0$ *(in-control-model),*

$$\frac{S_N}{h} \xrightarrow{P} \zeta \quad \textit{and} \quad \frac{S_N}{N} \xrightarrow{P} 1,$$

as $N \to \infty$.

(ii) *Assume (E1), (E2), (K), (M), and (KM). Under the alternative* $H_1 : m_0 \geq^* 0$ *(out-of-control model) we have for each* $0 < a \leq 1$

$$\frac{\widetilde{S}_N^{(a)}}{N} \xrightarrow{P} \inf \{ s \in [a,1] : \mu_\zeta(s) > c \}$$

as $N, h \to \infty$ *with* $N \to \infty$, *where*

$$\mu_\zeta(s) = \int_0^{\zeta s} K(z - \zeta s) m_0(z) dz, \qquad s \in [0,1],$$

provided $\mu_\zeta(s') > c$ *for some* $s' \in [a,1]$.

Remark 4.1. *(a) A sufficient condition for* $\mu_\zeta = I(\zeta \circ) \in C(\mathbb{R}_0^+)$ *is given in Section 2. (b) For* $\zeta = 1$ *we obtain an analogue to Steland (2003b, Th. 2.2), where untruncated stopping rules are studied. In that paper, the functional optimization w.r.t. the kernel* K *is also discussed. For a Bayesian view on this issue see Steland (2002b).*

Proof. In Theorem 4.1 we have shown that

$$\|\mathbb{M}_N\|_{\infty,[0,1]} \xrightarrow{d} \|\mathbb{M}_\zeta\|_{\infty,[0,1]},$$

as $N \to \infty$. Hence,

$$\|\mathbb{M}_N\|_{\infty,[0,1]} = O_P(1)$$

which implies

$$d(N^{1/2} h^{-1} \mathbb{M}_N, 0) \xrightarrow{P} 0, \tag{15}$$

as $N \to \infty$. Define the functional $\varphi : D[0,1] \to [0,1]$,

$$\varphi(f) = \inf\{0 \leq s \leq 1 : f(s) > c\}, \qquad f \in D[0,1].$$

Eq. (15) implies

$$S_N/N = \varphi(N^{1/2} h^{-1} \mathbb{M}_N) \Rightarrow \varphi(0), \qquad \text{in } (D[0,1], d), \tag{16}$$

as $N \to \infty$. Since $\varphi(0) = 1$ is a constant, (16) is equivalent to $S_N/N \xrightarrow{P} 1$, as $N \to \infty$. The corresponding result for $S_N/h = (N/h) S_N/N$ is now straightforward.

It remains to prove (ii). Fix $0 < a \leq 1$. Define the functional $\varphi_a : D[a,1] \to [a,1]$,

$$\varphi_a(f) = \inf\{a \leq s \leq 1 : f(s) > c\}.$$

Clearly, $\varphi_a|\mathcal{E}_c$ is continuous w.r.t. $\| \circ \|_\infty$ and d, where

$$\mathcal{E}_c = \{f \in C[0,1] : f(x^*) > c \text{ for some } x^*\}.$$

The assertion is equivalent to

$$\varphi_a(N^{1/2}h^{-1}\mathbb{M}_N) \xrightarrow{P} \varphi_a(\mu_\zeta),$$

as $N \to \infty$. Since Theorem 3.3 yields

$$\sup_{a \le s \le 1} \left| \frac{\sqrt{N}}{h}\mathbb{M}_N(s) - \int_0^{\zeta s} K(z - \zeta s) m_0(z)\, dz \right| \xrightarrow{P} 0, \tag{17}$$

as $N \to \infty$, we obtain

$$\varphi_a(N^{1/2}h^{-1}\mathbb{M}_N) \Rightarrow \varphi_a(\mu_\zeta), \qquad \text{in } (D[a,1], d), \tag{18}$$

as $N \to \infty$, because $\mu_\zeta \in C[0,1] \cap \mathcal{E}_c$ is a continuity point of φ_a. Since μ_ζ is a deterministic function, (18) is equivalent to $\varphi_a(N^{1/2}h^{-1}\mathbb{M}_N) \xrightarrow{P} \varphi_a(\mu_\zeta)$, as $N \to \infty$, which proves the assertion. □

5. Conclusions

We derived the asymptotic distributions of kernel-weighted partial sum processes and related sequential stopping rules for time series satisfying a weak α-mixing condition. We discussed a stopping rule, S_N, which mimics the real behavior of non-statisticians, and related procedures which are suggested by the asymptotic results. From an applied viewpoint it is important to note that our results yield approximations which depend on the underlying distribution only through second moments. Further, as a by-product, we obtain the asymptotic distribution of the Nadaraya-Watson estimator under the monitoring sampling design of our settting, which differs from the design usually assumed in nonparametric regression. Working with a special kind of local alternatives yields interesting insights into the joint asymptotic influence of the smoothing kernel and the generic alternative defining the sequence of local alternatives.

Acknowledgements

The author is grateful to two anonymous referees for their constructive comments, which improved the paper. One referee draws my attention to Lifshits' paper.

The financial support of the Deutsche Forschungsgemeinschaft (SFB 475, ,,Reduction of complexity in multivariate data structures") is gratefully acknowledged.

References

[1] Adler, R. J. and Taylor, J. E. (2003). *Random Fields and their Geometry.* Under preparation for Birkhäuser-Verlag.

[2] Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[3] Bosq, D. (1996). *Nonparametric Statistics for Stochastic Processes.* Springer, New York.

[4] Bradley, R. (1983). Approximation theorems for strongly mixing random variables. *Michigan Mathematics Journal*, **30**, 69-81.

[5] Brockwell, P. J. and Davies, R. A. (1991). *Time Series: Theory and Methods*, Springer, New York.

[6] Brodsky, B. E. and Darkhovsky B. S. (1993). *Nonparametric Methods in Change-Point Problems.* Mathematics and its applications, **243**, Kluwer Academic Publishers, Dordrecht.

[7] Brodsky, B. E. and Darkhovsky B. S. (1993). *Non-Parametric Statistical Diagnosis Problems and Methods.* Mathematics and its applications, **509**, Kluwer Academic Publishers, Dordrecht.

[8] Ferger, D. (1994). On the power of nonparametric change-point tests. *Metrika*, **41**, 277-292.

[9] Ferger, D. (1997). Optimal bounds for the Prokhorov distance of the Miller-Sen process and Brownian motion. *Theory Probab. Appl.*, **42**, 155-162.

[10] Gombay, E. and Horvath, L. (1995). An application of the U-statistics to change-point analysis. *Acta Sci. Math (Szeged)*, **60**, 345-357.

[11] Huskova, M. and Slaby, A. (2001). Permutation tests for multiple changes. *Kybernetika*, **37**, 5, 605-622.

[12] Koksma, J. F. (1942/43). Een algemeene stelling uit de theorie der gelijkmatige verdeeling modula 1, *Mathematica B (Zutphen)*, **11**, 7-11.

[13] Lifshits, M. A. (1982). On the absoulte continuity of distributions of functionals of random processes, *Theory Probab. Appl.*, **27**, 600-607.

[14] Niederreiter, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods.* SIAM, Philadelphia, 1992.

[15] Schmid, W. and Steland, A. (2000). Sequential control of non-stationary processes by nonparametric kernel control charts. *Journal of the German Statistical Association*, **84**, 3, 315-336.

[16] Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes With Applications to Statistics*, Wiley, New York.

[17] Siegmund, D. (1985). *Sequential Analysis.* Springer, New York.

[18] Steland, A. (2002a). Nonparametric monitoring of financial time series by jump-preserving estimators. *Statistical Papers*, **43**, 361-377.

[19] Steland, A. (2002b). A Bayesian view on detecting drifts by nonparametric methods. *Economic Quality Control*, **17**, 177-186.

[20] Steland, A. (2003a). Jump-preserving sequential monitoring of dependent time series under local alternatives. *Submitted.*

[21] Steland, A. (2003b). Optimal sequential kernel smoothers under local nonparametric alternatives for dependent processes. *Submitted.*

[22] Vaart, A w. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer, New York.

[23] Yokohama, R. (1980). Moment bounds for stationary mixing sequences. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **52**, 45-57.