Constancy of distributions: asymptotic efficiency of certain nonparametric tests of constancy

Alex J. Koning
Econometric Institute
Erasmus University Rotterdam
P.O. Box 1738
NL-3000 DR Rotterdam
The Netherlands
`koning@few.eur.nl`

Nils Lid Hjort
Department of Mathematics and Statistics
University of Oslo
P.O. Box 1053 Blindern
N-0316 Oslo 3
Norway
`nils@math.uio.no`

Econometric Institute Report EI 2002-33

Abstract

In this paper we study stochastic processes which enable monitoring the possible changes of probability distributions over time. These so-called monitoring processes are bivariate functions of time and position at the measurement scale, and in particular be used to test the null hypothesis of no change: one may then form Kolmogorov–Smirnov or other type of tests as functionals of the processes. In Hjort and Koning (2001) Cramér-type deviation results were obtained under the constancy null hypothesis for [bootstrapped versions of] such "derived" test statistics.

Here the behaviour of derived test statistics is investigated under alternatives in the vicinity of the constancy hypothesis. When combined with Cramér-type deviation results, the results in this paper enable the computation of efficiencies of the corresponding tests. The discussion of some examples of yield guidelines for the choice of the test statistic, and hence for the underlying monitoring process.

1 Introduction and summary

Assume that independent data are available for each of n consecutive occasions, perhaps measurements of some quantity taken on separate dates. The null hypothesis to be tested here is that of

$$H_0 \ : \ F_1 = F_2 = \ldots = F_n, \tag{1}$$

where F_i is the cumulative distribution function specifying the distribution of data $X_{i,1}, \ldots, X_{i,m_i}$ on occasion i. We shall refer to $X_{i,1}, \ldots, X_{i,m_i}$ as the i^{th} subsample. Together, the subsamples form the full sample.

One may think of (1) as the hypothesis that an infinite dimensional parameter F remains constant. In this perspective, F_i is the value of F in the i^{th} subsample.

We shall denote the size $m_1 + \ldots + m_n$ of the full sample by m. Although it is not reflected in notation, note that m depends on n, and tends to infinity as n tends to infinity. The subsample sizes m_i are allowed to be random, and are conveniently represented by the random probability measure

$$\mu_n(t) = m^{-1} \sum_{i=1}^{[nt]} m_i, \quad t \in [0, 1].$$

Under the assumption that $\mu_n(t)$ converges to a deterministic function $\mu(t)$ in some predescribed manner as $n \to \infty$, null hypothesis theory for stochastic processes which enable monitoring (1) is presented in Hjort and Koning (2001); in particular, Komlós-Major-Tusnády type inequalities are employed to obtain deviation results for [bootstrapped versions of] test statistics based on these monitoring processes. In the sequel we shall refer to these statistics as "derived test statistics".

In this paper we develop "local alternatives theory"; that is, theory for the behaviour of the monitoring processes under alternatives in the vicinity of the null hypothesis. In combination with null hypothesis theory as in Hjort and Koning (2001), the local alternatives theory enables us to assess the ability of a monitoring process to detect departures from the null hypothesis. In fact, we shall investigate the performance of a derived test statistic by evaluating various "local" efficiency measures which pertain to the behaviour of the power curve in the vicinity of the null hypothesis.

Local efficiency measures are typically used as a selection device for statistical tests, as for any two tests which differ in efficiency there is a vicinity of the null hypothesis in which the more efficient one is more powerful than the less efficient one. For an enthusiastic review of the role of efficiency measures in the development of nonparametric statistics, we refer to Nikitin (1995).

In order to compute the various local efficiencies in a unified manner, we first show that the derived test statistic satisfies Condition III* in Wieand (1976) [cf. Definition 2(c) in Section 3]. The combination of a moderate deviation result under the null hypothesis and Condition III* in the vicinity of the null hypothesis enables the computation of limiting [as the alternative approaches the null hypothesis] approximate Bahadur efficiency, limiting [as the size of the test tends to zero] Pitman efficiency, and weak asymptotic i-efficiency. Moreover, replacing the moderate deviation result by a Cramér type deviation result [Chernoff type deviation result] yields asymptotic i-efficiency [strong asymptotic i-efficiency].

Following inspection of the structure of the evaluated efficiency, we formulate guidelines for constructing highly efficient derived tests. Indirectly, these guidelines shed light on the performance of the underlying monitoring process as well.

The outline of the paper is as follows. In Section 2 we introduce the monitoring processes, and study their behaviour under alternatives in the vicinity of the null hypothesis. In Section 3 we use the results of Section 2 to compute local efficiencies of derived test statistics. In Section 4 the methods are applied to sea water level data. Proofs are gathered in Section 5.

2 The alternative hypothesis

2.1 Notation and preliminaries

In this section we assume that the cumulative distribution functions $F_1, \dots, F_n$ are not equal to each other, but instead coincide with the n^{th} row $F_{n1}(\cdot;\delta), \dots, F_{nn}(\cdot;\delta)$ of a triangular array indexed by $\delta > 0$. Let $\mathcal{P}$ the class of all probability measures under consideration. It is convenient to think of $F_{ni}(\cdot;\delta)$ as the cumulative distribution function belonging to the i^{th} subsample at stage n under the probability measure $P_\delta \in \mathcal{P}/\mathcal{P}_0$, where δ indicates the distance of the alternative to the null hypothesis. The null hypothesis is assumed to correspond to $\delta = 0$, and P_δ "approaches" P_0 as δ tends to zero.

In this section we provide approximations of the monitoring processes under the alternative hypothesis, which hold true in the vicinity of the null hypothesis; that is, there exists a $\delta^* > 0$ such that the approximation holds true for $0 < \delta < \delta^*$. In particular, our intention is to show that the non-negative random variables Δ_n governing the approximation belong to a certain class $\mathcal{C}_a$. This class, which was inspired by Condition III* in Wieand (1976), is defined below.

Definition 1 *A sequence of random variables Δ_n is said to belong to the class $\mathcal{C}_a = \mathcal{C}_a(\{P_\delta : \delta > 0\})$ [notation: $\Delta_n \in \mathcal{C}_a$] if there exists a positive constant δ^* such that for every $\epsilon_1, \epsilon_2 > 0$ there exists a constant $C_{\epsilon_1,\epsilon_2} > 0$ such that*

$$P_\delta(|\Delta_n| > \epsilon_1) < \epsilon_2$$

for all $0 < \delta < \delta^*$ and n such that $m^{1/2}\delta > C_{\epsilon_1,\epsilon_2}$.

The following facts may facilitate computations.

- If $\Delta_n \in \mathcal{C}_a$, and if there exists a universal positive constant c such that $\Delta_n' \leq c\Delta_n$, then $\Delta_n' \in \mathcal{C}_a$ [take $C'_{\epsilon_1,\epsilon_2}$ equal to $C_{\epsilon_1/c,\epsilon_2}$].
- If $\Delta_n \in \mathcal{C}_a$ and $\Delta_n' \in \mathcal{C}_a$, then $\Delta_n + \Delta_n' \in \mathcal{C}_a$ and $\Delta_n\Delta_n' \in \mathcal{C}_a$.
- If $\Delta_n = |Y_n - 1| \in \mathcal{C}_a$ and $\Delta_n' = |Y_n' - 1| \in \mathcal{C}_a$, then $|Y_nY_n' - 1| \in \mathcal{C}_a$ [as follows from $|Y_nY_n' - 1| \leq \Delta_n + \Delta_n' + \Delta_n\Delta_n'$].
- [Wieand (1976)] If for every $\epsilon > 0$ there exists $C_\epsilon > 0$ such that $P_\delta(\Delta_n > C_\epsilon) < \epsilon$, then $\left(m^{1/2}\delta\right)^{-1}\Delta_n \in \mathcal{C}_a$.

The results of this paper depend for a large part on the way in which the F_{ni}'s differ from the "average" cumulative distribution function

$$F_{n\cdot}(x;\delta) = m^{-1}\sum_{i=1}^{n} m_i F_{ni}(x;\delta)$$

[and thus on the way in which the F_{ni}'s differ from each other]. Under Condition 1 the differences between the F_{ni}'s is conveniently described by the function $D(t,x;\delta)$.

Condition 1 *There exists a function $D(t,x;\delta)$, $t \in [0,1]$, $x \in I\!R$, such that*

$$\sup_{t\in[0,1]}\sup_{x\in\boldsymbol{R}}\left|(m\delta)^{-1}\sum_{i=1}^{[nt]} m_i\left\{F_{ni}(x;\delta) - F_{n\cdot}(x;\delta)\right\} - D(t,x;\delta)\right| \in \mathcal{C}_a.$$

If there exists a cumulative distribution function $F^*(x;\delta)$ such that

$$\delta^{-1}\sup_{x\in\boldsymbol{R}}|F_{n\cdot}(x;\delta) - F^*(x;\delta)| \in \mathcal{C}_a \tag{2}$$

[as is often the case], then for the verification of Condition 1 it suffices to show that

$$\sup_{t\in[0,1]}\sup_{x\in\boldsymbol{R}}\left|(m\delta)^{-1}\sum_{i=1}^{[nt]} m_i\left\{F_{ni}(x;\delta) - F^*(x;\delta)\right\} - D(t,x;\delta)\right| \in \mathcal{C}_a;$$

observe that in these circumstances we have $D(1,x;\delta) = 0$ for every $x \in I\!R$.

In the process of verifying Condition 1, it is often necessary to show that

$$\sup_{t\in[0,1]}|\mu_n(t) - \mu(t)| \in \mathcal{C}_a, \tag{3}$$

If the distribution of $m^{1/2}\sup_{t\in[0,1]}|\mu_n(t) - \mu(t)|$ remains bounded in probability, uniformly in $P \in \mathcal{P}$, then (3) holds as a consequence of the fourth fact mentioned following Definition 1.

The function $D(t,x;\delta)$ describes the nature of the departure from the null hypothesis. For instance, if under P_δ there is a sudden change [for $t = \gamma$, say] in the value of the cumulative distribution function in x, then $D(t,x;\delta)$ is proportional to the triangular function with value $-(1-\gamma)\mu(t)$ for $0 \le t \le \gamma$ and $\gamma(\mu(t) - \mu(1))$ for $\gamma < t \le 1$. On the other hand, if under P_δ there is a linear [in $\mu(t)$] trend present in the value of the cumulative distribution function in x, then $D(t,x;\delta)$ is proportional to $\mu(t)(\mu(1) - \mu(t))$.

2.2 Some examples

In this paragraph we discuss two examples of practical interest. The examples have in common that there exists a cumulative distribution function F_0 which is contaminated by a cumulative distribution F_c with probability $\delta v(i/n)$. The function $v(t)$ describes how the degree of contamination varies over time, and δ indicates the overall magnitude of contamination. That is, we have

$$F_{ni}(x;\delta) = F_0(x) + \delta v(i/n)\left(F_c(x) - F_0(x)\right) \tag{4}$$

for $i = 1, 2, \ldots, n$. To ensure that $\delta v(i/n)$ indeed is a probability, we shall assume that $0 \leq \delta v(t) \leq 1$ for every $0 \leq t \leq 1$. Define

$$V_n(t) = m^{-1}\sum_{i=1}^{[nt]} m_i v(i/n) = \int_0^t v(s)d\mu_n(s), \quad V(t) = \int_0^t v(s)d\mu(s).$$

Observe that the cumulative distribution function

$$F^*(x;\delta) = F_0(x) + \delta V(1)\left(F_c(x) - F_0(x)\right)$$

satisfies

$$m^{-1}\sum_{i=1}^{[nt]}\left\{F_{ni}(x;\delta) - F^*(x;\delta)\right\} = \delta\left\{V_n(t) - V(1)\mu_n(t)\right\}\left(F_c(x) - F_0(x)\right).$$

Hence, if

$$\sup_{t\in[0,1]}|V_n(t) - V(t)| \in \mathcal{C}_a, \qquad \sup_{t\in[0,1]}|\mu_n(t) - \mu(t)| \in \mathcal{C}_a,$$

then Condition 1 holds with

$$D(t,x;\delta) = W(t)\left(F_c(x) - F_0(x)\right),$$

where

$$W(t) = \int_0^t \left\{v(s) - V(1)\right\}d\mu(s) = V(t) - V(1)\mu(t).$$

Plots of $D(t,x;\delta)$ versus t all exhibit the same shape, determined by $W(t)$. The shape may reveal straight lines [indicating that $v(t)$ remains constant on the corresponding time interval], curvature [indicating a gradually changing $v(t)$] or angles [indicating abrupt changes in $v(t)$].

Example: "at most one" change-point The first example is the "at most one" change-point problem, well studied in literature [see Csörgő and Horváth (1997)]. In this example we have that the change point marks a sudden shift, say from $F_0(x)$ to $F(x;\delta)$. If

$$F(x;\delta) = F_0(x) + \delta\left\{F_c(x) - F_0(x)\right\},$$

then we have that (4) holds with

$$v_\gamma(t) = \begin{cases} 0 & \text{if } t \le \gamma, \\ 1 & \text{if } t > \gamma, \end{cases}$$

which yields

$$W_\gamma(t) = \int_0^t w_\gamma(s) d\mu(s) = \begin{cases} \mu(t)\,(1-\mu(\gamma)) & \text{if } t \le \gamma, \\ \mu(\gamma)\,(1-\mu(t)) & \text{if } t > \gamma. \end{cases}$$

If $\sup_{t\in[0,1]} |\mu_n(t) - \mu(t)| \in \mathcal{C}_a$, then Condition 1 holds with

$$D(t,x;\delta) = W_\gamma(t)\,(F_c(x) - F_0(x))\,,$$

as $\sup_{t\in[0,1]} |V_n(t) - V(t)| \in \mathcal{C}_a$ automatically holds due to the fact that $v_\gamma(t)$ is bounded by 1.

Example: linear trend In this example we have that the change point marks the onset of a trend, linearly in $\mu(t)$. Hence, we have that (4) holds with

$$v_\gamma(t) = \begin{cases} 0 & \text{if } t \le \gamma, \\ \mu(t) - \mu(\gamma) & \text{if } t > \gamma, \end{cases}$$

which yields

$$W_\gamma(t) = \begin{cases} -\frac{1}{2}\mu(t)\,(1-\mu(\tau))^2 & \text{if } t \le \gamma, \\ \frac{1}{2}\,(\mu(t)-\mu(\tau))^2 - \frac{1}{2}\mu(t)\,(1-\mu(\tau))^2 & \text{if } t > \gamma. \end{cases}$$

If $\sup_{t\in[0,1]} |\mu_n(t) - \mu(t)| \in \mathcal{C}_a$, then Condition 1 holds with

$$D(t,x;\delta) = \delta^{-1} W_\gamma(t)\,\{F(x;0) - F(x;\delta))\,,$$

as $\sup_{t\in[0,1]} |V_n(t) - V(t)| \in \mathcal{C}_a$ automatically holds due to the fact that $v_\gamma(t)$ is bounded by 1.

2.3 The basic process

In this paragraph we investigate the asymptotic behaviour of the basic process under fixed alternatives in the vicinity of the null hypothesis. Define $A_n(t,x;\delta)$ by

$$A_n(t,x;\delta) = m^{-1/2} \sum_{i \le [nt]} \sum_{j=1}^{m_i} \left(1_{\{X_{i,j} \le x\}} - F_{n\cdot}(x;\delta)\right) \quad t \in [0,1],\ x \in I\!R. \qquad (5)$$

Definition (5) coincides for $\delta = 0$ with the definition of the basic process $A_n(t,x)$ in Hjort and Koning (2001).

Theorem 1 is our key result under the alternative hypothesis. Its proof is deferred to Section 5.

Theorem 1 *If Condition 1 holds, then*

$$\sup_{t\in[0,1]}\sup_{x\in\boldsymbol{R}}\left|\left(m^{1/2}\delta\right)^{-1}A_n(t,x;\delta)-D(t,x;\delta)\right|\in\mathcal{C}_a. \tag{6}$$

2.4 Monitoring cumulative distribution functions

Let

$$\hat{F}_i(x)=\frac{1}{m_i}\sum_{j=1}^{m_i}1_{\{X_{i,j}\le x\}}$$

be the empirical estimator of $F(x)$ in the i^{th} subsample, and let

$$\bar{F}_n(x)=\frac{1}{m}\sum_{i=1}^{n}m_i\hat{F}_i(x)=\frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{m_i}1_{\{X_{i,j}\le x\}}$$

be the empirical estimator of $F(x)$ in the full sample. Lemma 1 in Hjort and Koning (2001) implies that under the null hypothesis (1) the process

$$B_n(t,x)=\frac{1}{\sqrt{m}}\sum_{i=1}^{[nt]}m_i\left(\hat{F}_i(x)-\bar{F}_n(x)\right),\quad t\in[0,1],\ x\in I\!R, \tag{7}$$

converges in distribution to a zero mean Gaussian process with covariance function $\{\mu\left(t\wedge t'\right)-\mu(t)\mu(t')\}\left\{F_0\left(x\wedge x'\right)-F_0(x)F_0(x')\right\}$.

In this paragraph we investigate the asymptotic behaviour of the monitoring process $B_n(t,x)$ under fixed alternatives in the vicinity of the null hypothesis.

Lemma 1 *If Condition 1 holds, then*

$$\sup_{t\in[0,1]}\sup_{x\in\boldsymbol{R}}\left|\left(m^{1/2}\delta\right)^{-1}B_n(t,x)-D(t,x;\delta)\right|\in\mathcal{C}_a. \tag{8}$$

Lemma 1 shows that the behaviour of $B_n(t,x)$ under a fixed alternative in the vicinity of the null hypothesis is largely determined by $m^{1/2}\delta D(t,x;\delta)$. This suggests that plots of $B_n(t,x)$ versus t and x may contain important information with respect to the nature of the departure from the null hypothesis. In particular, plots of $B_n(t,x)$ versus t for fixed x may reveal straight lines [indicating that the value of $F(x)$ remains constant throughout the corresponding time interval], curvature [indicating a gradually changing value of $F(x)$] or angles [indicating abrupt changes in the value of $F(x)$].

Lemma 2 in Hjort and Koning (2001) implies that under the null hypothesis (1) the process $C_n(t,x)$ defined by

$$C_n(t,x)=L(t)B_n(t,x)-\int_0^t B_n(s,x)\,dL(s),\quad t\in[0,1],\ x\in I\!R. \tag{9}$$

converges in distribution to a zero mean Gaussian process with covariance function

$$\left\{\int_0^{t\wedge t'} L(v)^2 d\mu(v) - \int_0^t L(v)d\mu(v)\int_0^{t'} L(v)d\mu(v)\right\}\left\{F(x\wedge x') - F(x)F(x')\right\}.$$

Lemma 2 extends Lemma 1 to the weighted monitoring process $C_n(t,x)$.

Condition 2 *There exists a finite constant $c_1 > 0$ such that $L(t)$ is bounded by c_1, and has variation bounded by c_1.*

Lemma 2 *Let $L(t)$ satisfy Condition 2, and define*

$$L * D(t,x;\delta) = L(t)D(t,x;\delta) - \int_0^t D(s,x;\delta)\,dL(s).$$

If Condition 1 holds, then

$$\sup_{t\in[0,1]}\sup_{x\in\mathbb{R}}\left|\left(m^{1/2}\delta\right)^{-1} C_n(t,x) - L * D(t,x;\delta)\right| \in \mathcal{C}_a.$$

2.5 Monitoring probability density functions

Let

$$\hat{f}_i(x) = \frac{1}{m_i h}\sum_{j=1}^{m_i} K\left(\frac{X_{i,j}-x}{h}\right)$$

be the kernel density estimator in subsample i, and let

$$\bar{f}_{n,h}(x) = \frac{1}{mh}\sum_{i=1}^{n}\sum_{j=1}^{m_i} K\left(\frac{X_{i,j}-x}{h}\right) = \frac{1}{m}\sum_{i=1}^{n} m_i\hat{f}_i(x)$$

be the full sample kernel density estimator under the null hypothesis (1); here, $K(x)$ is a symmetric density, and h a smoothing parameter.

Condition 3 *The kernel function $K(x)$ is a symmetric probability density function satisfying*

$$\int |K'(x)|\,dx < c_2,$$

where $K'(x)$ denotes the derivative of $K(x)$, and c_2 is a finite constant.

Lemma 3 in Hjort and Koning (2001) implies that under the null hypothesis (1) the monitoring process

$$B_{n,h}(t,x) = h^{1/2}m^{-1/2}\sum_{i=1}^{[nt]} m_i\left(\hat{f}_i(x) - \bar{f}_{n,h}(x)\right), \quad t\in[0,1],\ x\in\mathbb{R},$$

converges in distribution to a zero mean Gaussian process with covariance function $\{\mu\,(t \wedge t') - \mu(t)\mu(t')\}\,\sigma_h(x, x')$, where

$$\sigma_h(x, x') = h^{-1} \int K\left(\frac{v - x}{h}\right) K\left(\frac{v - x'}{h}\right) f_0(v)dv - h\mathcal{E}\bar{f}_h(x)\mathcal{E}\bar{f}_h(x'),$$

where

$$\mathcal{E}\bar{f}_h(x) = h^{-1} \int K\left(\frac{v - x}{h}\right) f_0(v)dv = \int K(u)f_0(x + hu)du.$$

One may interpret $\sigma_h(x, x')/h$ as the covariance function of the full sample estimator $\bar{f}_{n,h}(x)$ under the null hypothesis (1).

In this paragraph we investigate the asymptotic behaviour of the monitoring process $B_{n,h}(t, x)$ under fixed alternatives in the vicinity of the null hypothesis.

Lemma 3 *If Conditions 1 and 3 hold, then*

$$h^{1/2} \sup_{t\in[0,1]} \sup_{x\in\boldsymbol{R}} \left| \left(m^{1/2}\delta\right)^{-1} B_{n,h}(t, x) - h^{1/2}D_h(t, x; \delta)\right| \in \mathcal{C}_a, \tag{10}$$

where

$$D_h(t, x; \delta) = -h^{-2} \int K'\left(\frac{v - x}{h}\right) D(t, v; \delta)dv.$$

Lemma 3 shows that the behaviour of $B_{n,h}(t, x)$ under a fixed alternative in the vicinity of the null hypothesis is largely determined by $(mh)^{1/2}\delta D_h(t, x; \delta)$. Similar to plots of $B_n(t, x)$ discussed earlier, plots of $B_{n,h}(t, x)$ versus t for fixed x may reveal the nature of the departure from the null hypothesis: straight lines indicate that the value of $f(x)$ remains constant throughout the corresponding time-interval [here f denotes the probability density function belonging to F], curvature indicates a gradually changing value of $f(x)$, and angles indicate abrupt changes in the value of $f(x)$.

For a sequence of bandwidths h_n tending to zero, the situation is less simple due to the irregular asymptotic behaviour of $V_{n,h_n}(t, x; \delta)$ stemming from the structure of $\sigma_0(x, x')$.

Lemma 4 in Hjort and Koning (2001) implies that under the null hypothesis (1) the weighted monitoring process $C_{n,h}(t, x)$ defined by

$$C_{n,h}(t, x) = L(t)B_{n,h}(t, x) - \int_0^t B_{n,h}(s, x)\,dL(s), \quad t \in [0, 1],\ x \in I\!R \tag{11}$$

converges in distribution to a zero mean Gaussian process with covariance function

$$\left\{\int_0^{t\wedge t'} L(v)^2 d\mu(v) - \int_0^t L(v)d\mu(v) \int_0^{t'} L(v)d\mu(v)\right\} \sigma_h(x, x').$$

Lemma 4 shows that the results for $B_{n,h}(t, x)$ extend to the process $C_{n,h}(t, x)$.

Lemma 4 *Let $L(t)$ satisfy Condition 2, and define*

$$L * D_h(t,x;\delta) = L(t)D_h(t,x;\delta) - \int_0^t D_h(s,x;\delta)\,dL(s).$$

If Conditions 1 and 3 hold, then

$$\sup_{t\in[0,1]}\sup_{x\in\mathbb{R}}\left|\left(m^{1/2}\delta\right)^{-1} C_{n,h}(t,x) - h^{1/2}L * D_h(t,x;\delta)\right| \in \mathcal{C}_a.$$

3 Tests of constancy

3.1 Notation and preliminaries

In this section we show that the theory of the previous section is relevant for verifying whether a sequence of test statistics T_n based on either C_n or $C_{n,h}$ may be classified as a Wieand sequence and/or a weak Kallenberg sequence. Wieand sequences allow the computation of limiting [as the size of the test tends to zero] Pitman efficiency, and weak Kallenberg sequences allow the computation of weak i-efficiency. Throughout this section we shall assume that the test based on a test statistic T_n rejects the null hypothesis for large values of T_n.

In the first instance we shall restrict ourselves to testing the null hypothesis versus the alternative that P belongs to a path of probability measures $\{P_\delta : \delta > 0\}$ approaching $P_0 \in \mathcal{P}_0$ as $\delta \downarrow 0$. Restricting the alternative hypothesis is not uncommon in efficiency computations. For instance, in Nikitin (1995), p. 106, p. 122, the Bahadur efficiency of nonparametric tests is computed by restricting the alternative hypothesis to a simple hypothesis.

Definition 2 *A sequence of test statistics T_n is said to be a Wieand sequence if the following three conditions are satisfied.*

(a) For every $P \in \mathcal{P}_0$ the sequence T_n converges in P-distribution to a random variable T.

(b) There exists a positive constant a such that

$$\lim_{y\to\infty} y^{-2}\log\sup_{P\in\mathcal{P}_0} P\left(T > y\right) = -a/2.$$

(c) There exists a constant δ^* and a function $b : \mathbb{R}^+ \to \mathbb{R}$ such that for every $\epsilon_1 > 0$ and $\epsilon_2 \in (0,1)$ there exists a constant $C_{\epsilon_1,\epsilon_2}$ such that

$$P_\delta\left(\left|\left(m^{1/2}b(\delta)\right)^{-1} T_n - 1\right| > \epsilon_1\right) < \epsilon_2$$

for all $0 < \delta < \delta^*$ and n such that $m^{1/2}b(\delta) > C_{\epsilon_1,\epsilon_2}$.

Typically, we have that $b(\delta)$ satisfies $\lim_{\delta\downarrow 0} \delta^{-1} b(\delta) > 0$. If this is indeed the case, then

$$\left(m^{1/2} b(\delta)\right)^{-1} T_n - 1 \in \mathcal{C}_a \tag{12}$$

implies Definition 2(c).

If T_n is a Wieand sequence, then we shall refer to $a\,(b(\delta))^2$ as the Wieand slope of T_n. The following lemma is the composite null hypothesis version of the simple null hypothesis lemma given in Wieand (1976), and follows from Theorem 1 in Kallenberg and Koning (1995).

Lemma 5 *Let T_n and T_n' be two Wieand sequences with respective slopes $a\,(b(\delta))^2$ and $a'\,(b'(\delta))^2$. Suppose $\lim_{\delta\downarrow 0} b(\delta) = \lim_{\delta\downarrow 0} b'(\delta) = 0$, and suppose that the limit*

$$\lim_{\delta\downarrow 0} \frac{a\,(b(\delta))^2}{a'\,(b'(\delta))^2} \tag{13}$$

exists. Then the limiting [as the size of the tests tend to zero] asymptotic Pitman efficiency of T_n with respect to T_n' exists, and is equal to the limit given in (13).

It should be noted that the asymptotic Pitman efficiency of T_n with respect to T_n' does not depend on the size or the power of the test if both T_n and T_n' are asymptotically normal. However, the test statistics that we are considering typically have nonnormal limiting distributions, and hence the asymptotic Pitman eficiency may depend on the power and the size of the test, which makes the concept of Pitman efficiency less attractive as a performance measure. Lemma 5 shows that by letting the size of the test tend to zero, we arrive at a criterion that does not depend on the size and the power anymore.

The Wieand approach to efficiency is based on separately letting the size of the test tend to zero, and the alternative tend to the null hypothesis. In Kallenberg (1983) the concept of asympotic i-efficiency was proposed, in which both operations are performed simultaneously.

Definition 3 *A sequence of test statistics T_n is said to be a Kallenberg sequence if the following two conditions are satisfied.*

(a) There exists a positive constant a such that

$$\lim_{n\to\infty} (y_n)^{-2} \log \sup_{P\in\mathcal{P}_0} P\left(T_n > y_n\right) = -a/2 \tag{14}$$

holds for all sequences y_n such that $y_n \to \infty$ and $y_n = o\left(m^{1/6}\right)$ as $n \to \infty$.

(b) There exists a positive function $b(\delta)$ such that $\left(m^{1/2} b(\delta_n)\right)^{-1} T_n$ tends to 1 in P_{δ_n}-probability for all sequences δ_n such that $\delta_n \to 0$ and $m^{1/2} b(\delta_n) \to \infty$ as $n \to \infty$.

Observe that Definition 3(b) is implied by Definition 2(c) [since $\delta_n < \delta^*$ and $m^{1/2}b(\delta_n) > C_{\epsilon_1,\epsilon_2}$, eventually]. Recall that if $\lim_{\delta\downarrow 0} \delta^{-1}b(\delta) > 0$, then Definition 2(c) is in turn implied by (12).

Definition 3 is motivated by Lemma 2.1 in Kallenberg (1983), which uses the notion of Hellinger distance to identify sequences δ_n for which T_n is consistent under P_{δ_n}. Lemma 6 shows that if conditions (a) and (b) of Definition 2 hold, we may alternatively use $b(\delta)$ itself to identify those "consistent" sequences [as is done in Definition 3] if the size of the test is sufficiently small. Moreover, Lemma 6(a) implies that a Wieand sequence of test statistics is consistent under P_{δ_n} for all sequences δ_n such that $0 < \delta_n < \delta^*$ and $m^{1/2}b(\delta_n) \to \infty$ as $n \to \infty$.

Lemma 6 *Suppose that the sequence of test statistics T_n satisfies conditions (a) and (b) of Definition 2, and let δ_n be a sequence such that $\left(m^{1/2}b(\delta_n)\right)^{-1} T_n$ tends to 1 in P_{δ_n}-probability. Then*

(a) if $\lim_{n\to\infty} m^{1/2}b(\delta_n) = \infty$, then T_n is consistent under P_{δ_n};

(b) if $\limsup_{n\to\infty} m^{1/2}b(\delta_n) < \infty$, then the test based on T_n is not consistent under P_{δ_n} if its size is sufficiently small.

If T_n is a Kallenberg sequence, then we shall refer to $a\,(b(\delta))^2$ as the intermediate slope of T_n. If T_n and T_n' are two Kallenberg sequences with respective slopes $a\,(b(\delta))^2$ and $a'\,(b'(\delta))^2$, and if the limit $\lim_{\delta\downarrow 0} a\,(b(\delta))^2 / a'\,(b'(\delta))^2$ exists, then the asymptotic i-efficiency of T_n with respect to T_n' is defined as this limit.

Weak asymptotic i-efficiency, also proposed in Kallenberg (1983), is a variant of asymptotic i-efficiency which replaces the Cramér type deviation result (a) by a moderate deviation result: (14) should hold for all sequences y_n such that $y_n \to \infty$ and $y_n = O\left((\log m)^{1/2}\right)$ as $n \to \infty$.

For the sake of completeness, we mention that there is also strong asymptotic i-efficiency, which replaces the Cramér type deviation result by a Chernoff type deviation result: (14) should hold for all sequences y_n such that $y_n \to \infty$ and $y_n = o\left(m^{1/2}\right)$ as $n \to \infty$.

Lemma 6 in Hjort and Koning (2001) and Lemma 7 together provide a framework for verifying whether a sequence of test statistics is Wieand and/or [weak] Kallenberg. The sequence of test statistics is obtained by standardizing an initial sequence of test statistics $\tilde{T}_n$ by means of a random variable $\hat{\nu}$. Using standardized test statistics is quite natural in the light of condition (b) of Definition 2(b) and condition (a) of Definition 3.

Lemma 7 *Let $\{P_\delta : \delta > 0\}$ be a path of probability measures approaching $P_0 \in \mathcal{P}_0$ as $\delta \downarrow 0$. Suppose $\tilde{T}_n$ and $\hat{\nu}$ satisfy (i) and (ii) below.*

(i) For every $\delta > 0$ there exists $\tilde{b}(\delta)$ such that

$$\left(m^{1/2}\tilde{b}(\delta)\right)^{-1} \tilde{T}_n - 1 \in \mathcal{C}_a$$

holds, and $\lim_{\delta\downarrow 0} \delta^{-1}\tilde{b}(\delta) > 0$.

(ii) For every $\delta > 0$ there exists ν_δ such that

$$|\nu_\delta/\hat{\nu} - 1| \in \mathcal{C}_a,$$

and $\lim_{\delta\downarrow 0} \nu_\delta > 0$.

Then the test statistic $T_n = \hat{\nu}^{-1}\tilde{T}_n$ satisfies condition (c) of Definition 2 with $b(\delta) = \tilde{b}(\delta)/\nu_\delta$.

3.2 A general approach for sublinear tests

In this paragraph we briefly outline the verification of the conditions of Lemma 7 for test statistics based on the monitoring processes $C_n(t,x)$ and $C_{n,h}(t,x)$.

Let $D([0,1] \times I\!R)$ denote the space of real-valued functions defined on $[0,1] \times I\!R$ which are cadlag in both components, and let $T : D([0,1] \times I\!R) \to I\!R^+$ be a functional which is positive-homogeneous [that is, $T(c\xi) = cT(\xi)$ for every constant $c > 0$ and every $\xi \in D([0,1] \times I\!R)$] and Lipschitz [that is, there exists a constant $c_3 > 0$ such that $|T(\xi) - T(\xi')| \leq c_3 \sup_{t\in[0,1]} \sup_{x\in \boldsymbol{R}} |\xi(t,x) - \xi'(t,x)|$ for every $\xi, \xi' \in D([0,1] \times I\!R)$].

If we set

$$\tilde{T}_n = T(C_n),$$

and define $b(\delta)$ as $\delta T(L * D(\cdot,\cdot;\delta))$, then it follows from Lemma 2 that

$$\left|\left(m^{1/2}\delta\right)^{-1}\tilde{T}_n - \delta^{-1}b(\delta)\right| \leq c_3 \sup_{t\in[0,1]} \sup_{x\in\boldsymbol{R}} \left|\left(m^{1/2}\delta\right)^{-1} C_n(t,x) - L * D(t,x;\delta)\right| \in \mathcal{C}_a,$$

and hence condition (i) of Lemma 7 is satisfied if $\lim_{\delta\downarrow 0} T(D(\cdot,x;\delta))$ is positive.

Similarly, if we set

$$\tilde{T}_n = T(C_{n,h}),$$

and define $b(\delta)$ as $\delta h^{1/2} T(L * D_h(\cdot,\cdot;\delta))$, then it follows from Lemma 3 and Lemma 4 that condition (i) of Lemma 7 is satisfied if $\lim_{\delta\downarrow 0} T(D_h(\cdot,x;\delta))$ is positive. Again, it only remains to show that condition (ii) of Lemma 7 is satisfied.

3.3 Supremum type tests

To illustrate the general approach described in the previous paragraph, we now verify condition (ii) of Lemma 7 for the special case where T takes the form

$$T(\xi) = \sup_{x\in\boldsymbol{R}} S(\xi(\cdot,x)),$$

where

$$S(\xi_1) = \sup_{v\in V} \sqrt{Q_v(\xi_1,\xi_1)}$$

for every $\xi_1 \in D([0,1])$; here V is some index set, and Q_v is a symmetric bounded bilinear form on $D([0,1])$ for every $v \in V$ [see also Koning and Protasov (2001)]. Deviation results were obtained in Hjort and Koning (2001).

Typical examples of S are the Kolmogorov functional S_{Kol}, the Cramér-von Mises functional S_{CvM} and the Andersen-Darling functional S_{AD}. These functionals are respectively defined by

$$S_{\text{Kol}}(\xi_1) = \sup_{t\in[0,1]} |\xi_1(t)|,$$

$$S_{\text{CvM}}(\xi_1) = \left\{ \int (\xi_1(s))^2 \, d\mu(s) \right\}^{1/2},$$

$$S_{\text{AD}}(\xi_1) = \left\{ \int \frac{(\xi_1(s))^2}{\mu(s)(1-\mu(s))} d\mu(s) \right\}^{1/2}.$$

For each of these choices of S, there is an associated positive constant a_S: $a_{S_{\text{Kol}}} = 4$, $a_{S_{\text{CvM}}} = \pi^2$ and $a_{S_{\text{AD}}} = 2$ [cf. Koning and Protasov (2001)].

As in Hjort and Koning (2001), we shall mainly consider test statistics of the form T_{S,B_n} or $T_{S,B_{n,h}}$, as T_{S,C_n} and $T_{S,C_{n,h}}$ may be expressed as T_{S_L,B_n} and $T_{S_L,B_{n,h}}$ for a convenient choice of S_L.

Let T_{S,B_n} denote the test statistic $\sup_{x\in\boldsymbol{R}} S(B_n(\cdot,x))$. If a moderate deviation result holds for T_{S,B_n} with $a = 4a_S$ [cf. Lemma 7 in Hjort and Koning (2001)], and if $\lim_{\delta\downarrow 0} \sup_{x\in\boldsymbol{R}} S(D(\cdot,x;\delta))$ is positive, then it follows by Lemma 7 that T_{S,B_n} is both a Wieand and a weak Kallenberg sequence with slope $e\delta^2 + o(\delta^2)$, where

$$e = 4a_S \left\{ \lim_{\delta\downarrow 0} \sup_{x\in\boldsymbol{R}} S(D(\cdot,x;\delta)) \right\}^2$$

[we shall refer to e as the efficacy of T_{S,B_n}]. If a Cramér type deviation result holds for T_{S,B_n}, then T_{S,B_n} is a Kallenberg sequence. If a Chernoff type deviation result holds for T_{S,B_n}, then T_{S,B_n} is a strong Kallenberg sequence. Moreover, it follows by Lemma 6(a) that the test based on T_{S,B_n} is consistent for fixed alternatives in the vicinity of the null hypothesis, and for local alternatives which satisfy $m^{1/2}b(\delta_n) \to \infty$ as $n \to \infty$.

Let $T_{S,B_{n,h}}$ denote the test statistic $\sup_{x\in\boldsymbol{R}} S(B_{n,h}(\cdot,x))$. Lemma 8 provides a necessary additional result for $T_{S,B_{n,h}}$.

Lemma 8 *For $P \in \mathcal{P}_0$, define $\nu_0 = \nu_0(P)$ by $\nu_0^2 = \sup_{x\in\boldsymbol{R}} \sigma_h(x,x)$. Define the estimator $\hat{\nu}$ by*

$$\hat{\nu}^2 = \sup_{x\in\boldsymbol{R}} \frac{1}{mh} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left\{ K\left(\frac{X_{ij}-x}{h}\right) - \bar{f}_n(x) \right\}^2.$$

Assume Condition 3 holds, and assume that there exists a positive constant c_8 such that $c_8\nu_0 \geq 1$ for every $P \in \mathcal{P}_0$. If (2) holds, and if

$$\lim_{\delta\downarrow 0} \sup_{x\in\boldsymbol{R}} |F^*(x;\delta) - F_0(x)| = 0, \tag{15}$$

then condition (ii) of Lemma 7 is satisfied with $\lim_{\delta\downarrow 0}\nu_\delta = \nu_0$.

The assumption that ν_0 is bounded by below for $P \in \mathcal{P}_0$ may not be fulfilled in general for $T_{S,B_{n,h}}$. In such a case, one could consider the technical solution of removing from $\mathcal{P}_0$ those probability measures for which ν_0 becomes too small.

If the test statistic $T_{S,B_{n,h}}$ satisfies a moderate deviation result with $a = a_S$ and $\nu_0 = \sup_{x\in\boldsymbol{R}}\sigma_h(x,x)$ [cf. Lemma 8 in Hjort and Koning (2001)], and if the quantity $\lim_{\delta\downarrow 0}\sup_{x\in\boldsymbol{R}} S\left(D_h(\cdot,x;\delta)\right)$ is positive, then it follows by Lemma 7 that $\hat{\nu}^{-1}T_{S,B_{n,h}}$ is both a Wieand and a weak Kallenberg sequence with slope $e_h\delta^2 + o(\delta^2)$, where

$$e_h = \frac{h a_S}{\sup\limits_{x\in\boldsymbol{R}}\sigma_h(x,x)}\left\{\lim_{\delta\downarrow 0}\sup_{x\in\boldsymbol{R}} S\left(D_h(\cdot,x;\delta)\right)\right\}^2$$

[we shall refer to e_h as the efficacy of $T_{S,B_{n,h}}$]. If a Cramér type deviation result holds for $T_{S,B_{n,h}}$, then $T_{S,B_{n,h}}$ is a Kallenberg sequence. If a Chernoff type deviation result holds for $T_{S,B_{n,h}}$, then $T_{S,B_{n,h}}$ is a strong Kallenberg sequence. Moreover, it follows by Lemma 6(a) that the test based on $\hat{\nu}^{-1}T_{S,B_{n,h}}$ is consistent for fixed alternatives in the vicinity of the null hypothesis, and for local alternatives which satisfy $m^{1/2}b(\delta_n)\to\infty$ as $n\to\infty$.

Observe that e_h does not depend on n, and may be used as a criterion for selecting the bandwidth. Moreover, as h tends to zero, then we typically have that $D_{2,h}(x)$ as well as $\sigma_h(x,x)$ tends to a finite constant not equal to zero, and hence e_h tends to zero. Thus, letting h_n tend to zero as n tends to infinity yields an inefficient procedure.

3.4 Some examples

In this paragraph we return to the situation discussed in paragraph 2.2. Recall that if

$$\sup_{t\in[0,1]}|V_n(t)-V(t)|\in\mathcal{C}_a,\qquad \sup_{t\in[0,1]}|\mu_n(t)-\mu(t)|\in\mathcal{C}_a,$$

then Condition 1 holds with

$$D(t,x;\delta) = W(t)\left(F_c(x)-F_0(x)\right).$$

It follows that the slope $a(b(\delta))^2$ of the test statistic T_{S,B_n} behaves as $\delta^2 e_S e_{B_n}$ for δ tending to zero, where

$$e_S = a_S\left\{S\left(W\right)\right\}^2,\quad\text{and}\quad e_{B_n} = 4\sup_{x\in\boldsymbol{R}}\left\{F_c(x)-F_0(x)\right\}^2.$$

Moreover, the slope $a(b(\delta))^2$ of the test statistic $T_{S,B_{n,h}}$ behaves as $\delta^2 e_S e_{B_{n,h}}$ for δ tending to zero, where e_S is as before, and

$$e_{B_{n,h}} = \frac{\sup\limits_{x\in\boldsymbol{R}}\left\{\left(\int K'\left(\frac{v-x}{h}\right)\left(F_c(v)-F_0(v)\right)dv\right)\right\}^2}{h^3\sup\limits_{x\in\boldsymbol{R}}\sigma_h(x,x)}.$$

Apparently, the efficacy of test statistics of the form T_{S,B_n} or $T_{S,B_{n,h}}$ is the product of two factors. The first factor e_S is determined by the choice of S and the way the degree of contamination varies over time [as reflected by $W(t)$]. The second factor, either e_{B_n} or $e_{B_{n,h}}$, is determined by the choice of the monitoring process and the type of contamination [as reflected by $F_c(x) - F_0(x)$].

Recall that a test statistic of the form T_{S,C_n} may be expressed as T_{S_L,B_n} for a convenient choice of S_L. In particular, we may express $T_{S_{\text{Kol}},C_n} = \sup_{x\in\boldsymbol{R}} S_{\text{Kol}}(C_n(\cdot,x))$ as $\sup_{x\in\boldsymbol{R}} S_{L,\text{Kol}}(B_n(\cdot,x))$, where

$$S_{L,\text{Kol}}(\xi_1) = \sup_{t\in[0,1]} \left| L(t)\xi_2(t) - \int_0^t \xi_2(s)dL(s) \right|.$$

One may show that

$$a_{S_{L,\text{Kol}}}^{-1} = \sup_{t\in[0,1]} \int_0^t L(s)^2 d\mu(s) - \left(\int_0^t L(s)d\mu(s) \right)^2.$$

$$e_{S_{L,\text{Kol}}} = a_{S_{L,\text{Kol}}} \left\{ S_{L,\text{Kol}}\left(W\right) \right\}^2,$$

The freedom still remaining in the choice of L allows us to construct a test which has high power for a specific alternative of special interest. If w is known explicitly, then Lemma 9 shows that taking the weight function $L(t)$ equal to $w(t)$ yields a test statistic which is optimal within the class of tests statistics $\sup_{x\in\boldsymbol{R}} S_{L,\text{Kol}}(B_n(\cdot,x))$.

Lemma 9 *The ratio $e_{S_{L,Kol}}$ does not exceed $e_{\text{opt}} = \int_0^1 \{w(s)\}^2 \, d\mu(s)$, and this upper bound is attained by $L(t) = w(t)$.*

Typically, in practical circumstances the function $v(t)$ [and hence $w(t)$] is not fully specified. For instance, in the examples given in paragraph 2.2 we have that $v(t) = v_\gamma(t)$ depends on the changepoint γ, which is usually unknown. Nevertheless, the quantity e_{opt} is an upper bound for $e_{S_{L,\text{Kol}}}$, and hence we shall use it as a yardstick for e_S in the sequel.

Example: "at most one" change point problem In the "at most one" changepoint problem we have $D(t,x;\delta) = W_\gamma(t) \{F_c(x) - F_0(x)\}$, where

$$W_\gamma(t) = \int_0^t w_\gamma d\mu(s) = \begin{cases} \mu(t)\left(1-\mu(\gamma)\right) & \text{if } t \le \gamma, \\ \mu(\gamma)\left(1-\mu(t)\right) & \text{if } t > \gamma. \end{cases}$$

We have

$$e_{\text{opt}} = \int_0^1 w_\gamma(s)^2 ds = \mu(\gamma)\left(1-\mu(\gamma)\right),$$

$$e_{S_{\text{Kol}}} = a_{S_{\text{Kol}}} \left\{ S_{\text{Kol}}\left(W_\gamma\right) \right\}^2 = 4 \left\{ \sup_{t\in[0,1]} |W_\gamma(t)| \right\}^2 = 4\mu(\gamma)^2 \left(1-\mu(\gamma)\right)^2,$$

$$e_{S_{\text{CvM}}} = a_{S_{\text{CvM}}} \left\{ S_{\text{CvM}} \left(W_\gamma \right) \right\}^2 = \tfrac{1}{3}\pi^2 \left(1 - \mu(\gamma)\right)^2 \mu(\gamma)^2,$$

$$\begin{aligned} e_{S_{\text{AD}}} &= a_{S_{\text{AD}}} \left\{ S_{\text{AD}} \left(W_\gamma \right) \right\}^2 \\ &= -2\mu(\gamma)^2 \ln \mu(\gamma) - 2\left(1 - \mu(\gamma)\right)^2 \ln\left(1 - \mu(\gamma)\right) - 2\mu(\gamma)\left(1 - \mu(\gamma)\right). \end{aligned}$$

Figure 1 displays $e_{S_{\text{Kol}}}/e_{\text{opt}}$, $e_{S_{\text{CvM}}}/e_{\text{opt}}$ and $e_{S_{\text{AD}}}/e_{\text{opt}}$ versus $\mu(t)$, and has the following implications for the comparison of test statistics T_{S_{Kol},B_n}, T_{S_{AD},B_n} and T_{S_{CvM},B_n} derived from the monitoring process $B_n(t,x)$. The test statistic T_{S_{Kol},B_n} should always be preferred over T_{S_{CvM},B_n}, as the efficiency $e_{S_{\text{CvM}}}/e_{S_{\text{Kol}}} = \pi^2/12 = 0.8225$ of T_{S_{CvM},B_n} with respect to T_{S_{Kol},B_n} is less than 1 and does not depend on the position of the changepoint. However, for changepoints γ close to 0 or 1 [more precisely, satisfying $\mu(\gamma) \leq 0.15$ or $\mu(\gamma) \geq 0.85$], T_{S_{AD},B_n} shows a stronger performance than T_{S_{Kol},B_n}.

Figure 1 has exactly the same implications for the comparison of the test statistics $T_{S_{\text{Kol}},B_{n,h}}$, $T_{S_{\text{AD}},B_{n,h}}$ and $T_{S_{\text{CvM}},B_{n,h}}$ derived from the monitoring process $B_{n,h}(t,x)$.

Example: linear trend Recall that in the linear trend example we have

$$D(t,x;\delta) = \delta^{-1} W_\gamma(t) \left\{ F(x;0) - F(x;\delta) \right\},$$

with

$$W_\gamma(t) = \begin{cases} -\frac{1}{2}\mu(t)\left(1 - \mu(\tau)\right)^2 & \text{if } t \leq \gamma, \\ \frac{1}{2}\left(\mu(t) - \mu(\tau)\right)^2 - \frac{1}{2}\mu(t)\left(1 - \mu(\tau)\right)^2 & \text{if } t > \gamma. \end{cases}$$

Figure 2 displays $e_{S_{\text{Kol}}}/e_{\text{opt}}$, $e_{S_{\text{CvM}}}/e_{\text{opt}}$ and $e_{S_{\text{AD}}}/e_{\text{opt}}$ versus $\mu(t)$. The statistic T_{S_{AD},B_n} shows the strongest performance, and clearly outperforms T_{S_{Kol},B_n}; the efficiency of T_{S_{CvM},B_n} with respect to T_{S_{AD},B_n} is close to 1 for changepoints γ satisfying $\mu(\gamma) \leq 0.5$, but deteriorates fast for $\mu(\gamma) > 0.5$. Similar conclusions hold for $T_{S_{\text{Kol}},B_{n,h}}$, $T_{S_{\text{AD}},B_{n,h}}$ and $T_{S_{\text{CvM}},B_{n,h}}$.

Example: normal contamination Suppose that $F_{ni}(x;\delta)$ satisfies (4), with

$$F_0(x) = \Phi\left(\frac{x - \tau_0}{\rho_0}\right), \quad F_c(x) = \Phi\left(\frac{x - \tau}{\rho}\right),$$

where $\Phi(z)$ is the standardnormal cumulative distribution function. We have

$$e_{B_n} = 4 \sup_{x \in \boldsymbol{R}} \left\{ \Phi\left(\frac{x - \tau}{\rho}\right) - \Phi\left(\frac{x - \tau_0}{\rho_0}\right) \right\}^2. \tag{16}$$

When the kernel function $K(x)$ is taken equal to the standard normal probability density function $\phi(x)$, we may derive that

$$e_{B_{n,h}} = \frac{2\pi h \sup_{x\in\boldsymbol{R}} \left\{ \frac{1}{\sqrt{\rho^2+h^2}} \phi\left(\frac{x-\tau}{\rho}\right) - \frac{1}{\sqrt{\rho_0^2+h^2}} \phi\left(\frac{x-\tau_0}{\rho_0}\right) \right\}^2}{\sup_{x\in\boldsymbol{R}} \frac{1}{\sqrt{2\rho_0^2+h^2}} \phi\left(\frac{x-\tau_0}{\sqrt{\rho_0^2+h^2/2}}\right) - \frac{h}{\rho_0^2+h^2} \phi\left(\frac{x-\tau_0}{\sqrt{\rho_0^2/2+h^2/2}}\right)}. \tag{17}$$

By exploring the ratio $e_{B_n}/e_{B_{n,c\rho_0}}$ [with e_{B_n} and $e_{B_{n,c\rho_0}}$ given by (16) and (17), respectively] numerically for various values of c, we found that setting c equal to 0.75 gives reasonable results for $\rho \leq 1$. Figure 3 evaluates the performance of $T_{S,B_{n,.75\rho_0}}$ relative to T_{S,B_n} by plotting the ratio $e_{B_n}/e_{B_{n,.75\rho_0}}$ versus ρ for $\tau_0 = 0$, $\rho_0 = 1$ and various values of τ. Although the ratio depends on τ, Figure 3 suggests that there exists an upper bound which only depends on ρ. Observe that $T_{S,B_{n,.75\rho_0}}$ outperforms T_{S,B_n} for values of ρ between 0.06 and 1.

As it is quite difficult to obtain similar results in the general situation, we can only rely on the findings in the "normal contamination" example. Fortunately, these findings are in line with expectation: there is an advantage in using the monitoring process $B_{n,h}(t,x)$ when the contamination is reasonably concentrated. This leads us to conjecture that our conclusions extend to the general situation: $T_{S,B_{n,.75\rho_0}}$ [where ρ_0 is now the variance of the observations under the null hypothesis] outperforms T_{S,B_n} when the contamination under the alternative has a reasonably [but not too extremely] concentrated character.

In actual applications we should replace ρ_0 by an estimator. The usual estimator s_{pooled}, defined by

$$s^2_{\text{pooled}} = \frac{1}{m-n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left(X_{i,j} - \bar{X}_i\right)^2,$$

[here $\bar{X}_i$ denotes the mean of the i^{th} sample] becomes degenerate when the observations are individual [that is, $m_i = 1$ for every $i = 1, \ldots, n$]. Alternative variance estimators for individual observations are discussed in Wetherill and Brown (1991), p. 114–121; in particular, we mention the "successive differences" estimator

$$s_{\text{succ-diff}} = \frac{1}{n-1} \sum_{i=1}^{n} \frac{|X_{i,1} - X_{i-1,1}|}{2/\sqrt{\pi}}.$$

[see Kamat (1953)]. Observe that $2/\sqrt{\pi}$ coincides with the "control chart constant" 1.128, often encountered in industrial statistics.

3.5 Bootstrap tests

Let T_n be a test statistic, and T_n^* a bootstrap replication of T_n. The bootstrap test based on T_n employs the distribution of T_n^* to evaluate the achieved significance level of

T_n. Hence, to investigate the asymptotic limiting [as the size of the test tends to zero] Pitman efficiency of the bootstrap test based on T_n in the manner described above, we should require that T_n^* satisfies conditions (a) and (b), and T_n satisfies condition (c) of Definition 2. Likewise, to investigate the i-efficiency of the bootstrap test based on T_n, we should require that T_n^* satisfies condition (a), and T_n satisfies condition (b) of Definition 3.

To usual way of implementing the bootstrap test is to generate a number of bootstrap replications, and count the replications greater than or equal to the achieved value of T_n [cf. Efron and Tibshirani (1993), p. 232]. However, for a bootstrap replication T_n^* satisfying condition (a) and (b) of Definition 2 we may benefit from the fact that its distribution under the null hypothesis approximately has a normal right hand tail. Thus, a normal probability plot of the bootstrap replications should become linear for large values of the normal score. One may interpret the location where the normal probability plot exceeds the attained value of the test statistic as a "z-score" corresponding to the achieved significance level. Determining the achieved significance level of a bootstrap test via a normal probability plot has the advantage that the number of bootstrap replications can be kept relatively low [for instance, in accordance with rule of thumb (2) in paragraph 6.4 in Efron and Tibshirani (1993), p. 52].

Note that both implementations sketched above are scale invariant, in the sense that for every fixed constant $c > 0$ the achieved significance level of the test based on cT_n does not depend on c.

Now, let T_n be equal to [a standardized version of] $T\left(C_n\right)$ or $T\left(C_{n,h}\right)$. In earlier paragraphs we have already discussed how to verify whether T_n satisfies condition (c) of Definition 2 and condition (b) of Definition 3. The theory in Hjort and Koning (2001) with respect to the bootstrap replications C_n^* and $C_{n,h}^*$ of the monitoring processes C_n and $C_{n,h}$ may be used to verify whether condition (a) and (b) of Definition 2 and condition (b) of Definition 3 hold for T_n^*. In general, if condition (a) and (b) of Definition 2 hold for T_n, then they also hold for T_n^*. With respect to the verification of condition (b) of Definition 3 is more complicated, stemming from the fact that the "original" rate $\left((r_n^*)^{1/2} \vee m^{1/2}/\log m\right)$ [appearing in (12) and (17) in Hjort and Koning (2001)] is slightly better from the "bootstrap" rate $\left((r_n^*)^{1/2} \vee m^{1/4}\right)$ [appearing in (18) and (19) in Hjort and Koning (2001)].

Despite this difference in rate, bootstrap tests have clear advantages in applications. Due to the scale invariance of both bootstrap implementations, standardization of the test statistics is not needed [and hence, estimation of ν_0 can be avoided]. Moreover, the achieved significance level can be determined without explicit knowledge with respect to the [asymptotic] distribution of T_n.

4 Applications

In this section we apply the methods of the previous sections to sea water level data, and discuss the patterns of nonconstancy which show up in the monitoring plots. Of particular interest for the interpretation of the monitoring plots are the presence of straight lines [indicating periods of constancy], curvature [indicating periods of gradual change] or angles [indicating moments of abrupt change].

4.1 Sea water levels at Vlissingen, The Netherlands

The sea water level data involve a series of high tide sea water levels at Vlissingen, The Netherlands, starting at January 1, 1882 and ending at December 31, 1985. A total number of 73397 high tide sea water levels were recorded during the measurement period. The data were grouped in 104 subsamples, each covering a one-year period. The pooled standard deviation s of the sea water levels is 39.84 centimeter. The sea water levels ranged from -16 to 455 centimeter, and are displayed in Figure 4. A close inspection of Figure 4 reveals that there are no abrupt changes in the distribution of the sea water levels, but there is a small positive trend.

Figure 5 displays $B_n(\cdot, x)$ for the values of x which correspond to the 25 equidistant horizontal dotted "scan lines" in Figure 4. As test statistic we selected T_{S_{AD}, B_n} [which would have been a logical choice in the presence of advance knowledge that only gradual changes were to be expected]. To evaluate this test statistic, 1000 equidistant scan lines are used. The supremum over x is attained for $x_{\text{opt}} = 199.01$; the solid line in Figure 5 corresponds to $B_n(\cdot, x_{\text{opt}})$. The test statistic T_{S_{AD}, B_n} takes the values 18.978, well exceeding the asymptotic critical values listed in Table 1 in Koning and Protasov (2001). A normal probability plot of 200 bootstrap replications, shows that the value 18.978 of T_{S_{AD}, B_n} is highly significant [an example of such a "bootstrap plot" will be discussed later].

The quadratic shapes in Figure 5 reveal the existence of a linear trend in the data. Note that we should not attach any meaning to the fact that $B_n(t, x_{\text{opt}})$ reaches its maximum value around 1935, as we obviously are not dealing with changepoints here.

Figure 6 displays $B_{n,0.75s}(\cdot, x) = B_{n,29.88}(\cdot, x)$ for the values of x which correspond to the 25 equidistant horizontal dotted "scan lines" in Figure 4. To evaluate the test statistic $T_{S_{\text{AD}}, B_{n,0.75s}}$, 1000 equidistant scan lines were used. The supremum over x is attained for $x_{\text{opt}} = 234.81$; the solid line in Figure 6 corresponds to $B_{n,0.75s}(\cdot, x_{\text{opt}})$.

The test statistic $T_{S_{\text{AD}}, B_{n,0.75s}}$ takes the value 0.3319, which should be compared to the value 0.0646 taken by $\hat{\nu}$. Again, we avoid the problem of limited knowledge with respect to the distribution of $\hat{\nu}^{-1} T_{S_{\text{AD}}, B_{n,0.75s}}$ by resorting to the bootstrap test based on $T_{S_{\text{AD}}, B_{n,0.75s}}$: a normal probability plot of 200 bootstrap replications shows that 0.3319 is a highly significant value of $T_{S_{\text{AD}}, B_{n,0.75s}}$. The quadratic shapes in Figure 6 reveal the existence of a linear trend in the data.

For the sake of completeness, we mention that the respective values 12.205, 8.557, 0.2126 and 0.1494 of T_{S_{Kol}, B_n}, T_{S_{CvM}, B_n}, $T_{S_{\text{Kol}}, B_{n,0.75s}}$ and $T_{S_{\text{CvM}}, B_{n,0.75s}}$ are also highly significant. For these test statistics, the values of x_{opt} are 199.01, 199.01, 234.81 and 234.34, respectively.

4.2 Annual sea water level maxima at Vlissingen, The Netherlands

In this paragraph we study the annual sea water level maxima instead of the original high tide sea water level data. Note that we are now dealing with the "individual observations" situation, where every m_i is equal to 1. The "successive differences" standard deviation s of the sea water level maxima is 27.9384. The annual sea water level maxima ranged from 271 to 455 centimeter, and are displayed in Figure 7.

Figure 8 displays $B_n(\cdot, x)$ for the values of x which correspond to the 25 equidistant horizontal dotted "scan lines" in Figure 7. As test statistic we selected T_{S_{Kol},B_n}. To evaluate this test statistic, 1000 equidistant scan lines are used. The supremum over x is attained for $x_{\text{opt}} = 310.1$; the solid line in Figure 8 corresponds to $B_n(\cdot, x_{\text{opt}})$. The test statistic T_{S_{Kol},B_n} takes the values 1.1776, exceeding the asymptotic critical values listed in Table 1 in Koning and Protasov (2001).

Figure 8 suggests that throughout the first part of the twentieth century the distribution of the yearly sea water level maximum remains relatively constant. Around 1952 there is an abrupt change, after which the distribution remains relatively constant again.

Figure 9 displays $B_{n,0.75s}(\cdot, x) = B_{n,20.95}(\cdot, x)$ for the values of x which correspond to the 25 equidistant horizontal dotted "scan lines" in Figure 7. To evaluate the test statistic $T_{S_{\text{Kol}},B_{n,0.75s}}$, 1000 equidistant scan lines were used. The supremum over x is attained for $x_{\text{opt}} = 294.27$; the solid line in Figure 9 corresponds to $B_{n,0.75s}(\cdot, x_{\text{opt}})$.

The test statistic $T_{S_{\text{Kol}},B_{n,0.75s}}$ takes the value 0.0360, which should be compared to the value 0.0776 taken by $\hat{\nu}$. Again, we avoid the problem of limited knowledge with respect to the distribution of $\hat{\nu}^{-1}T_{S_{\text{AD}},B_{n,0.75s}}$ by resorting to the bootstrap test based on $T_{S_{\text{Kol}},B_{n,0.75s}}$: the normal probability plot of 200 bootstrap replications in Figure 10 shows that 0.0360 is a significant value of $T_{S_{\text{AD}},B_{n,0.75s}}$.

Figure 9 also suggests that throughout the first part of the twentieth century the distribution of the yearly sea water level maximum remains relatively constant. Around 1950 there is an abrupt change, after which the distribution remains relatively constant again.

For the sake of completeness, we mention that the respective values 0.743, 1.666, 0.0216 and 0.0490 of T_{S_{CvM},B_n}, T_{S_{AD},B_n}, $T_{S_{\text{CvM}},B_{n,0.75s}}$ and $T_{S_{\text{AD}},B_{n,0.75s}}$ are also significant. For these test statistics, the values of x_{opt} are 310.1, 310.1, 293.91 and 293.54, respectively.

5 Proofs

This section contains the proofs of Theorem 1, and Lemma's 1, 3, 6, 7, 8 and 9. The proofs of Lemma's 2 and 4 are straightforward, and hence not included. The proofs in this section make use of the technical results collected in Section 5 in Hjort and Koning (2001), and of the DKW-inequality [Dvoretzky, Kiefer and Wolfowitz (1956)]. Below we present the extended version of Bretagnolle (1980) [cf. Inequality 25.1.2 in Shorack and Wellner (1986), p. 797] which allows the random variables $X_1, \ldots, X_m$ to have different distributions. In case these random variables have a common distribution, one may replace $2e \exp\{-2y^2\}$ by $2 \exp\{-2y^2\}$ [cf. Csörgő and Horváth (1993), p. 119].

Inequality 1 (DKW-inequality) *Let $X_1, \ldots, X_m$ be independent random variables, and let $F_\ell(x)$ denote the cumulative distribution function of X_ℓ. Then, for every $y > 0$,*

$$P\left(\sup_{x\in\boldsymbol{R}}\left|m^{-1/2}\sum_{\ell=1}^{m}\left(1_{\{X_\ell\le x\}} - F_\ell(x)\right)\right| \ge y\right) \le 2e\exp\{-2y^2\}.$$

Proof of Theorem 1 We may write

$$\sup_{t\in[0,1]}\sup_{x\in\boldsymbol{R}}\left|A_n(t,x;\delta) - m^{1/2}\delta D(t,x;\delta)\right| \le \Delta_{n1} + \Delta_{n2},$$

where

$$\Delta_{n1} = \sup_{t\in[0,1]}\sup_{x\in\boldsymbol{R}}\left|m^{-1/2}\sum_{i=1}^{[nt]}\sum_{j=1}^{m_i}\left(1_{\{X_{ij}\le x\}} - F_{ni}(x;\delta)\right)\right|,$$

$$\Delta_{n2} = m^{1/2}\delta\sup_{t\in[0,1]}\sup_{x\in\boldsymbol{R}}\left|(m\delta)^{-1}\sum_{i=1}^{[nt]}\sum_{j=1}^{m_i}\left(F_{ni}(x;\delta) - F_{n\cdot}(x;\delta)\right) - D(t,x;\delta)\right|.$$

Combining Inequality 1 and the argument given in the proof of Proposition 1.1.2 in de la Peña and Giné (1999) yields

$$\begin{aligned} P\left(\Delta_{n1} > y\right) &\le 3\max_{k=1,\ldots,m} P\left(\left|m^{-1/2}\sum_{i=1}^{k}\sum_{j=1}^{m_i}\left(1_{\{X_{ij}\le x\}} - F_{ni}(x;\delta)\right)\right| > y/3\right) \\ &\le 3\max_{k=1,\ldots,m} P\left(\left|\left(\sum_{i=1}^{k} m_i\right)^{-1/2}\sum_{i=1}^{k}\sum_{j=1}^{m_i}\left(1_{\{X_{ij}\le x\}} - F_{ni}(x;\delta)\right)\right| > y/3\right) \\ &\le 3\max_{k=1,\ldots,m} 2e\exp\left\{-2y^2\right\} \le 6e\exp\left\{-2y^2\right\}. \end{aligned}$$

Since for every $\epsilon > 0$ there exists C_ϵ such that $P_\delta\left(\Delta_{n1} > C_\epsilon\right) < \epsilon$ for every $\delta > 0$, it follows by one of facts mentioned after Definition 1 that

$$\left(m^{1/2}\delta\right)^{-1}\Delta_{n1} \in \mathcal{C}_a. \qquad (18)$$

Moreover, observe that Condition 1 directly yields that

$$\left(m^{1/2}\delta\right)^{-1}\Delta_{n2} \in \mathcal{C}_a. \qquad (19)$$

Combining (18) and (19) yields (6). This completes the proof of Theorem 1. □

Proof of Lemma 1 Since

$$\begin{aligned}
&\sup_{t\in[0,1]}\sup_{x\in\boldsymbol{R}}\left|B_n(t,x)-m^{1/2}\delta D(t,x;\delta)\right| \\
&\quad\le \sup_{t\in[0,1]}\sup_{x\in\boldsymbol{R}}\left|A_n(t,x;\delta)-m^{1/2}\delta D(t,x;\delta)\right| \\
&\quad\quad+\sup_{t\in[0,1]}\sup_{x\in\boldsymbol{R}}\left|A_n(1,x;\delta)-m^{1/2}\delta D(1,x;\delta)\right|\mu_n(t),
\end{aligned}$$

Lemma 1 is an immediate consequence of Theorem 1. □

Proof of Lemma 3 Recall from the proof of Lemma 3 in Hjort and Koning (2001) that

$$B_{n,h}(t,x)=-h^{-3/2}\int B_n(t,v)K'\left(\frac{v-x}{h}\right)dv.$$

Since

$$\begin{aligned}
&h^{1/2}\sup_{t\in[0,1]}\sup_{x\in\boldsymbol{R}}\left|\left(m^{1/2}\delta\right)^{-1}B_{n,h}(t,x;\delta)-h^{1/2}D_h(t,x;\delta)\right| \\
&\le\left\{\sup_{t\in[0,1]}\sup_{x\in[0,1]}\left|\left(m^{1/2}\delta\right)^{-1}B_n(t,x)-D(t,x;\delta)\right|\right\} \\
&\quad\times h^{1/2}\left\{\sup_{x\in[0,1]}h^{-3/2}\int\left|K'\left(\frac{v-x}{h}\right)\right|dv\right\} \\
&\le\left\{\sup_{t\in[0,1]}\sup_{x\in[0,1]}\left|\left(m^{1/2}\delta\right)^{-1}B_n(t,x)-D(t,x;\delta)\right|\right\}\times\left\{\int|K'(v)|\,dv\right\},
\end{aligned}$$

Condition 3 and (8) together yield (10). This concludes the proof of Lemma 3. □

Proof of Lemma 6 To verify part (a) of Lemma 6, let α denote the size of the test. By conditions (a) and (b) of Definition 2 there exists $n_\alpha,\ y_\alpha>0$ such that $\sup_{P\in\mathcal{P}_0}P\left(T_n>y_\alpha\right)<\alpha$ for $n\ge n_\alpha$. Observe that the actual critical value of T_n does not exceed y_α, and hence $P_{\delta_n}\left(T_n>y_\alpha\right)$ is a lower bound for the power of the test. As $y_\alpha<\frac{2}{3}m^{1/2}b(\delta)$ eventually, it suffices to show that $P_{\delta_n}\left(T_n>\frac{2}{3}m^{1/2}b(\delta)\right)$ tends to 1 as $n\to\infty$. Condition (b) of Definition 3 yields that the right hand side of the inequality

$$P_{\delta_n}\left(T_n>\tfrac{2}{3}m^{1/2}b(\delta_n)\right)\ge 1-P_{\delta_n}\left(\left|\left(m^{1/2}b(\delta_n)\right)^{-1}T_n-1\right|>\tfrac{1}{3}\right)$$

tends to 1 for $n\to\infty$, which completes the proof of Lemma 6(a).

Next, we turn to the proof of Lemma 6(b). Since $\limsup_{n\to\infty}m^{1/2}b(\delta_n)<\infty$, it follows that there exists a constant c such that $\lim_{n\to\infty}P_{\delta_n}\left(T_n>c\right)=0$. By conditions (a) and (b) of Definition 2, it follows that for sufficiently small size of the test, the critical value of the test exceeds c; as $P_{\delta_n}(T_n > c)$ is an upper bound to the power of the test, it follows that the power of the test tends to zero as $n \to \infty$. This completes the proof of Lemma 6(b). □

Proof of Lemma 7 Introduce

$$Y_n = \left(m^{1/2}\tilde{b}(\delta)\right)^{-1}\tilde{T}_n, \quad Y_n' = \nu_\delta/\hat{\nu}.$$

Since both $Y_n - 1$ and $Y_n' - 1$ belong to $\mathcal{C}_a$, it follows by the third fact mentioned after Definition 1 that

$$\left(m^{1/2}b(\delta)\right)^{-1}T_n - 1 = Y_nY_n' - 1 \in \mathcal{C}_a.$$

Since $\lim_{\delta\downarrow 0}\delta^{-1}b(\delta) = \lim_{\delta\downarrow 0}\delta^{-1}\tilde{b}(\delta)/\nu_\delta > 0$ [both $\lim_{\delta\downarrow 0}\nu_\delta$ and $\lim_{\delta\downarrow 0}\delta^{-1}\tilde{b}(\delta)$ are positive], the test statistic T_n satisfies condition (c) of Definition 2. This completes the proof of Lemma 7. □

Before turning to the proof of Lemma 8, we first state and prove the auxillary result Lemma 10.

Lemma 10 *Assume Condition 3 holds, and let $\mathcal{P}^* \subset \mathcal{P}$ be a collection of probability measures. Define $\nu_{\delta,n}$ by $\nu_{\delta,n}^2 = \sup_{x\in\boldsymbol{R}}\tilde{\sigma}_{n,h}(x,x)$, where*

$$\tilde{\sigma}_{n,h}(x,x;\delta) = h^{-1}\int\left\{K\left(\frac{v-x}{h}\right)\right\}^2 f_{n\cdot}(v;\delta)dv - h\left(\tilde{f}_{n,h}(x)\right)^2,$$

$$\tilde{f}_{n,h}(x;\delta) = h^{-1}\int K\left(\frac{v-x}{h}\right)f_{n\cdot}(v;\delta)dv,$$

and $f_{n\cdot}(x;\delta)$ is the derivative of $F_{n\cdot}(x;\delta)$. If there exists a constant c_4 such that $c_4\nu_{\delta,n}^2 \geq 1$ for every $P \in \mathcal{P}^*$, then there exist positive constants c_5–c_7 such that

$$\sup_{P\in\mathcal{P}^*} P\left(h\left|\nu_{\delta,n}/\hat{\nu} - 1\right| > m^{-1/2}y\right) \leq c_5\exp\{-c_6 y\}$$

for $0 < y < (1 \wedge c_7 h)^2 m$.

Proof of Lemma 10 As

$$K\left(\frac{X_{i,j}-x}{h}\right) = -h^{-1}\int 1_{\{X_{i,j}\leq v\}}K'\left(\frac{v-x}{h}\right)dv.$$

[see (29) in Hjort and Koning (2001)], we may write

$$\bar{f}_n(x) = \frac{1}{mh}\sum_{i=1}^{n}\sum_{j=1}^{m_i}K\left(\frac{X_{i,j}-x}{h}\right) = -\frac{1}{mh^2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\int 1_{\{X_{i,j}\leq v\}}K'\left(\frac{v-x}{h}\right)dv.$$

Similarly, we have

$$\hat{\sigma}_{n,h}(x,x) + h\left(\bar{f}_n(x)\right)^2 = \frac{1}{mh}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left\{K\left(\frac{X_{ij}-x}{h}\right)\right\}^2$$
$$= -\frac{2}{mh^2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\int 1_{\{X_{i,j}\le v\}}K\left(\frac{v-x}{h}\right)K'\left(\frac{v-x}{h}\right)dv.$$

Integration by parts yields

$$\tilde{\sigma}_{n,h}(x,x;\delta) + h\left(\tilde{f}_{n,h}(x;\delta)\right)^2 = -2h^{-2}\int F_{n\cdot}(v;\delta)K\left(\frac{v-x}{h}\right)K'\left(\frac{v-x}{h}\right)dv,$$
$$\tilde{f}_{n,h}(x) = -h^{-2}\int F_{n\cdot}(v;\delta)K'\left(\frac{v-x}{h}\right)dv. \tag{20}$$

Introduce

$$\Delta_n = m^{1/2}\sup_{x\in\boldsymbol{R}}\left|m^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i}1_{\{X_{i,j}\le x\}} - F_{n\cdot}(x;\delta)\right|.$$

Since Condition 3 implies

$$\int|K'(u)|\,du < c_2, \quad 2\int K(u)\,|K'(u)|\,du < c_2^2,$$

and $\sup_{x\in\boldsymbol{R}}\tilde{f}_{n,h}(x) \le c_2/h$ by (20), we may write

$$h\left|\hat{\nu}^2 - \nu_{\delta,n}^2\right| \le h\sup_{x\in\boldsymbol{R}}\left|\hat{\sigma}_{n,h}(x,x) - \tilde{\sigma}_{n,h}(x,x;\delta)\right|$$
$$\le 2c_2\Delta_{n1} + (\Delta_{n1})^2 + \Delta_{n2},$$
$$\le 3c_2\Delta_{n1} + \Delta_{n2} \quad \text{if} \quad \Delta_{n1} \le c_2,$$

where

$$\Delta_{n1} = h\sup_{x\in\boldsymbol{R}}\left|\bar{f}_n(x) - \tilde{f}_{n,h}(x;\delta)\right|$$
$$\le \sup_{x\in\boldsymbol{R}}\left|h^{-2}\int\left(\frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{m_i}1_{\{X_{i,j}\le v\}} - F_{n\cdot}(v;\delta)\right)K'\left(\frac{v-x}{h}\right)dv\right|$$
$$\le c_2 m^{-1/2}\Delta_n,$$

and

$$\Delta_{n2} = h\sup_{x\in\boldsymbol{R}}\left|\left\{\hat{\sigma}_{n,h}(x,x) + h\left(\bar{f}_n(x)\right)^2\right\} - \left\{\tilde{\sigma}_{n,h}(x,x) + h\left(\tilde{f}_{n,h}(x)\right)^2\right\}\right|$$
$$\le \sup_{x\in\boldsymbol{R}}\left|2h^{-2}\int\left(\frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{m_i}1_{\{X_{i,j}\le v\}} - F_{n\cdot}(v;\delta)\right)K\left(\frac{v-x}{h}\right)K'\left(\frac{v-x}{h}\right)dv\right|$$
$$\le c_2^2 m^{-1/2}\Delta_n.$$

Hence, we obtain

$$h\left|\hat{\nu}^2-\nu_{\delta,n}^2\right| \leq (2c_2)^2 m^{-1/2}\Delta_n \quad \text{if} \quad \Delta_n \leq m^{1/2}. \tag{21}$$

Since $(2c_4)^2\left|\hat{\nu}^2-\nu_{\delta,n}^2\right| \leq 3$ implies

$$\left|(\hat{\nu}/\nu_{\delta,n})^2-1\right| \leq (\nu_{\delta,n})^{-2}\left|\hat{\nu}^2-\nu_{\delta,n}^2\right| \leq c_4^2\left|\hat{\nu}^2-\nu_{\delta,n}^2\right| \leq \tfrac{3}{4},$$

and $\left|x^{-1/2}-1\right|$ is bounded by $4\,|x-1|$ for $x \geq \frac{1}{4}$, it follows that

$$|\nu_{\delta,n}/\hat{\nu}-1| \leq 4\left|(\hat{\nu}/\nu_{\delta,n})^2-1\right| \leq (2c_4)^2\left|\hat{\nu}^2-\nu_{\delta,n}^2\right| \quad \text{if} \quad (2c_4)^2\left|\hat{\nu}^2-\nu_{\delta,n}^2\right| \leq 3. \tag{22}$$

Combining (21) and (22) yields

$$h\,|\nu_{\delta,n}/\hat{\nu}-1| \leq (4c_2c_4)^2 m^{-1/2}\Delta_n \quad \text{if} \quad m^{-1/2}\Delta_n \leq 1 \wedge 3h(4c_2c_4)^{-2}.$$

As

$$\sup_{P\in\mathcal{P}} P\left(\Delta_n > y\right) \leq 2e\exp\left\{-2y^2\right\}$$

for every $y > 0$ by Inequality 1, Lemma 10 now follows [take $c_5 = 2e$, $c_6 = 2/(4c_2c_4)^2$ and $c_7 = 3/(4c_2c_4)^2$]. □

Proof of Lemma 8 Define $\nu_{\delta,n}$ and $\tilde{\sigma}_{n,h}(x,x)$ as in Lemma 10, and define ν_δ by $\nu_\delta^2 = \sup_{x\in\boldsymbol{R}}\sigma_{\delta,h}(x,x)$, where

$$\sigma_{\delta,h}(x,x) = h^{-1}\left(-2h^{-1}\int\left(F^*(v;\delta)\right)K\left(\frac{v-x}{h}\right)K'\left(\frac{v-x}{h}\right)dv\right) + \left\{h^{-1}\int\left(F^*(v;\delta)\right)K'\left(\frac{v-x}{h}\right)dv\right\}^2\Bigg).$$

Observe that (15) yields that ν_δ tends to ν_0 as δ tends to zero. Thus, we may choose $\delta^{**} > 0$ so as to satisfy $\nu_\delta^{-1} \leq 2c_8$ for $0 < \delta < \delta^{**}$.

As $\nu_\delta/\hat{\nu}$ may be written as the product of $Y_n = \nu_\delta/\nu_{\delta,n}$ and $Y_n' = \nu_{\delta,n}/\hat{\nu}$, it suffices to show that both $|Y_n-1|$ and $|Y_n-1|$ belong to $\mathcal{C}_a$.

We start by showing that $|\nu_{\delta,n}/\hat{\nu}-1| \in \mathcal{C}_a$. Applying Lemma 10 with $\mathcal{P}^* = \{P_\delta : 0 < \delta < \delta^{**}\}$ and $c_4 = 2c_8$, yields that for every $\epsilon > 0$ there exists C_ϵ such that

$$P_\delta\left(m^{1/2}\delta\,|\nu_{\delta,n}/\hat{\nu}-1| > C_\epsilon\right) \leq P_\delta\left(m^{1/2}\delta^{**}\,|\nu_{\delta,n}/\hat{\nu}-1| > C_\epsilon\right) < \epsilon$$

for every $0 < \delta < \delta^{**}$. It follows by the last fact mentioned after Definition 1 that $|\nu_{\delta,n}/\hat{\nu}-1|$ belongs to the class $\mathcal{C}_a$.

Next, we turn to verifying that $|\nu_\delta/\nu_{\delta,n}-1| \in \mathcal{C}_a$. Since

$$h^{-1}\int F_{n\cdot}(v)K'\left(\frac{v-x}{h}\right)dv \leq c_2,$$

we may write

$$\left|\nu_{\delta,n}^2 - \nu_\delta^2\right| \le \frac{1}{h}\left\{2c_2\Delta_{n1} + (\Delta_{n1})^2 + \Delta_{n2}\right\},$$

where

$$\begin{aligned}\Delta_{n1} &= \sup_{x\in\boldsymbol{R}}\left|\frac{1}{h}\int \left(F_{n\cdot}(x;\delta) - F^*(x;\delta)\right) K'\left(\frac{v-x}{h}\right) dv\right| \\ &\le c_2 \sup_{x\in\boldsymbol{R}} \left|F_{n\cdot}(x;\delta) - F^*(x;\delta)\right|,\end{aligned}$$

$$\begin{aligned}\Delta_{n2} &= \sup_{x\in\boldsymbol{R}}\left|\frac{2}{h}\int \left(F_{n\cdot}(x;\delta) - F^*(x;\delta)\right) K\left(\frac{v-x}{h}\right) K'\left(\frac{v-x}{h}\right) dv\right| \\ &\le c_2^2 \sup_{x\in\boldsymbol{R}} \left|F_{n\cdot}(x;\delta) - F^*(x;\delta)\right|.\end{aligned}$$

Observe that (2) yields that both Δ_{n1} and Δ_{n2} belong to $\mathcal{C}_a$, which yields $\left|\nu_{\delta,n}^2 - \nu_\delta^2\right| \in \mathcal{C}_a$. Since $(4c_8)^2\left|\nu_\delta^2 - \nu_{\delta,n}^2\right| \le 3$ implies

$$\left|(\nu_{\delta,n}/\nu_\delta)^2 - 1\right| \le \nu_\delta^{-2}\left|\nu_\delta^2 - \nu_{\delta,n}^2\right| \le 4c_8^2\left|\nu_\delta^2 - \nu_{\delta,n}^2\right| \le \tfrac{3}{4},$$

and $\left|x^{-1/2} - 1\right|$ is bounded by $4\left|x-1\right|$ for $x \ge \frac{1}{4}$, it follows that

$$\left|\nu_\delta/\nu_{\delta,n} - 1\right| \le 4\left|(\nu_{\delta,n}/\nu_\delta)^2 - 1\right| \le (4c_8)^2\left|\nu_\delta^2 - \nu_{\delta,n}^2\right| \quad \text{if} \quad (4c_8)^2\left|\nu_\delta^2 - \nu_{\delta,n}^2\right| \le 3.$$

Hence, $\left|\nu_\delta^2 - \nu_{\delta,n}^2\right| \in \mathcal{C}_a$ implies $\left|\nu_\delta/\nu_{\delta,n} - 1\right| \in \mathcal{C}_a$, which completes the proof of Lemma 8. □

Proof of Lemma 9 Let t^* be the value of t that maximizes $\left\{\int_0^t L(s)w(s)d\mu(s)\right\}^2$. Then e_1 is bounded by

$$\frac{\left\{\int_0^{t^*} L(s)w(s)d\mu(s)\right\}^2}{\left\{\int_0^{t^*} (L(s))^2\, d\mu(s) - \left(\int_0^{t^*} L(s)d\mu(s)\right)^2\right\}}.$$

By a Cauchy-Schwarz argument as in the Appendix of Hjort and Koning (2002), it follows that e_1 is bounded by

$$\int_0^{t^*} \{w(s)\}^2\, d\mu(s) \le \int_0^1 \{w(s)\}^2\, d\mu(s).$$

Since

$$\int_0^t (w(s))^2\, d\mu(s) - \{W(t)\}^2 \le \int_0^1 (w(s))^2\, d\mu(s),$$

it follows that

$$\frac{\sup_{t\in[0,1]}\left\{\int_0^t (w(s))^2\,d\mu(s)\right\}^2}{\sup_{t\in[0,1]}\left\{\int_0^t (w(s))^2\,d\mu(s)-(W(t))^2\right\}}=\int_0^1 (w(s))^2\,d\mu(s),$$

which concludes the proof of Lemma 9. □

Acknowledgements

The Vlissingen sea water level data set used in this paper was kindly made available to us by the *National Institute for Coastal and Marine Management* [RIKZ], a department of the Dutch Ministry of Transport, Public Works and Water Management.

Nils Lid Hjort is grateful to the Tinbergen Institute at the Erasmus University Rotterdam for hospitality and financial support. Alex J. Koning is grateful for hospitality and financial support in connection with visits to the Department of Mathematics at the University of Oslo.

References

[1] Bretagnolle, J. (1980). Statistique de Kolmogorov-Smirnov pour un enchantillon nonequireparti. *Colloq. internat. CNRS* **307**, 39–44.

[2] Csörgő, M., Horváth, L. (1993). *Weighted approximations in probability and statistics*. Wiley, New York.

[3] Csörgő, M., Horváth, L. (1997). *Limit theorems in change-point analysis*. Wiley, New York.

[4] de la Peña, V.H., Giné, E. (1999). *Decoupling. From dependence to independence.* Springer-Verlag, New York.

[5] Dvoretzky, A., Kiefer, J., Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Annals of Mathematical Statistics* **27**, 642–669.

[6] Efron, B., Tibshirani, R.J. (1993). *An introduction to the bootstrap*. Chapman and Hall, New York.

[7] Hjort, N.L., Koning, A.J. (2001). Constancy of distributions: nonparametric monitoring of probability distributions over time. *EI Report 2001-50*. `http://www.eur.nl/WebDOC/doc/econometrie/feweco20020115123608.pdf`.

[8] Kallenberg, W.C.M. (1983). Intermediate efficiency, theory and examples. *Annals of Statistics* **11**, 170-182.

[9] Kallenberg, W.C.M., Koning, A.J. (1995). On Wieand's theorem. *Statistics & Probability Letters* **25**, 121–132.

[10] Kamat, A.R. (1953). On the mean successive difference and its ratio to the root mean square. *Biometrika* **40**, 116–127.

[11] Koning, A.J., Protasov, V. (2001). Tail behaviour of Gaussian processes with applications to the Brownian pillow. *EI Report 2001-49*. `http://www.eur.nl/WebDOC/doc/econometrie/feweco20020115122231.pdf`.

[12] Nikitin, Ya. (1995). *Asymptotic efficiency of nonparametric tests*. Cambridge University Press.

[13] Rubin, H., Sethuraman, J. (1965). Bayes risk efficiency. *Sankhyā Ser. A* **27**, 347–356.

[14] Shorack, G.R., Wellner, J.A. (1986). *Empirical processes with applications to statistics*. Wiley, New York.

[15] Wetherill, G.B, Brown, D.W. (1991). *Statistical process control. Theory and practice*. Chapman and Hall, London.

[16] Wieand, H.S. (1976). A condition under which the Pitman and Bahadur approaches to efficiency coincide. *Annals of Statistics* **4** , 1003-1011.

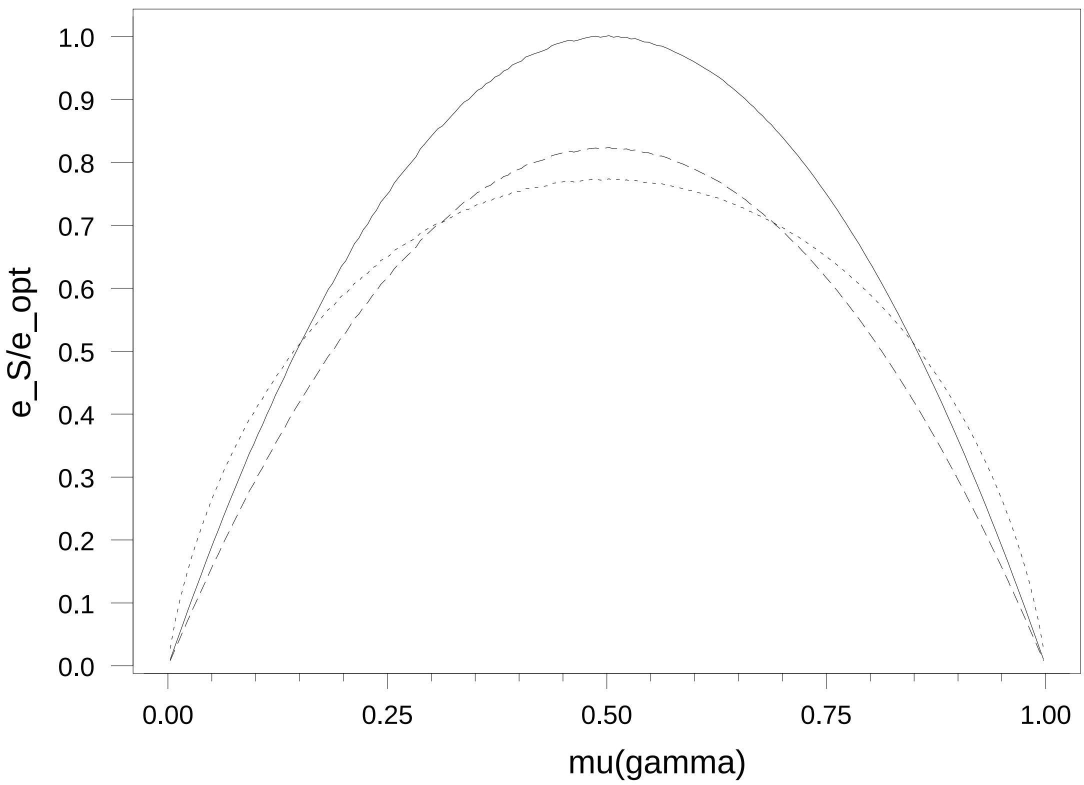

Figure 1: Plots of relative efficiency e_S/e_{opt} versus $\mu(\gamma)$ in "at most one" changepoint example, where S is S_{Kol} [solid line], S_{CvM} [dashed line] or S_{AD} [dotted line]. Here γ denotes the location of the abrubt change. Test statistics involving S_{Kol} are superior to test statistics involving S_{CvM}. Test statistics involving S_{AD} outperform test statistics involving S_{Kol} if $\mu(\gamma) \leq 0.150$ or $\mu(\gamma) \geq 0.850$.

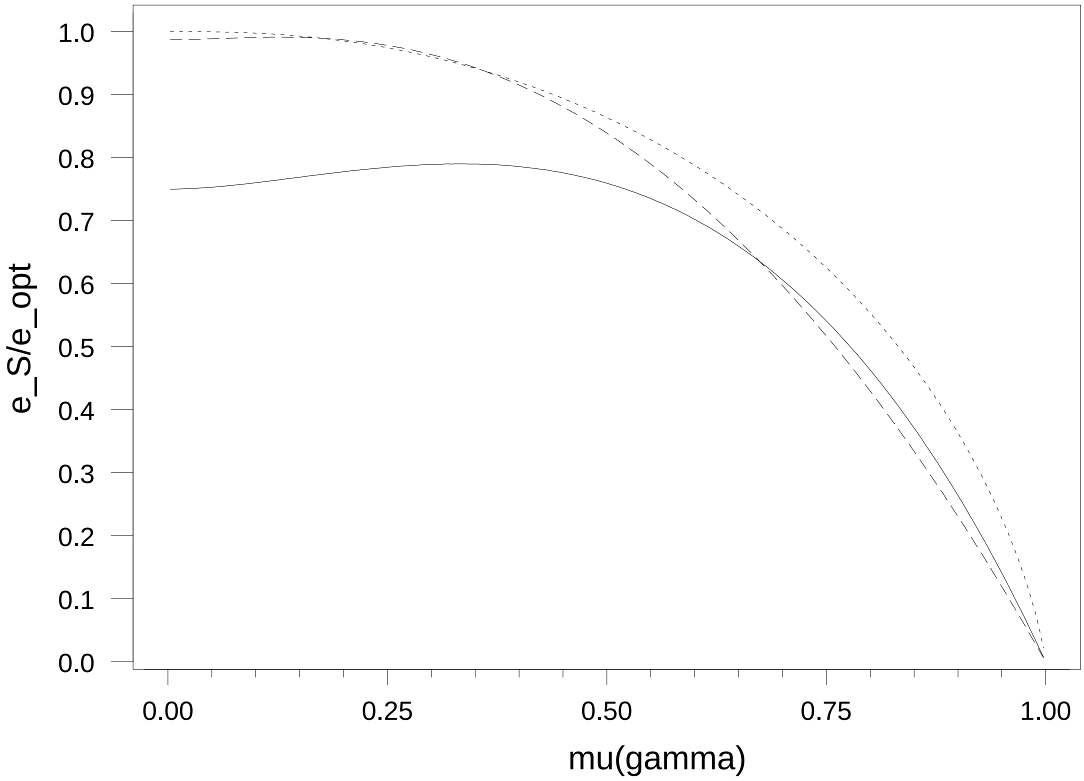

Figure 2: Plots of relative efficiency e_S/e_{opt} versus $\mu(\gamma)$ in linear trend example, where S is either S_{Kol} [solid line], S_{CvM} [dashed line] or S_{AD} [dotted line]. Here γ denotes the location where the linear trend first emerges. Test statistics involving S_{AD} are superior to test statistics involving S_{Kol}, and outperform test statistics involving S_{CvM} if either $\mu(\gamma) \leq 0.175$ or $\mu(\gamma) \geq 0.360$.

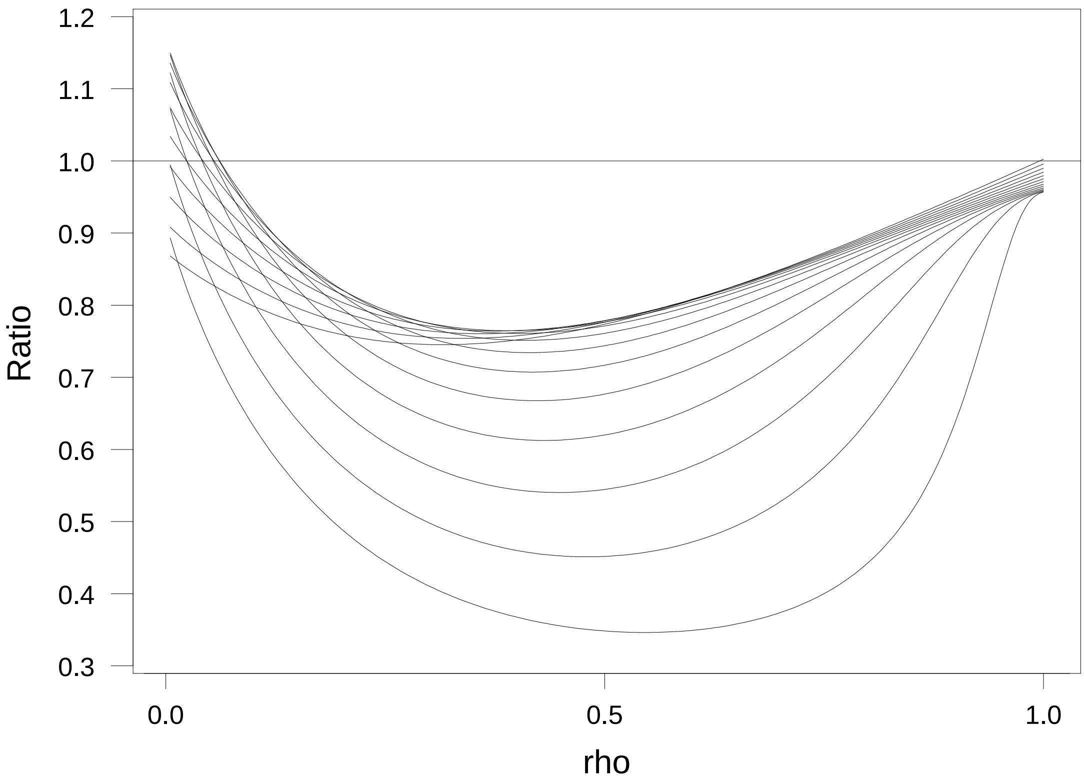

Figure 3: Plots of the ratio $e_{B_n}/e_{2,.75\sigma}$ versus ρ for $\tau = 0.1, 0.2, \ldots, 1.4$. The lowest curve corresponds to $\tau = 0.1$. For values of ρ between 0.06 and 1, test statistics derived from the monitoring process $B_{n,0.75\sigma}(t, x)$ are more efficient than their counterparts derived from $B_n(t, x)$.

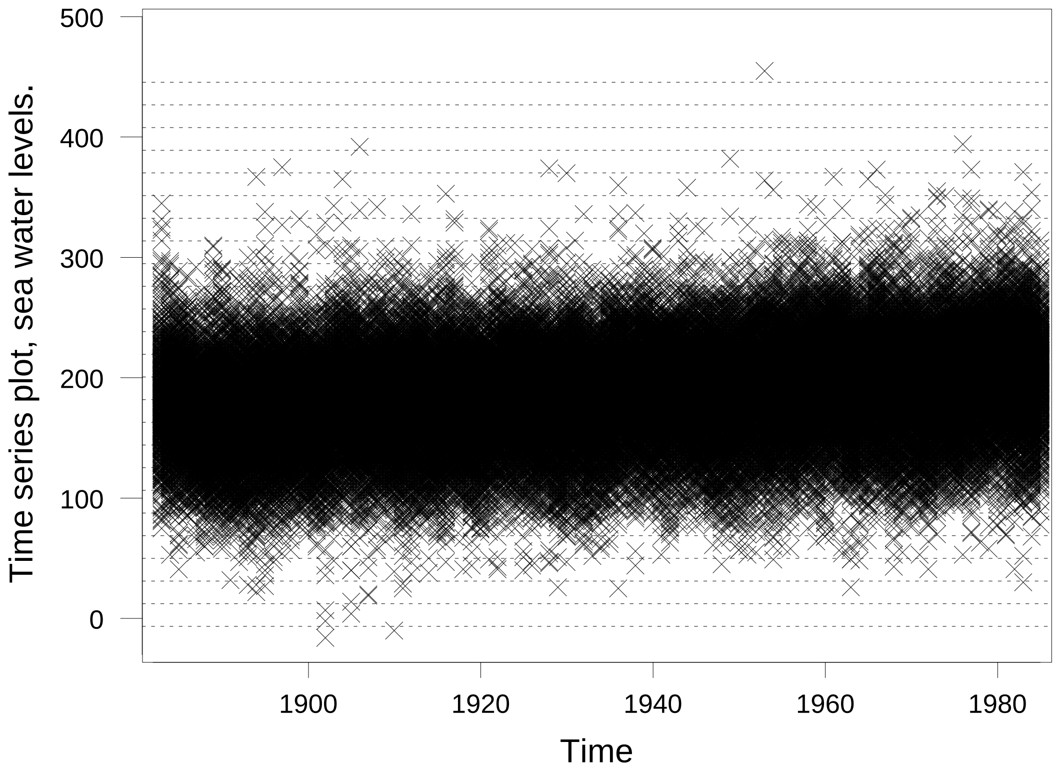

Figure 4: Time series plot of high tide sea water level at Vlissingen, The Netherlands. A total number of 73397 high tide sea water levels were recorded during the measurement period starting at January 1, 1882 and ending at December 31, 1985. The data were grouped in 104 subsamples, each covering one calendar year.

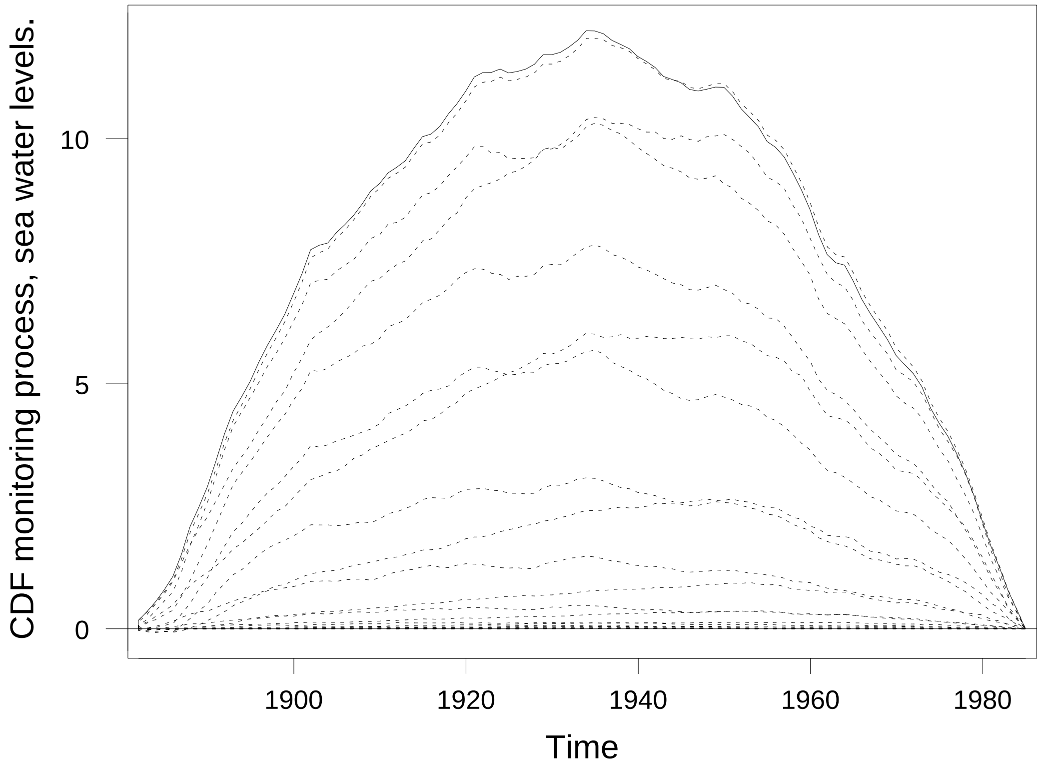

Figure 5: The monitoring process $B_n(t,x)$ for fixed x, sea water level data. The dotted lines and the solid line are the results of "scanning" the monitoring process $B_n(t,x)$ along the dotted lines in Figure 4 and the line $x = x_{\text{opt}} = 199.01$, respectively. The quadratic shapes reveal the existence of a linear trend in the cumulative distribution function of the sea water levels.

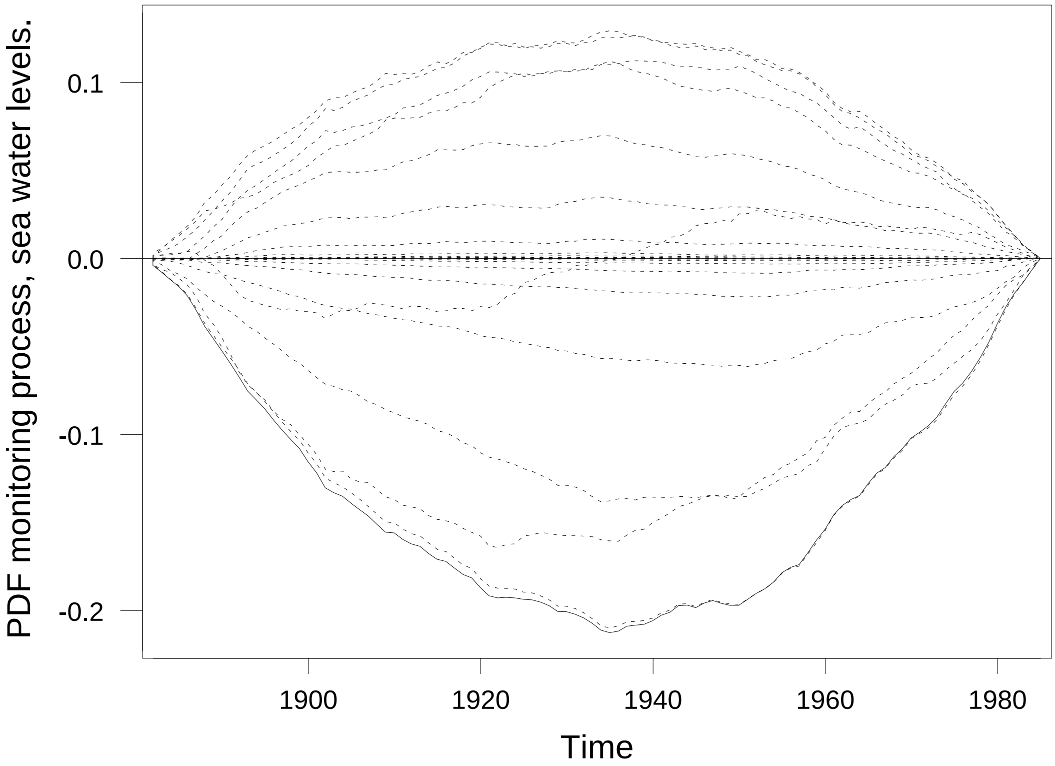

Figure 6: The monitoring process $B_{n,0.75s}(t,x)$ for fixed x, sea water level data. The dotted lines and the solid line are the results of “scanning” the monitoring process $B_{n,0.75s}(t,x)$ along the dotted lines in Figure 4 and the line $x = x_{\text{opt}} = 234.89$, respectively. The quadratic shapes reveal the existence of a linear trend in the probability density function of the sea water levels.

Figure 7: Time series plot of the annual sea water level maxima at Vlissingen, The Netherlands, 1882–1985. Clearly visible are two important events in the Dutch fight against the arch-enemy: the “watersnoodramp” of 1953 caused 1835 deaths in a flooded area of around 1500 square kilometres in the south-western part of the Netherlands; the storm surge of 1916 caused huge damage to the surroundings of the Zuider Zee. These two national disasters prompted the construction of the Delta works and the IJsselmeer causeway, respectively.

Figure 8: The monitoring process $B_n(t,x)$ for fixed x, annual sea water level maxima. The dotted lines and the solid line are the results of "scanning" the monitoring process $B_n(t,x)$ along the dotted lines in Figure 7 and the line $x = x_{\text{opt}} = 310.1$, respectively. The angular shapes around 1952 suggest the existence of an abrupt change in the cumulative distribution function of the annual sea water level maxima.

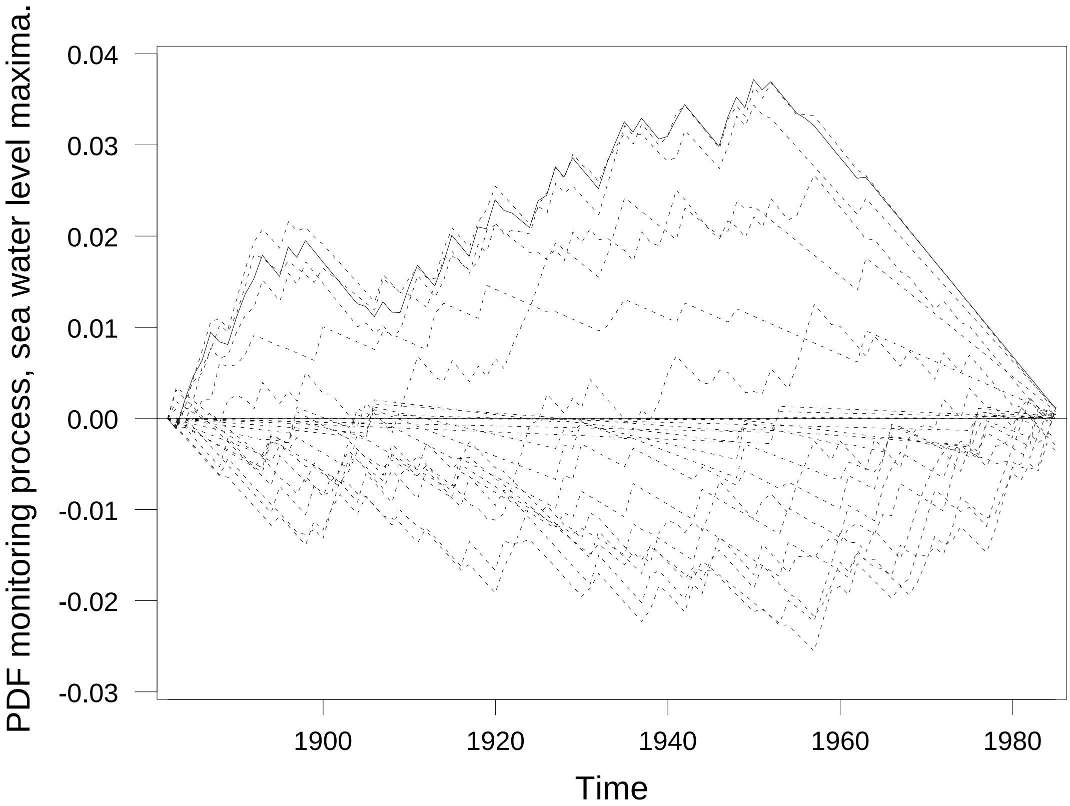

Figure 9: The monitoring process $B_{n,0.75s}(t,x)$ for fixed x, annual sea water level maxima. The dotted lines and the solid line are the results of "scanning" the monitoring process $B_{n,0.75s}(t,x)$ along the dotted lines in Figure 7 and the line $x = x_{\text{opt}} = 294.28$, respectively. The angular shapes around 1950 in the upper part of the plot [which corresponds to the lower sea water levels] suggest the existence of an abrupt change in the probability density function of the annual sea water level maxima. In the lower part of the plot the value of 0.02552 is attained, which is just significant at the 5% level according to the "bootstrap plot" in Figure 10; the quadratic shapes in the lower part [which corresponds to the higher sea water levels] suggest the existence of linear trend in the probability density function.

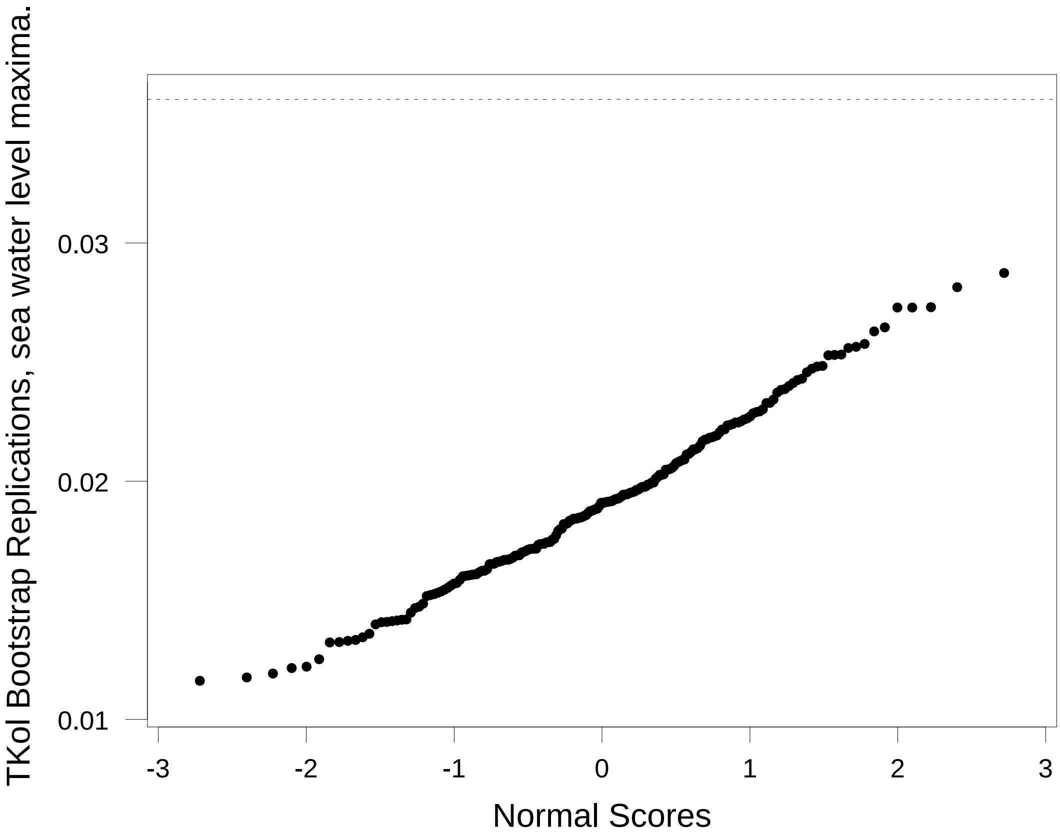

Figure 10: Normal probability plot of 200 bootstrap replications of $T_{S_{\text{Kol}},B_{n,0.75s}}$, annual sea water level maxima. The critical value of $T_{S_{\text{Kol}},B_{n,0.75s}}$ at the 5% level is estimated to be 0.02548, the "value" of the normal probability plot corresponding to a normal score of 1.645. The dotted line indicates the value 0.0360 taken by the test statistic. According to the theory in paragraph 3.5, the normal probability plot should become linear for larger values of the normal score. As one may interpret the location where the normal probability plot exceeds 0.0360 as an estimate of the "z-score" corresponding to the attained significance level, the plot shows that 0.0360 is indeed a highly significant value of T_{S_{AD},B_n}.