Supplement to
Heterogeneous Change Point Inference
by Florian Pein, Hannes Sieling and Axel Munk

Appendix A. Computation

In this section we detail the computation of the estimator H-SMUCE (Section A.1) and of the critical values $q_1, \ldots, q_{d_n}$ (Section A.2). We also examine the computation time (Section A.3) theoretically and empirically. An R-package is available online[1].

A.1. **Computation of the estimator.** First of all, we obtain from the multiscale test the bounds

$$[\underline{b}_{i,j}, \overline{b}_{i,j}] := \left[\overline{Y}_{ij} - \sqrt{\frac{q_{ij}\hat{s}_{ij}^2}{j-i+1}}, \overline{Y}_{ij} + \sqrt{\frac{q_{ij}\hat{s}_{ij}^2}{j-i+1}}\right] \tag{A.1}$$

for μ on the interval $[i/n, j/n] \in \mathcal{D}$. Therefore, H-SMUCE can be computed as in (Frick et al., 2014, Section 3) for SMUCE described. However, in what follows we give a modification of the algorithm which reduces the computation time remarkably due to the small number of intervals $\mathcal{O}(n)$ in the dyadic partition $\mathcal{D}$. Here, we compute first left and right limits for the location of the change-points and then start the dynamic program restricted to these intervals. A notable difference to (Killick et al., 2012; Frick et al., 2014) is that this approach leads also to pruning in the forward step of the dynamic program. More precisely, we define the intersected bounds as

$$\underline{B}_{i,j} := \max_{\substack{i \leq s < t \leq j \\ [s/n, t/n] \in \mathcal{D}}} \underline{b}_{s,t} \text{ and } \overline{B}_{i,j} := \min_{\substack{i \leq s < t \leq j \\ [s/n, t/n] \in \mathcal{D}}} \overline{b}_{s,t}$$

and set recursively

$$L_k := \min\left\{1 < r \leq L_{k+1} - 1 \ : \ \underline{B}_{r, L_{k+1}-1} \leq \overline{B}_{r, L_{k+1}-1}\right\},$$

for $k = \hat{K}, \ldots, 1$, with $L_{\hat{K}+1} := n + 1$. The right limits are defined as

$$R_k := \min\left\{R_{k-1} < r \leq n \ : \ \underline{B}_{R_{k-1}, r} > \overline{B}_{R_{k-1}, r}\right\},$$

for $k = 1, \ldots, \hat{K}$, with $R_0 := 1$. In other words, the left limit for the k-th change-point L_k is the smallest number $1 < r \leq n$ such that between Y_r and Y_n a piecewise constant solution with $\hat{K} - k$ change-points exists which respects the bounds (A.1). Analogously, the right limit R_k is the smallest number $1 < r \leq n$ such that between Y_1 and Y_r no piecewise constant solution with $k - 1$ change-points exists which fulfils the bounds (A.1). Note, that we do not have to compute the right limits separately, since we can just start the dynamic program at L_k and stop if another change-point has to be included. It follows that the k-th change-point $\hat{\tau}_k$ has to be in the confidence interval $[L_k/n, R_k/n]$, since otherwise an additional change-point would be necessary to fulfil the multiscale constraints.

[1] `http://www.stochastik.math.uni-goettingen.de/hsmuce`

A.2. **Computation of the critical values.** In this section we show how the critical values can be computed by Monte-Carlo simulations. Note first that the following method uses only the continuity and the monotonicity of the cumulative distribution functions of the statistics $T_1, \dots, T_{d_n}$ and therefore the methodology can also be used for other multiscale tests, see for instance the extension to other interval sets in Remark 2.2.

Let M be the number of simulations and $(T_{1,1}, \dots, T_{d_n,1}), \dots, (T_{1,M}, \dots, T_{d_n,M})$ be i.i.d. copies of the vector $(T_1, \dots, T_{d_n})$. Moreover, we denote by $F_M(\cdot)$ the empirical distribution function of $(T_1, \dots, T_{d_n})$ and by $F_{M,k}(\cdot)$ the empirical distribution function of the random variable T_k. Then, we aim to find a vector of critical values $\widehat{\mathbf{q}}_M = (\widehat{q}_{M,1}, \dots, \widehat{q}_{M,d_n})$ which satisfies with

$$\alpha - \frac{1}{M} < 1 - F_M\left(\widehat{\mathbf{q}}_M\right) \leq \alpha, \tag{A.2}$$

an empirical version of condition (2.3), and with

$$\frac{1 - F_{M,j_1}(\widehat{q}_{M,j_1})}{\beta_{j_1}} \leq \frac{1 - F_{M,j_2}(\widehat{q}_{M,j_2}) + \frac{1}{M}}{\beta_{j_2}} \quad \text{for all} \quad j_1, j_2 \in \{1, \dots, d_n\}, \tag{A.3}$$

an empirical version of condition (2.5). In the following we propose an iterative method to determine such a vector and show afterwards that this vector converges almost surely to the vector of critical values defined by (2.3) and (2.5). As the k-th entry of the starting vector we choose the empirical $(1-\alpha\beta_k)$-quantile of the statistic T_k, since the vector with these values satisfies condition (A.3) and the inequality

$$1 - F_M\left(\cdot\right) \leq \alpha.$$

Afterwards, we reduce the entries until the lower bound from condition (A.2) is satisfied, too. To ensure condition (A.3) in every iteration, we always reduce the entry which has the smallest ratio

$$\frac{1 - F_{M,k}(\widehat{q}_{M,k})}{\beta_k}.$$

In Algorithm 1 the determination of the critical values is summarized in pseudocode.

The method has the advantage that we do not need specific assumptions on the distribution of the vector $(T_1, \dots, T_{d_n})$ and still get critical values which are adapted to the exact finite sample distribution of $(T_1, \dots, T_{d_n})$ and ensure therefore even for a finite number of observations the significance level α.

The following theorem shows the convergence of this algorithm to $\mathbf{q} = (q_1, \dots, q_{d_n})$.

Theorem A.1 (Consitency of Monte-Carlo critical values)**.** *The empirical vector of critical values $\widehat{\boldsymbol{q}}_M = (\widehat{q}_{M,1}, \dots, \widehat{q}_{M,d_n})$ converges almost surely in the number of simulations M to the vector of critical values $\boldsymbol{q} = (q_1, \dots, q_{d_n})$ defined by* (2.3) *and* (2.5)*.*

The computation time is dominated by the generation of the M i.i.d. copies of the vector $(T_1, \dots, T_{d_n})$. Therefore, we store the generated realizations and recycle them. To avoid memory problems we only store the realizations for every dyadic number, because the significance level α is still satisfied if we determine the critical values based on realizations with a larger number of observations, since then the maxima in $(T_1, \dots, T_{d_n})$ are taken

Algorithm 1 Determination of the critical values.

Input: The statistics $T_1, \ldots, T_{d_n}$ as well as the significance level $\alpha \in (0,1)$, the weights $\beta_1, \ldots, \beta_{d_n} > 0$, with $\sum_{k=1}^{d_n} \beta_k = 1$, and the number of simulations $M \in \mathbb{N}$.

Output: The vector of critical values $\widehat{\mathbf{q}}_M = (\widehat{q}_{M,1}, \ldots, \widehat{q}_{M,d_n})$ which fulfils the conditions (A.2) and (A.3).

for $i = 1, \ldots, M$ **do**
  $(T_{1,i}, \ldots, T_{d_n,i}) \leftarrow$ realisation of $(T_1, \ldots, T_{d_n})$
end for
for $k = 1, \ldots, d_n$ **do**
  $(S_{k,1}, \ldots, S_{k,M}) \leftarrow \mathrm{sort}\,((T_{k,1}, \ldots, T_{k,M}))$
  $w_k \leftarrow M - \lfloor \alpha \beta_k M \rfloor$
end for
repeat
  $\hat{k} \leftarrow \mathrm{argmin}_{k=1,\ldots,d_n} \beta_k^{-1} (1 - F_{M,k}(S_{k,w_k}))$
  $w_{\hat{k}} \leftarrow w_{\hat{k}} - 1$
until $1 - F_M\left(S_{1,w_1}, \ldots, S_{m,w_{d_n}}\right) > \alpha$
$w_{\hat{k}} \leftarrow w_{\hat{k}} + 1$
return $S_{1,w_1}, \ldots, S_{m,w_{d_n}}$

over more intervals. To this end, the choice $M = 10\,000$ seems to be a good trade-off between computation time and approximation accuracy.

A.3. **Computation time.** In this section we discuss the theoretical computation time of H-SMUCE and compare it later in simulations with CBS, cumSeg and LOOVF. We stress that the computation time for the bounds, for the limits $L_1, \ldots, L_{\hat{K}}$ (and so for $\hat{K}$) and for the optimization problem (1.4), and therefore of all confidence sets, is always $\mathcal{O}(n)$. Hence, the computation time is dominated by the determination of the restricted maximum likelihood estimator by dynamic programming.

Lemma A.2 (Computation time)**.** *The algorithm has data depended computation time*

$$\mathcal{O}\left(n + \sum_{k=1}^{\hat{K}-1} (R_k - L_k + 1)(R_{k+1} - L_{k+1} + 1)\right). \tag{A.4}$$

This can be bounded by $\mathcal{O}(n^2)$ in the worst case, but the computation time is in many cases much smaller. In particular, if the signal to noise ratios are large enough such that the change-points are easy to detect, i.e. $R_k - L_k$ is small. This is for instance the case for a fixed signal, where $R_k - L_k$ stays more or less constant. More precisely, by combining (A.4) with equation (3.10) we see that with probability tending to one the computation time of H-SMUCE is even linear, if $\alpha_n \to 0$, but $n^{-\frac{1}{2}} \log((\alpha_n \beta_{k_n,n})^{-1}) \to 0$. In comparison to the computation time of SMUCE, see (Sieling, 2013, (4.3)), which is dominated by the term

$$\mathcal{O}\left(\sum_{k=1}^{\hat{K}-1} (R_k - R_{k-1})(R_{k+1} - R_k)\right),$$

we see that the computation time is further reduced. In particular, if no change-point is present the computation time is $\mathcal{O}(n)$ instead of $\mathcal{O}(n^2)$. The computation time is also

$\mathcal{O}(n)$ if the number of change-points increases linear in the number of observations and the change-points are evenly enough distributed.

In the following we examine the computation time empirically in a similar simulation study as in (Maidstone and Pickering, 2014). More precisely, we generate data with varying number of observations n and equidistant change-points. Thereby, we consider $K = 10$, $K = \sqrt{n}$ and $K = n/100$. In all scenarios we choose the values of the mean and the standard deviation function randomly like in Section 4, once again with $C = 200$. All simulations are repeated 100 times and terminated after ten seconds. The simulations were performed on a single core system with 1.8 GHz and 8 GB RAM in a 64-bit OS.

We fix the significance level $\alpha = 0.1$ as well as the weights $\beta_1 = \cdots = \beta_{d_n} = 1/d_n$ and compare H-SMUCE with CBS, LOOVF and cumSeg. Note, that we restore the Monte-Carlo simulations at the first use to reduce further loading times, here we only take the already restored simulations into account. Furthermore, we set for cumSeg the maximal number of change-points $k = \max(2K, 10)$, since for the default parameter $k = \min(30, n/10)$ the program requires manual increase of k for many simulations runs. Note, that the choice above already incorporates prior knowledge about the true signal. We stress (not displayed) that the computation time (and the required memory space) increases severely in the parameter k.

From Figure 8 we draw that H-SMUCE is much faster than the other methods, in particular if the number of change-points increases. For $K = n/100$ the computation time increases almost linearly in the number of observations. For example, when $n = 10^7$ it is still less than a minute. The second shortest computation time has CBS for larger numbers of observations, whereas cumSeg is superior for smaller numbers of observations. The computation time of CBS for $n = 10^5$ observations is still less than a minute in all scenarios, whereas cumSeg has a similar computation time for $K = 10$, but lasts several minutes in the other cases. Lastly, LOOVF exceeds ten seconds already for $n = 400$ observations and is always found to be the slowest method.

Appendix B. Additional Figures and Tables

In this section we collect additional figures and tables.

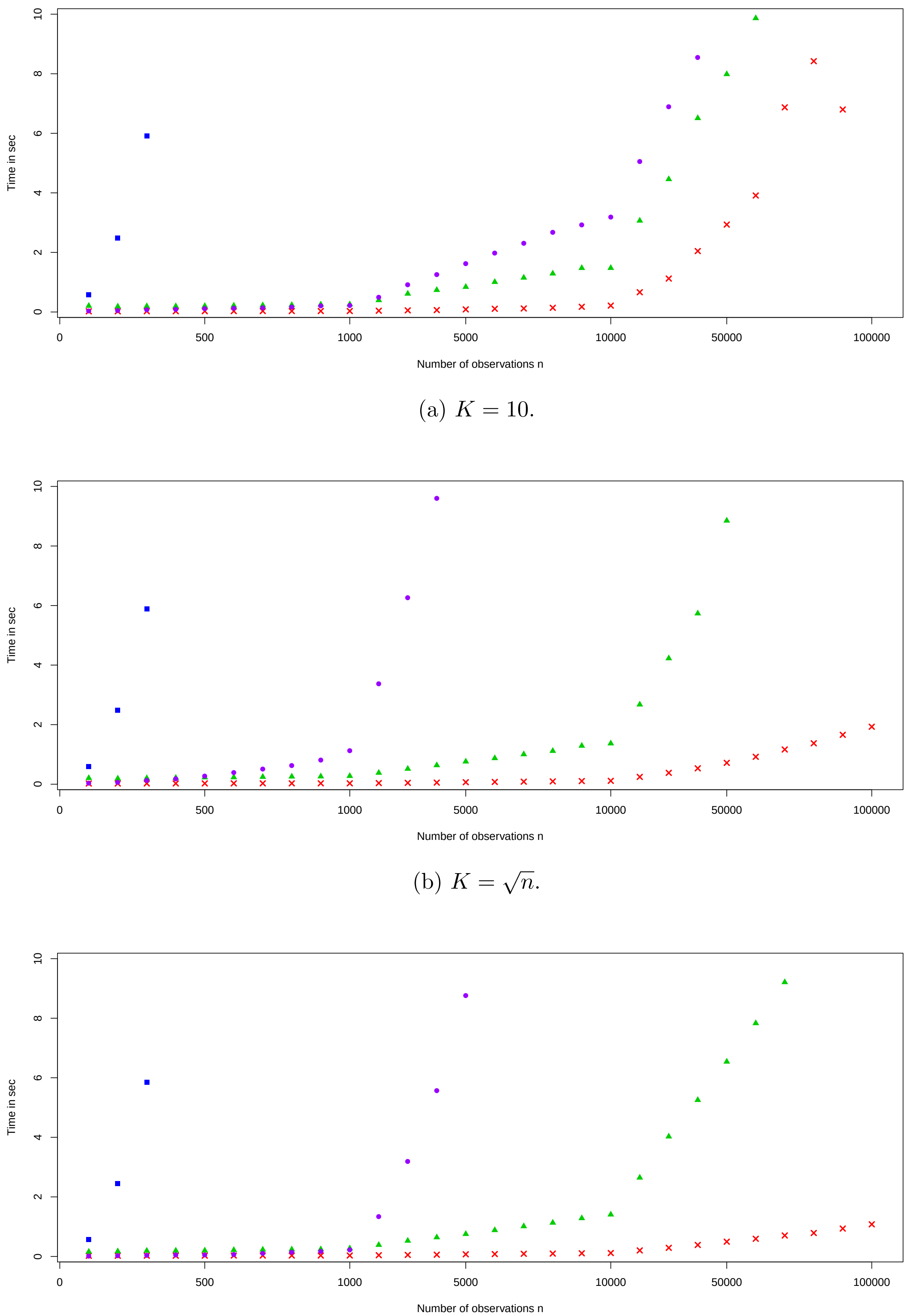
B.1. **Simulations.** We start with estimates by CBS, cumSeg and LOOVF for the data from Figure 2.

The following three tables collect the results of the simulations in Section 4.1. Recall the random pair $(\mu_R, \sigma_R^2) \in \mathcal{S}$ (all random variables are independent from each other):

(a) We fix the number of observations n, the number of change-points K, a constant C and a minimum value for the smallest scale $\lambda_{\min}$.

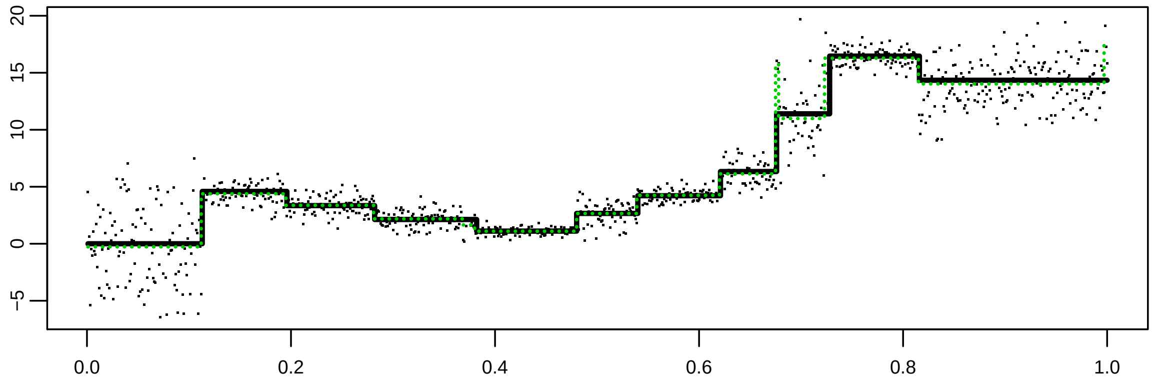
(b) We draw the locations of the change-points $\tau_0 := 0 < \tau_1 < \cdots < \tau_K < 1 =: \tau_{K+1}$ uniformly distributed with the restriction that $\lambda := \min_{k=0,\ldots,K} |\tau_{k+1} - \tau_k| \geq \lambda_{\min}$.

(c) We choose the function values $s_0, \ldots, s_K$ of the standard deviation function σ_R by $s_k := 2^{U_k}$, where $U_0, \ldots, U_K$ are uniform distributed on $[-2, 2]$.

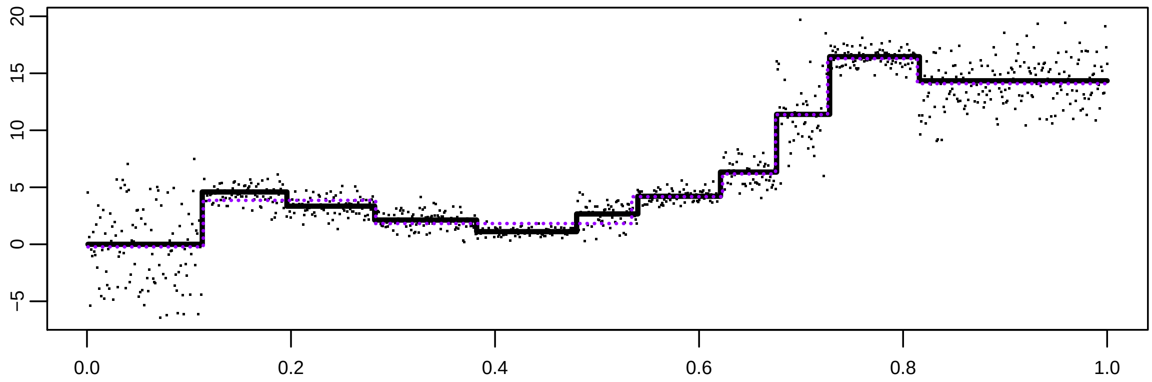


(c) $K = n/100$.

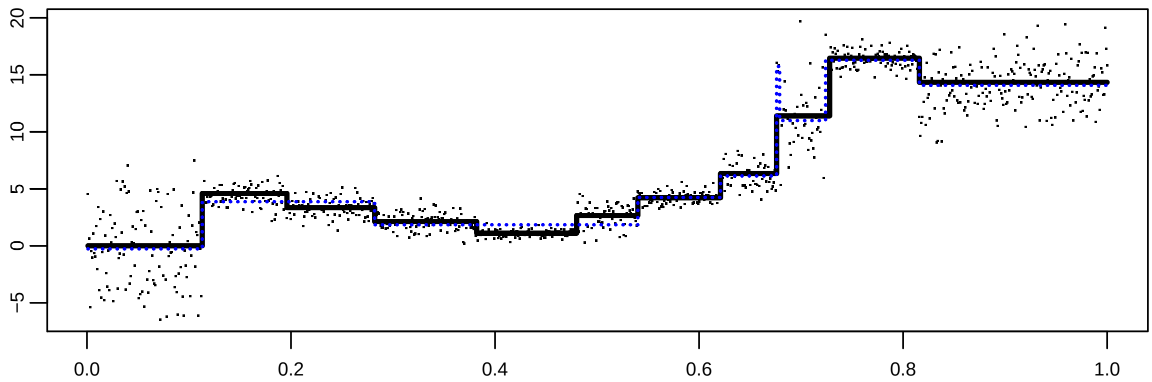
FIGURE 8. Mean computation time of H-SMUCE (red crosses), CBS (green triangles), cumSeg (purple circles) and LOOVF (blue squares) for different number of observations n and different number of change-points K. Note that for purposes of visualization the x-axis is displayed non-equidistantly.



(a) CBS.



(b) cumSeg.



(c) LOOVF.

FIGURE 9. Observations (black points) and true signal (black line) together with estimates by CBS, cumSeg and LOOVF for the data from Figure 2. All parameters are chosen as described in Section 4.

(d) We determine the function values $m_0, \ldots, m_K$ of the signal μ_R such that

$$|m_k - m_{k-1}| = \sqrt{\frac{C}{n} \min\left(\frac{\tau_{k+1} - \tau_k}{s_k^2}, \frac{\tau_k - \tau_{k-1}}{s_{k-1}^2}\right)^{-1}} \ \forall\ k = 1, \ldots, K.$$

Thereby, we start with $m_0 = 0$ and choose randomly with probability $1/2$ whether the expectation increases or decreases.

All simulations are repeated 10 000 times.

Setting	Method	-1	0	+1	$\geq +2$	$\lvert\hat{K}-K\rvert$	FPSLE	FNSLE	MISE	MIAE
$\sigma_0 = 0.5$,	HS(0.1)	0.000	0.995	0.004	0.000	0.005	0.82	0.74	0.0119	0.0644
$\sigma_1 = 0.5$,	HS(0.3)	0.000	0.975	0.025	0.000	0.025	1.38	0.95	0.0129	0.0672
	HS(0.5)	0.000	0.929	0.070	0.001	0.072	2.67	1.40	0.0144	0.0706
	CBS	0.000	0.949	0.036	0.015	0.066	2.31	0.94	0.0128	0.0660
	cumSeg	0.000	0.995	0.005	0.000	0.005	1.37	1.28	0.0172	0.0707
	LOOVF	0.000	0.774	0.142	0.084	0.378	10.10	2.42	0.1402	0.2897
$\sigma_0 = 0.5$,	HS(0.1)	0.112	0.886	0.002	0.000	0.114	3.99	6.77	0.0543	0.1405
$\sigma_1 = 1$,	HS(0.3)	0.020	0.961	0.019	0.000	0.039	2.38	2.56	0.0321	0.1086
	HS(0.5)	0.005	0.940	0.054	0.001	0.061	3.12	2.30	0.0314	0.1090
	CBS	0.042	0.873	0.068	0.017	0.147	5.87	4.93	0.0496	0.1315
	cumSeg	0.008	0.969	0.021	0.003	0.034	3.09	2.78	0.0375	0.1126
	LOOVF	0.006	0.791	0.112	0.091	0.373	11.30	3.90	0.1720	0.3004
$\sigma_0 = 0.5$,	HS(0.1)	0.484	0.515	0.001	0.000	0.485	12.77	24.89	0.1736	0.3110
$\sigma_1 = 1.5$,	HS(0.3)	0.209	0.778	0.012	0.000	0.222	6.81	11.92	0.1025	0.2075
	HS(0.5)	0.089	0.872	0.039	0.000	0.129	4.92	6.54	0.0725	0.1690
	CBS	0.417	0.454	0.105	0.024	0.577	17.63	25.40	0.1845	0.3385
	cumSeg	0.231	0.731	0.032	0.006	0.276	9.60	14.62	0.1149	0.2317
	LOOVF	0.135	0.683	0.098	0.085	0.490	15.78	11.92	0.2307	0.3322
$\sigma_0 = 1$,	HS(0.1)	0.453	0.547	0.001	0.000	0.453	13.49	24.97	0.1514	0.3140
$\sigma_1 = 1$,	HS(0.3)	0.171	0.818	0.011	0.000	0.182	8.11	12.53	0.0942	0.2170
	HS(0.5)	0.062	0.900	0.038	0.000	0.101	6.75	7.99	0.0745	0.1847
	CBS	0.156	0.744	0.091	0.008	0.265	9.51	11.41	0.0943	0.2127
	cumSeg	0.120	0.876	0.004	0.000	0.124	5.93	8.88	0.0748	0.1839
	LOOVF	0.039	0.749	0.132	0.081	0.405	13.29	6.87	0.1947	0.3472
$\sigma_0 = 1$,	HS(0.1)	0.727	0.272	0.000	0.000	0.728	19.44	37.71	0.2237	0.4244
$\sigma_1 = 1.5$,	HS(0.3)	0.410	0.584	0.006	0.000	0.416	13.35	23.77	0.1644	0.3256
	HS(0.5)	0.218	0.753	0.028	0.000	0.247	10.25	15.64	0.1283	0.2669
	CBS	0.491	0.406	0.096	0.008	0.604	18.00	28.42	0.2013	0.3741
	cumSeg	0.409	0.580	0.010	0.000	0.420	12.91	22.99	0.1571	0.3155
	LOOVF	0.184	0.638	0.105	0.072	0.501	16.55	14.92	0.2410	0.3626
$\sigma_0 = 1.5$,	HS(0.1)	0.844	0.156	0.000	0.000	0.844	22.41	43.65	0.2581	0.4713
$\sigma_1 = 1.5$,	HS(0.3)	0.574	0.423	0.003	0.000	0.577	18.21	33.12	0.2219	0.4101
	HS(0.5)	0.352	0.629	0.018	0.000	0.371	15.47	25.01	0.1915	0.3582
	CBS	0.659	0.258	0.079	0.003	0.746	20.73	35.81	0.2449	0.4379
	cumSeg	0.629	0.369	0.002	0.000	0.631	17.56	33.32	0.2147	0.4067
	LOOVF	0.297	0.534	0.104	0.066	0.589	19.34	21.26	0.2715	0.4046

TABLE 1. Simulations with a single change (fixed signal and variances): $n = 100$ observations and a single change at 0.5, from 0 to 1 for different standard deviations changing from σ_0 to σ_1 at 0.5, too. Columns from left to right: setting, method, proportions of $\hat{K} - K$ and averages of the corresponding error criteria. HS(α) denotes H-SMUCE at significance level α.

B.2. **Prior information on scales.** To demonstrate the effect of incorporating prior knowledge about those scales where change-points are likely to happen we consider again the observations from Table 3 with $n = 10\,000$, $K = 10$ and $\lambda_{\min} = 50$. To this end, we use the adapted weights, where we eliminate the smallest three scales $k = 1, 2, 3$, since all constant segments contain at least 50 observations and therefore these small scales are not needed for detection. Moreover, we choose $\tilde{\beta}_4 = 1/4$, $\tilde{\beta}_5 = 1/4$, $\tilde{\beta}_6 = 1/6$, $\tilde{\beta}_7 = 1/6$, $\tilde{\beta}_8 = 1/12$, $\tilde{\beta}_9 = 1/12$ in decreasing order, since change-points on smaller scales are more likely and harder to detect. For the same reasons we eliminate the four largest scales $k = 10, 11, 12, 13$, too.

A comparison of Table 3 and 4 shows that the modified weights increase the detection power of H-SMUCE for all significance levels, so we encourage the user to adapt the

Setting	Method	≤ -2	-1	0	+1	$\ge +2$	$\vert\hat{K}-K\vert$	FPSLE	FNSLE	MISE	MIAE
n = 1000,	HS(0.1)	-	-	0.965	0.035	0.000	0.035	17.75	4.73	0.0035	0.0365
K = 0,	HS(0.3)	-	-	0.867	0.128	0.005	0.138	68.95	18.42	0.0045	0.0401
$\mu = \mu_R \equiv 0$,	HS(0.5)	-	-	0.719	0.256	0.025	0.307	153.45	41.25	0.0061	0.0454
$\sigma = \sigma_R$	S(0.1)	-	-	0.965	0.034	0.001	0.036	17.90	5.03	0.0039	0.0371
$\equiv$ const	S(0.3)	-	-	0.832	0.160	0.008	0.177	88.45	24.80	0.0059	0.0435
	S(0.5)	-	-	0.667	0.298	0.035	0.370	184.90	50.94	0.0082	0.0499
	CBS	-	-	0.991	0.000	0.009	0.018	8.90	1.26	0.0037	0.0351
	cumSeg	-	-	0.999	0.001	0.000	0.001	0.30	0.06	0.0029	0.0345
n = 1000,	HS(0.1)	0.010	0.174	0.802	0.014	0.000	0.208	26.32	72.66	0.0132	0.0613
K = 2,	HS(0.3)	0.004	0.108	0.819	0.067	0.002	0.187	38.10	52.90	0.0114	0.0571
$\lambda_{\min} = 30$,	HS(0.5)	0.002	0.070	0.768	0.150	0.010	0.244	64.14	48.50	0.0111	0.0573
$\mu = \mu_R$,	S(0.1)	0.003	0.074	0.912	0.011	0.000	0.092	16.96	34.03	0.0092	0.0513
$\sigma \equiv 1$	S(0.3)	0.001	0.040	0.892	0.065	0.002	0.112	32.24	27.39	0.0090	0.0513
	S(0.5)	0.001	0.025	0.806	0.155	0.013	0.209	63.30	32.33	0.0095	0.0536
	CBS	0.005	0.060	0.821	0.082	0.033	0.221	37.55	37.57	0.0111	0.0527
	cumSeg	0.025	0.116	0.749	0.099	0.011	0.289	65.32	82.63	0.0364	0.0738
n = 1000,	HS(0.1)	0.009	0.160	0.815	0.015	0.000	0.194	27.14	68.91	0.0127	0.0611
K = 2,	HS(0.3)	0.004	0.098	0.829	0.067	0.001	0.176	37.77	49.63	0.0111	0.0572
$\lambda_{\min} = 50$,	HS(0.5)	0.002	0.063	0.774	0.152	0.009	0.237	63.46	46.06	0.0109	0.0573
$\mu = \mu_R$,	S(0.1)	0.003	0.068	0.919	0.009	0.000	0.084	16.82	31.94	0.0091	0.0515
$\sigma \equiv 1$	S(0.3)	0.001	0.035	0.899	0.063	0.002	0.104	31.19	25.81	0.0090	0.0515
	S(0.5)	0.001	0.020	0.819	0.147	0.013	0.195	59.86	30.23	0.0095	0.0537
	CBS	0.005	0.058	0.824	0.083	0.031	0.215	37.50	36.27	0.0112	0.0532
	cumSeg	0.023	0.110	0.769	0.090	0.008	0.262	59.74	79.25	0.0336	0.0741
n = 1000,	HS(0.1)	0.508	0.330	0.161	0.001	0.000	1.634	54.37	172.66	0.1112	0.1842
K = 10,	HS(0.3)	0.354	0.377	0.263	0.006	0.000	1.233	44.53	127.81	0.0817	0.1561
$\lambda_{\min} = 30$,	HS(0.5)	0.253	0.384	0.346	0.017	0.000	0.987	40.88	102.88	0.0679	0.1419
$\mu = \mu_R$,	S(0.1)	0.163	0.352	0.485	0.001	0.000	0.721	29.14	77.49	0.0424	0.1193
$\sigma \equiv 1$	S(0.3)	0.093	0.301	0.598	0.007	0.000	0.513	24.23	56.17	0.0366	0.1099
	S(0.5)	0.062	0.258	0.657	0.022	0.001	0.415	23.34	46.37	0.0342	0.1060
	CBS	0.033	0.129	0.531	0.204	0.102	0.644	42.69	45.08	0.0417	0.1078
	cumSeg	0.163	0.216	0.403	0.165	0.053	0.904	65.16	105.59	0.1107	0.1492
n = 1000,	HS(0.1)	0.445	0.356	0.198	0.001	0.000	1.474	59.32	162.03	0.0913	0.1801
K = 10,	HS(0.3)	0.303	0.384	0.307	0.005	0.000	1.104	47.34	120.10	0.0682	0.1532
$\lambda_{\min} = 50$,	HS(0.5)	0.213	0.379	0.390	0.018	0.001	0.881	41.98	96.70	0.0577	0.1398
$\mu = \mu_R$,	S(0.1)	0.155	0.351	0.494	0.000	0.000	0.697	32.51	77.29	0.0426	0.1235
$\sigma \equiv 1$	S(0.3)	0.085	0.299	0.612	0.004	0.000	0.485	26.14	55.78	0.0368	0.1131
	S(0.5)	0.054	0.252	0.680	0.014	0.000	0.381	23.81	45.39	0.0344	0.1086
	CBS	0.027	0.135	0.524	0.203	0.111	0.653	45.64	44.88	0.0425	0.1116
	cumSeg	0.165	0.217	0.389	0.179	0.050	0.904	63.73	104.37	0.1037	0.1522

TABLE 2. Simulations with constant variance and $C = 200$. Columns from left to right: setting, method, proportions of $\hat{K} - K$ and averages of the corresponding error criteria. HS(α) and S(α) denote H-SMUCE and SMUCE at significance level α, respectively.

weights if prior information on the scales where changes occur is available.

B.3. **Robustness.** Figure 10 shows the standard deviation functions in Table 5 to examine robustness against variance changes on constant segments. We consider the sinus-shaped standard deviation σ_1 (continuous changes), the piecewise linear standard deviation σ_2 (continuous and abrupt changes at the same time) and the piecewise constant standard deviation σ_3 (abrupt changes). Moreover, we analyse in Table 6 robustness against small periodic trends in simulations similar to those in (Venkatraman et al., 2004). More precisely, we generate the random pairs $(\mu_R, \sigma_R^2) \in \mathcal{S}$ as in (a)-(d) described, but

Setting	Method	≤ -2	-1	0	+1	$\geq +2$	$\lvert\hat{K}-K\rvert$	FPSLE	FNSLE	MISE	MIAE
n = 100,	HS(0.1)	0.000	0.125	0.873	0.002	0.000	0.128	1.51	4.07	0.8182	0.3308
K = 2,	HS(0.3)	0.000	0.042	0.945	0.013	0.000	0.055	1.04	1.70	0.4217	0.2482
$\lambda_{\min} = 15$,	HS(0.5)	0.000	0.016	0.940	0.043	0.000	0.060	1.63	1.26	0.2776	0.2291
$\mu = \mu_R$,	CBS	0.000	0.001	0.925	0.058	0.016	0.092	2.03	0.79	0.2220	0.2143
$\sigma = \sigma_R$	cumSeg	0.000	0.066	0.720	0.167	0.047	0.343	6.50	4.39	0.4898	0.3053
	LOOVF	0.000	0.031	0.700	0.163	0.106	0.683	12.83	3.36	0.3167	0.2639
n = 100,	HS(0.1)	0.608	0.364	0.028	0.000	0.000	1.610	13.51	32.33	9.5104	1.8626
K = 5,	HS(0.3)	0.212	0.577	0.211	0.000	0.000	1.003	8.63	19.80	6.5362	1.3263
$\lambda_{\min} = 15$,	HS(0.5)	0.061	0.466	0.473	0.001	0.000	0.588	5.27	11.65	3.9992	0.9047
$\mu = \mu_R$,	CBS	0.001	0.008	0.884	0.089	0.018	0.137	1.65	1.02	0.4539	0.3130
$\sigma = \sigma_R$	cumSeg	0.098	0.230	0.544	0.117	0.012	0.588	6.93	12.13	1.2454	0.5441
	LOOVF	0.031	0.112	0.520	0.152	0.184	1.648	14.61	6.92	0.5887	0.4042
n = 1000,	HS(0.1)	0.000	0.007	0.974	0.018	0.000	0.026	8.42	5.83	0.0195	0.0617
K = 2,	HS(0.3)	0.000	0.001	0.921	0.075	0.002	0.080	24.72	9.57	0.0193	0.0636
$\lambda_{\min} = 30$,	HS(0.5)	0.000	0.000	0.827	0.162	0.012	0.185	53.23	17.09	0.0204	0.0668
$\mu = \mu_R$,	CBS	0.005	0.019	0.774	0.146	0.056	0.298	52.95	21.17	0.0347	0.0711
$\sigma = \sigma_R$	cumSeg	0.022	0.161	0.683	0.103	0.030	0.387	64.04	92.66	0.0765	0.1112
n = 1000,	HS(0.1)	0.000	0.002	0.982	0.017	0.000	0.018	7.25	4.35	0.0182	0.0630
K = 2,	HS(0.3)	0.000	0.000	0.926	0.071	0.002	0.076	22.64	8.49	0.0196	0.0657
$\lambda_{\min} = 50$,	HS(0.5)	0.000	0.000	0.830	0.160	0.010	0.181	50.02	16.22	0.0214	0.0692
$\mu = \mu_R$,	CBS	0.003	0.011	0.776	0.153	0.057	0.296	53.69	15.85	0.0355	0.0730
$\sigma = \sigma_R$	cumSeg	0.016	0.155	0.699	0.098	0.031	0.370	60.63	84.69	0.0739	0.1132
n = 1000,	HS(0.1)	0.123	0.429	0.446	0.002	0.000	0.686	22.83	55.06	0.4045	0.2402
K = 10,	HS(0.3)	0.016	0.199	0.770	0.015	0.000	0.245	11.98	21.12	0.1863	0.1618
$\lambda_{\min} = 30$,	HS(0.5)	0.002	0.088	0.863	0.045	0.001	0.140	11.84	12.71	0.1220	0.1404
$\mu = \mu_R$,	CBS	0.002	0.008	0.463	0.316	0.211	0.843	47.26	15.20	0.1274	0.1435
$\sigma = \sigma_R$	cumSeg	0.439	0.243	0.187	0.085	0.046	1.674	94.91	228.44	0.3120	0.2806
n = 1000,	HS(0.1)	0.025	0.262	0.711	0.002	0.000	0.315	16.94	32.39	0.2102	0.1866
K = 10,	HS(0.3)	0.002	0.058	0.925	0.015	0.000	0.076	8.46	10.58	0.1009	0.1372
$\lambda_{\min} = 50$,	HS(0.5)	0.000	0.017	0.940	0.043	0.001	0.061	9.03	7.72	0.0860	0.1307
$\mu = \mu_R$,	CBS	0.001	0.007	0.451	0.319	0.222	0.868	47.81	15.10	0.1293	0.1463
$\sigma = \sigma_R$	cumSeg	0.433	0.254	0.197	0.082	0.035	1.601	97.00	223.47	0.2771	0.2794
n = 10000,	HS(0.1)	0.000	0.004	0.983	0.013	0.000	0.017	50.65	30.94	0.0016	0.0183
K = 2,	HS(0.3)	0.000	0.002	0.936	0.061	0.001	0.065	188.73	63.72	0.0016	0.0188
$\lambda_{\min} = 30$,	HS(0.5)	0.000	0.001	0.865	0.128	0.006	0.142	407.41	125.46	0.0016	0.0197
$\mu = \mu_R$,	CBS	0.012	0.036	0.532	0.200	0.220	0.886	1548.96	373.22	0.0057	0.0235
$\sigma = \sigma_R$	cumSeg	0.054	0.245	0.600	0.084	0.017	0.477	682.64	1457.08	0.0090	0.0379
n = 10000,	HS(0.1)	0.000	0.001	0.984	0.015	0.000	0.016	53.23	24.89	0.0014	0.0182
K = 2,	HS(0.3)	0.000	0.000	0.941	0.057	0.002	0.060	181.06	59.83	0.0014	0.0188
$\lambda_{\min} = 50$,	HS(0.5)	0.000	0.000	0.870	0.124	0.007	0.137	394.16	115.62	0.0016	0.0197
$\mu = \mu_R$,	CBS	0.012	0.035	0.521	0.208	0.225	0.917	1601.54	366.42	0.0058	0.0238
$\sigma = \sigma_R$	cumSeg	0.052	0.241	0.603	0.087	0.016	0.473	673.81	1430.47	0.0084	0.0377
n = 10000,	HS(0.1)	0.023	0.231	0.741	0.005	0.000	0.282	58.42	165.72	0.0178	0.0431
K = 10,	HS(0.3)	0.006	0.123	0.844	0.027	0.000	0.162	68.27	98.25	0.0122	0.0385
$\lambda_{\min} = 30$,	HS(0.5)	0.003	0.079	0.854	0.064	0.002	0.151	108.19	87.63	0.0103	0.0377
$\mu = \mu_R$,	CBS	0.024	0.043	0.180	0.222	0.531	2.088	1286.59	525.95	0.0198	0.0475
$\sigma = \sigma_R$	cumSeg	0.619	0.169	0.130	0.059	0.024	2.345	1000.55	3122.28	0.0433	0.0917
n = 10000,	HS(0.1)	0.009	0.165	0.819	0.007	0.000	0.190	59.11	124.05	0.0132	0.0418
K = 10,	HS(0.3)	0.001	0.064	0.905	0.029	0.001	0.097	67.32	65.54	0.0089	0.0375
$\lambda_{\min} = 50$,	HS(0.5)	0.000	0.029	0.900	0.067	0.003	0.102	103.42	60.04	0.0078	0.0368
$\mu = \mu_R$,	CBS	0.019	0.034	0.162	0.228	0.557	2.203	1317.31	467.47	0.0198	0.0475
$\sigma = \sigma_R$	cumSeg	0.607	0.188	0.131	0.051	0.023	2.277	997.64	3105.88	0.0405	0.0925
n = 10000,	HS(0.1)	0.609	0.284	0.107	0.001	0.000	1.908	155.65	504.02	0.1016	0.1031
K = 25,	HS(0.3)	0.278	0.399	0.318	0.006	0.000	1.044	94.53	263.30	0.0640	0.0789
$\lambda_{\min} = 30$,	HS(0.5)	0.140	0.371	0.470	0.019	0.000	0.696	84.07	182.54	0.0483	0.0703
$\mu = \mu_R$,	CBS	0.015	0.024	0.069	0.128	0.765	3.348	921.91	409.98	0.0411	0.0723
$\sigma = \sigma_R$	cumSeg	0.934	0.036	0.018	0.009	0.003	6.028	1043.82	3488.43	0.1159	0.1540
n = 10000,	HS(0.1)	0.396	0.383	0.220	0.001	0.000	1.334	146.74	387.66	0.0699	0.0945
K = 25,	HS(0.3)	0.103	0.359	0.528	0.010	0.000	0.591	85.33	175.03	0.0390	0.0715
$\lambda_{\min} = 50$,	HS(0.5)	0.038	0.241	0.690	0.030	0.001	0.352	78.74	114.01	0.0291	0.0647
$\mu = \mu_R$,	CBS	0.010	0.017	0.055	0.120	0.799	3.529	934.29	346.33	0.0405	0.0726
$\sigma = \sigma_R$	cumSeg	0.934	0.036	0.019	0.008	0.003	5.849	1053.35	3462.62	0.1022	0.1547

TABLE 3. Simulations with heterogeneous errors and $C = 200$. Columns from left to right: setting, method, proportions of $\hat{K} - K$ and averages of the corresponding error criteria. HS(α) denotes H-SMUCE at significance level α.

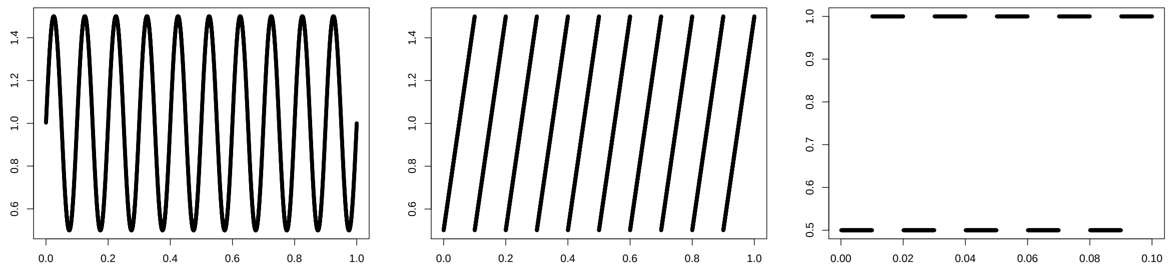
Method	≤ -2	-1	0	+1	$\geq +2$	$\lvert\hat{K}-K\rvert$	FPSLE	FNSLE	MISE	MIAE
HS(0.1)	0.005	0.117	0.876	0.002	0.000	0.130	50.82	113.50	0.0107	0.0406
HS(0.3)	0.000	0.032	0.952	0.016	0.000	0.049	48.39	49.84	0.0075	0.0368
HS(0.5)	0.000	0.013	0.940	0.045	0.001	0.061	78.86	48.19	0.0072	0.0368

TABLE 4. $n = 10\,000$ observations, $K = 10$ change-points, $C = 200$ and $\lambda_{\min} = 50$ from Table 3. Columns from left to right: setting, method, proportions of $\hat{K} - K$ and averages of the corresponding error criteria. HS(α) denotes H-SMUCE at significance level α, but with weights $\tilde{\beta}_4, \ldots, \tilde{\beta}_9$.

replace the signal μ_R by

$$\mu_T(i/n) = \mu_R + b\sin(a\pi i)$$
$$\text{and } \mu_{T_\sigma}(i/n) = \mu_R + b\sigma_R(i/n)\sin(a\pi i) + b(\sigma_R(i/n) - \sigma_R((i-1)/n)\sin(a\pi i),$$
$$i = 1, \ldots, n,$$

respectively. The signal μ_T reflects the situation of a fixed periodic trend, whereas in μ_{T_σ} the trend is scaled by the local standard deviation. The last term corrects the size of changes such that still σ_R determines the changes. We consider as in (Venkatraman et al., 2004) long ($a = 0.01$) and short ($a = 0.025$) trends. Finally, Table 7 reports result of t_3 distributed errors.



(a) (b) (c)

FIGURE 10. a: Continuous sinus-shaped standard deviation $\sigma_1(t) := 1 + 0.5\sin(20\pi t)$. b: Piecewise linear standard deviation $\sigma_2(t) := 0.5 + \sum_{i=0}^{9}(10t - i)\mathbb{1}_{(0.1i, 0.1(i+1)]}(t)$. c: Piecewise constant standard deviation $\sigma_3(t) := \sum_{i=1}^{n/200} 0.5\mathbb{1}_{(200(i-1)/n,\ 200(i-1)/n+100/n]}(t) + \mathbb{1}_{(200(i-1)/n+100/n,\ 200i/n]}(t)$, exemplary for $n = 1\,000$.

APPENDIX C. PROOFS

In this section we collect the proofs together with some auxiliary statements.

C.1. **Proof of Lemma 2.1.**

Proof of Lemma 2.1. A single statistic $T_i^j(Z, 0)$ has the c.d.f. $F_{1,j-i}(\cdot)$ of an F-distribution with $(1, j-i)$ degrees of freedom. Thus, $F_k(\cdot) = F_{1,2^k-1}(\cdot)^{|\mathcal{D}_k|}$ is continuous and strictly monotonically increasing for positive arguments. Now, it follows from equation (2.5) that

$$q_k = F_k^{-1}\left(1 - \frac{\beta_k}{\beta_1}(1 - F_1(q_1))\right) \text{ for } k = 2, \ldots, d_n. \tag{C.1}$$

Setting	Method	≤ -2	-1	0	+1	$\geq +2$	$\lvert\hat{K}-K\rvert$	FPSLE	FNSLE	MISE	MIAE
n = 1000,	HS(0.1)	-	-	0.968	0.032	0.000	0.033	16.30	4.12	0.0013	0.0277
K = 0,	HS(0.3)	-	-	0.876	0.118	0.005	0.129	64.60	15.91	0.0018	0.0306
$\mu = \mu_R \equiv 0$,	HS(0.5)	-	-	0.734	0.239	0.027	0.293	146.75	36.45	0.0023	0.0338
$\sigma = \sigma_1$	CBS	-	-	0.916	0.001	0.083	0.186	93.25	11.21	0.0045	0.0288
	cumSeg	-	-	1.000	0.000	0.000	0.000	0.20	0.04	0.0011	0.0264
n = 1000,	HS(0.1)	-	-	0.968	0.031	0.001	0.032	16.10	4.12	0.0013	0.0278
K = 0,	HS(0.3)	-	-	0.876	0.118	0.005	0.129	64.55	15.73	0.0017	0.0306
$\mu = \mu_R \equiv 0$,	HS(0.5)	-	-	0.734	0.241	0.024	0.292	145.80	35.28	0.0022	0.0340
$\sigma = \sigma_2$	CBS	-	-	0.937	0.004	0.060	0.135	67.70	8.96	0.0034	0.0281
	cumSeg	-	-	0.999	0.001	0.000	0.001	0.40	0.12	0.0011	0.0264
n = 1000,	HS(0.1)	-	-	0.969	0.030	0.001	0.032	15.75	3.91	0.0007	0.0210
K = 0,	HS(0.3)	-	-	0.875	0.119	0.006	0.130	65.10	16.31	0.0009	0.0227
$\mu = \mu_R \equiv 0$,	HS(0.5)	-	-	0.737	0.236	0.026	0.290	145.15	36.09	0.0012	0.0250
$\sigma = \sigma_3$	CBS	-	-	0.937	0.002	0.061	0.134	67.10	8.64	0.0019	0.0213
	cumSeg	-	-	0.999	0.001	0.000	0.001	0.35	0.10	0.0006	0.0199
n = 10000,	HS(0.1)	0.013	0.185	0.796	0.005	0.000	0.218	661.16	755.83	0.0212	0.0684
K = 10,	HS(0.3)	0.003	0.076	0.890	0.031	0.001	0.113	543.91	548.21	0.0167	0.0585
$\lambda_{\min} = 50$,	HS(0.5)	0.001	0.041	0.886	0.069	0.003	0.117	513.55	468.37	0.0147	0.0542
$\mu = \mu_R$,	CBS	0.000	0.001	0.191	0.155	0.653	2.636	1590.35	276.51	0.0092	0.0358
$\sigma = \sigma_1$	cumSeg	0.206	0.118	0.413	0.193	0.070	0.984	790.10	1054.73	0.0146	0.0502
n = 10000,	HS(0.1)	0.014	0.205	0.776	0.006	0.000	0.238	421.19	513.32	0.0156	0.0556
K = 10,	HS(0.3)	0.001	0.077	0.894	0.027	0.001	0.108	348.50	358.14	0.0119	0.0475
$\lambda_{\min} = 50$,	HS(0.5)	0.000	0.038	0.897	0.062	0.002	0.105	344.93	311.35	0.0106	0.0446
$\mu = \mu_R$,	CBS	0.000	0.000	0.215	0.174	0.611	2.362	1454.85	247.26	0.0085	0.0346
$\sigma = \sigma_2$	cumSeg	0.114	0.102	0.467	0.236	0.082	0.795	756.12	720.95	0.0136	0.0478
n = 10000,	HS(0.1)	0.019	0.233	0.744	0.004	0.000	0.276	161.27	251.06	0.0053	0.0301
K = 10,	HS(0.3)	0.002	0.069	0.904	0.025	0.000	0.099	137.86	136.79	0.0036	0.0254
$\lambda_{\min} = 50$,	HS(0.5)	0.000	0.029	0.906	0.062	0.003	0.096	170.29	128.56	0.0033	0.0248
$\mu = \mu_R$,	CBS	0.000	0.000	0.246	0.173	0.582	2.189	1134.85	214.71	0.0047	0.0263
$\sigma = \sigma_3$	cumSeg	0.054	0.051	0.516	0.279	0.101	0.669	749.33	499.10	0.0070	0.0346

TABLE 5. Simulations with standard deviations $\sigma_1(\cdot)$-$\sigma_3(\cdot)$ from Figure 10 and $C = 200$. Columns from left to right: setting, method, proportions of $\hat{K} - K$ and averages of the corresponding error criteria. HS(α) denotes H-SMUCE at significance level α.

This together with equation (2.3) yields

$$G(q_1) := F\left(q_1, F_2^{-1}\left(1 - \frac{\beta_2}{\beta_1}(1 - F_1(q_1))\right), \ldots, F_{d_n}^{-1}\left(1 - \frac{\beta_{d_n}}{\beta_1}(1 - F_1(q_1))\right)\right) = 1 - \alpha.$$

Note, that F is continuous and $\lim_{q_k \to 0} F(q_1, \ldots, q_{d_n}) = 0$ for all $k = 1, \ldots, d_n$ as well as $\lim_{q_1,\ldots,q_{d_n} \to \infty} F(q_1, \ldots, q_{d_n}) = 1$. Thus, the function G is continuous, strictly monotonically increasing on $[0, \infty)$ and attains all values in $[0, 1)$. Therefore, the existence of the vector of critical values follows from the intermediate value theorem and the vector is also unique. □

C.2. **Proof of Lemma 3.1.** First of all, recall from the proof of Lemma 2.1 that the statistic T_k has c.d.f. $F_{1,2^k-1}(\cdot)^{|\mathcal{D}_k|}$. For every $k = 1, \ldots, d_n$ we use the transformation

$$U_k := F_{1,2^k-1}(T_k)^{|\mathcal{D}_k|}$$

and the identity

$$T_k = F_{1,2^k-1}^{-1}\left(U_k^{|\mathcal{D}_k|^{-1}}\right).$$

Setting	Method	≤ -2	-1	0	+1	$\geq +2$	$\lvert\hat{K}-K\rvert$	FPSLE	FNSLE	MISE	MIAE
$\mu = \mu_T$,	HS(0.1)	0.137	0.421	0.439	0.003	0.000	0.709	24.74	58.06	0.3976	0.2449
$a = 0.01$,	HS(0.3)	0.019	0.209	0.741	0.032	0.000	0.279	15.65	24.13	0.1814	0.1681
$b = 0.1$,	HS(0.5)	0.003	0.091	0.822	0.081	0.003	0.184	17.32	15.63	0.1206	0.1474
$\sigma = \sigma_R$	CBS	0.001	0.011	0.383	0.290	0.315	1.141	72.93	22.14	0.1321	0.1535
	cumSeg	0.443	0.237	0.192	0.085	0.043	1.663	95.08	225.57	0.3080	0.2823
$\mu = \mu_T$,	HS(0.1)	0.149	0.360	0.370	0.104	0.016	0.821	73.52	94.89	0.4410	0.3070
$a = 0.01$,	HS(0.3)	0.029	0.181	0.496	0.226	0.067	0.611	78.00	62.38	0.2304	0.2376
$b = 0.3$,	HS(0.5)	0.007	0.092	0.466	0.306	0.129	0.697	88.73	53.42	0.1595	0.2169
$\sigma = \sigma_R$	CBS	0.001	0.006	0.082	0.135	0.776	3.243	249.67	74.97	0.1646	0.2325
	cumSeg	0.439	0.233	0.200	0.086	0.043	1.652	107.57	237.10	0.3394	0.3222
$\mu = \mu_T$,	HS(0.1)	0.140	0.287	0.323	0.176	0.075	0.936	134.92	135.90	0.5279	0.4052
$a = 0.01$,	HS(0.3)	0.032	0.146	0.327	0.298	0.197	0.970	146.40	103.03	0.3106	0.3393
$b = 0.5$,	HS(0.5)	0.009	0.076	0.258	0.329	0.328	1.223	162.18	92.24	0.2410	0.3207
$\sigma = \sigma_R$	CBS	0.002	0.004	0.020	0.043	0.932	5.623	435.32	124.66	0.2462	0.3440
	cumSeg	0.420	0.244	0.190	0.093	0.054	1.641	127.88	248.73	0.3921	0.3823
$\mu = \mu_T$,	HS(0.1)	0.128	0.424	0.446	0.002	0.000	0.693	23.50	56.36	0.4066	0.2415
$a = 0.025$,	HS(0.3)	0.017	0.201	0.759	0.023	0.001	0.259	13.43	22.21	0.1854	0.1628
$b = 0.1$,	HS(0.5)	0.003	0.086	0.843	0.066	0.002	0.162	14.51	13.77	0.1218	0.1416
$\sigma = \sigma_R$	CBS	0.002	0.008	0.395	0.287	0.308	1.135	64.07	19.55	0.1304	0.1471
	cumSeg	0.440	0.240	0.188	0.086	0.046	1.672	95.36	229.25	0.3058	0.2796
$\mu = \mu_T$,	HS(0.1)	0.108	0.344	0.411	0.111	0.027	0.738	58.57	73.26	0.4223	0.2606
$a = 0.025$,	HS(0.3)	0.016	0.138	0.468	0.252	0.126	0.715	77.74	50.19	0.2143	0.1929
$b = 0.3$,	HS(0.5)	0.003	0.058	0.382	0.315	0.243	0.989	101.66	48.03	0.1503	0.1749
$\sigma = \sigma_R$	CBS	0.002	0.003	0.050	0.065	0.880	5.370	356.54	85.59	0.1585	0.2027
	cumSeg	0.438	0.241	0.184	0.091	0.046	1.678	101.67	234.23	0.3243	0.2958
$\mu = \mu_T$,	HS(0.1)	0.054	0.180	0.276	0.226	0.264	1.247	164.83	114.26	0.4732	0.3127
$a = 0.025$,	HS(0.3)	0.007	0.060	0.195	0.229	0.509	1.945	214.89	100.21	0.2748	0.2586
$b = 0.5$,	HS(0.5)	0.001	0.027	0.127	0.191	0.654	2.591	256.78	101.30	0.2115	0.2465
$\sigma = \sigma_R$	CBS	0.000	0.001	0.006	0.011	0.982	10.383	709.91	149.02	0.2261	0.2993
	cumSeg	0.439	0.238	0.184	0.088	0.050	1.698	113.20	245.71	0.3520	0.3206
$\mu = \mu_{T_\sigma}$,	HS(0.1)	0.151	0.435	0.411	0.002	0.000	0.755	26.56	64.22	0.4469	0.3000
$a = 0.01$,	HS(0.3)	0.021	0.230	0.725	0.023	0.000	0.297	15.80	27.08	0.2289	0.2245
$b = 0.2$,	HS(0.5)	0.004	0.103	0.819	0.071	0.003	0.188	17.53	17.34	0.1622	0.2019
$\sigma = \sigma_R$	CBS	0.004	0.012	0.342	0.305	0.338	1.225	80.55	26.89	0.1689	0.2056
	cumSeg	0.422	0.233	0.193	0.095	0.056	1.653	107.36	234.27	0.3431	0.3292
$\mu = \mu_{T_\sigma}$,	HS(0.1)	0.254	0.410	0.298	0.036	0.001	1.012	68.27	115.70	0.6346	0.4582
$a = 0.01$,	HS(0.3)	0.055	0.261	0.488	0.176	0.019	0.591	81.26	80.78	0.4483	0.3953
$b = 0.5$,	HS(0.5)	0.015	0.129	0.466	0.321	0.070	0.624	100.70	72.34	0.3826	0.3694
$\sigma = \sigma_R$	CBS	0.002	0.004	0.022	0.059	0.914	4.231	357.42	117.42	0.2907	0.3140
	cumSeg	0.332	0.211	0.198	0.139	0.121	1.575	179.79	280.65	0.4522	0.4317
$\mu = \mu_{T_\sigma}$,	HS(0.1)	0.136	0.440	0.422	0.002	0.000	0.726	24.82	59.14	0.4567	0.3165
$a = 0.025$,	HS(0.3)	0.017	0.219	0.746	0.018	0.000	0.273	14.12	24.13	0.2432	0.2435
$b = 0.2$,	HS(0.5)	0.003	0.096	0.843	0.056	0.002	0.162	14.44	14.77	0.1756	0.2222
$\sigma = \sigma_R$	CBS	0.003	0.011	0.353	0.295	0.338	1.231	71.51	23.32	0.1892	0.2287
	cumSeg	0.432	0.238	0.182	0.094	0.054	1.688	103.19	236.34	0.3604	0.3501
$\mu = \mu_{T_\sigma}$,	HS(0.1)	0.181	0.433	0.370	0.016	0.000	0.831	37.89	75.31	0.7365	0.5244
$a = 0.025$,	HS(0.3)	0.033	0.240	0.594	0.125	0.009	0.450	42.28	44.37	0.5518	0.4736
$b = 0.5$,	HS(0.5)	0.007	0.110	0.541	0.281	0.061	0.534	64.52	41.39	0.4981	0.4582
$\sigma = \sigma_R$	CBS	0.002	0.002	0.023	0.043	0.929	5.589	365.42	103.32	0.4594	0.4362
	cumSeg	0.316	0.184	0.179	0.139	0.182	1.735	158.60	254.38	0.6153	0.5317

TABLE 6. Simulations with small periodic trends in the mean and $n = 1\,000$, $K = 10$, $\lambda_{\min} = 30$ and $C = 200$. Columns from left to right: setting, method, proportions of $\hat{K} - K$ and averages of the corresponding error criteria. HS(α) denotes H-SMUCE at significance level α.

Here, $F^{-1}_{1,2^k-1}(\cdot)$ denotes the quantile function of an F-distribution with $(1, 2^k - 1)$ degrees of freedom. Analogously, we define

$$q_{k,U} := F_{1,2^k-1}(q_k)^{|\mathcal{D}_k|}$$

Setting	Method	≤ -2	-1	0	+1	$\geq +2$	$\|\hat{K}-K\|$	FPSLE	FNSLE	MISE	MIAE
n = 1000,	HS(0.1)	-	-	0.982	0.018	0.000	0.018	9.05	2.53	0.0031	0.0347
K = 0,	HS(0.3)	-	-	0.927	0.071	0.001	0.074	37.05	10.31	0.0034	0.0361
$\mu = \mu_R \equiv 0$,	HS(0.5)	-	-	0.824	0.167	0.009	0.185	92.40	26.35	0.0043	0.0392
$\sigma = \sigma_R$,	S(0.1)	-	-	0.001	0.001	0.999	11.859	5929.70	369.55	0.8710	0.1491
	S(0.3)	-	-	0.000	0.000	1.000	14.803	7401.65	397.77	0.9338	0.1674
	S(0.5)	-	-	0.000	0.000	1.000	16.862	8431.00	411.30	0.9730	0.1787
	CBS	-	-	0.991	0.000	0.009	0.018	9.05	1.13	0.0058	0.0340
	cumSeg	-	-	0.955	0.001	0.044	0.188	93.90	11.98	0.0682	0.0375
n = 1000,	HS(0.1)	0.008	0.136	0.848	0.007	0.000	0.160	25.70	62.95	0.0120	0.0578
K = 2,	HS(0.3)	0.003	0.086	0.876	0.035	0.000	0.127	29.74	44.61	0.0103	0.0537
$\lambda_{\min} = 30$,	HS(0.5)	0.001	0.055	0.851	0.090	0.003	0.152	44.21	38.62	0.0097	0.0524
$\mu = \mu_R$,	S(0.1)	0.000	0.000	0.001	0.001	0.998	11.104	2683.40	250.21	0.3046	0.1232
$\sigma \equiv 1$,	S(0.3)	0.000	0.000	0.000	0.000	1.000	13.984	3361.80	283.17	0.3264	0.1340
	S(0.5)	0.000	0.000	0.000	0.000	1.000	15.991	3836.28	302.43	0.3400	0.1419
	CBS	0.053	0.161	0.726	0.043	0.018	0.346	46.69	119.74	0.0241	0.0712
	cumSeg	0.025	0.097	0.722	0.093	0.063	0.456	108.11	81.86	0.0557	0.0707
n = 10000,	HS(0.1)	0.002	0.079	0.916	0.004	0.000	0.086	93.09	119.69	0.0130	0.0425
K = 10,	HS(0.3)	0.000	0.025	0.957	0.017	0.000	0.043	86.32	81.78	0.0105	0.0397
$\lambda_{\min} = 50$,	HS(0.5)	0.000	0.012	0.950	0.038	0.000	0.050	99.93	76.24	0.0097	0.0389
$\mu = \mu_R$,	CBS	0.467	0.148	0.167	0.107	0.111	2.516	1356.25	6254.20	0.0877	0.1308
$\sigma = \sigma_R$	cumSeg	0.586	0.192	0.136	0.055	0.032	2.242	997.13	3005.71	0.0433	0.0906

TABLE 7. Simulations with t_3 distributed errors and $C = 200$. Columns from left to right: setting, method, proportions of $\hat{K} - K$ and averages of the corresponding error criteria. HS(α) and S(α) denote H-SMUCE and SMUCE at significance level α, respectively.

and have the identity

$$q_k = F^{-1}_{1,2^k-1}\left(q_{k,U}^{|\mathcal{D}_k|^{-1}}\right). \tag{C.2}$$

Then, the events $U_k > q_{k,\mathrm{U}}$ and $T_k > q_k$ are equivalent and therefore the vector $\mathbf{q}_U = (q_{1,\mathrm{U}}, \ldots, q_{d_n,\mathrm{U}})$ satisfies similar conditions to the equations (2.3) and (2.5), i.e.

$$1 - \mathbb{P}\left(U_1 \leq q_{1,U}, \ldots, U_{d_n} \leq q_{d_n,U}\right) = \alpha \tag{C.3}$$

and

$$\frac{1 - \mathbb{P}\left(U_1 \leq q_{1,U}\right)}{\beta_1} = \cdots = \frac{1 - \mathbb{P}\left(U_{d_n} \leq q_{d_n,U}\right)}{\beta_{d_n}}. \tag{C.4}$$

The following bounds can be interpreted as a weighted version of the Bonferroni-inequality.

Lemma C.1. *$q_{k,U} \leq 1 - \alpha\beta_k$ for $k = 1, \ldots, d_n$.*

Proof. We have $\mathbb{P}\left(U_j \leq q_{j,U}\right) = q_{j,U}$ for $j = 1, \ldots, d_n$, since U_j is uniformly distributed. Moreover, it follows from condition (C.4) that $1 - q_{j,U} = (1 - q_{k,U})\beta_j/\beta_k$. Combining this with equation (C.3) and $\sum_{j=1}^{d_n} \beta_j = 1$ yields

$$\begin{aligned}\alpha =& 1 - \mathbb{P}\left(U_1 \leq q_{1,U}, \ldots, U_{d_n} \leq q_{d_n,U}\right)\\ \leq& \sum_{j=1}^{d_n} \mathbb{P}\left(U_j > q_{j,U}\right) = \sum_{j=1}^{d_n} (1 - q_{j,U}) = \sum_{j=1}^{d_n} (1 - q_{k,U})\frac{\beta_j}{\beta_k} = \frac{1 - q_{k,U}}{\beta_k},\end{aligned}$$

which proves the assertion. □

Lemma C.2 bounds the quantile function of an F-distribution with $(1, c)$ degrees of freedom.

Lemma C.2 (Bounds on the F-quantiles)**.** *Let $F_{1,c}^{-1}(y)$ be the quantile function of an F-distribution with $(1,c)$ degrees of freedom, then*

$$c\left[\left(1-y^2\right)^{-\frac{1}{c}}-1\right] \leq F_{1,c}^{-1}(y) \leq c\left[\left(1-y^2\right)^{-\frac{2}{c-\frac{1}{2}}}-1\right].$$

Proof. We have from (Fujikoshi and Mukaihata, 1993, Theorem 4.2) that

$$c\left[\exp\left(\frac{\left(\chi_1^2\right)^{-1}(y)}{c}\right)-1\right] \leq F_{1,c}^{-1}(y) \leq c\left[\exp\left(\frac{\left(\chi_1^2\right)^{-1}(y)}{c-\frac{1}{2}}\right)-1\right],$$

with $\left(\chi_1^2\right)^{-1}(y)$ the quantile function of the chi-squared distribution with one degree of freedom. Moreover, we obtain for all $y \geq 0$

$$\begin{aligned} &\mathbb{P}\left(\chi_1^2 \leq y\right) = \mathbb{P}\left(-\sqrt{y} \leq Z \leq \sqrt{y}\right) = 2\Phi\left(\sqrt{y}\right)-1 \\ \iff &\left(\chi_1^2\right)^{-1}(y) = \Phi^{-1}\left(\frac{y+1}{2}\right)^2, \end{aligned}$$

where $\Phi^{-1}(y)$ is the quantile of the standard gaussian distribution. Furthermore, we have from (Johnson et al., 1994, (13.48), p. 115) that

$$\frac{1}{2}\left[1+\left(1-\exp\left(-\frac{x^2}{2}\right)\right)^{\frac{1}{2}}\right] \leq \Phi(x) \leq \frac{1}{2}\left[1+\left(1-\exp\left(-x^2\right)\right)^{\frac{1}{2}}\right]$$

and so for the quantile function one finds

$$\sqrt{-\log\left(1-(2y-1)^2\right)} \leq \Phi^{-1}(y) \leq \sqrt{-2\log\left(1-(2y-1)^2\right)}.$$

Combining the formulas proves the assertion. □

Proof of Lemma 3.1. First of all, (C.2) and the equation $|\mathcal{D}_k| = \lfloor n2^{-k}\rfloor$ yields

$$q_k = F_{1,2^k-1}^{-1}\left(q_{k,U}^{|\mathcal{D}_k|^{-1}}\right) = F_{1,2^k-1}^{-1}\left(q_{k,U}^{\lfloor n2^{-k}\rfloor^{-1}}\right) \leq F_{1,2^k-1}^{-1}\left(q_{k,U}^{2^k/n}\right).$$

Moreover, it follows from the Lemmas C.1 and C.2 that

$$\begin{aligned} q_k \leq F_{1,2^k-1}^{-1}\left(q_{k,U}^{2^k/n}\right) &\leq F_{1,2^k-1}^{-1}\left((1-\alpha\beta_k)^{2^k/n}\right) \\ &\leq \left(2^k-1\right)\left[\left(1-\left((1-\alpha\beta_k)^{2^k/n}\right)^2\right)^{-\frac{2}{(2^k-1)-\frac{1}{2}}}-1\right] \\ &\leq 2^k\left[\left(1-(1-\alpha\beta_k)^{2^k/n}\right)^{-\frac{4}{2^{k+1}-3}}-1\right]. \end{aligned}$$

Applying Bernoulli's inequality $(1-x)^c \leq 1-cx$ gives

$$q_k \leq 2^k\left[\left(1-(1-\alpha\beta_k)^{2^k/n}\right)^{-\frac{4}{2^{k+1}-3}}-1\right] \leq 2^k\left[\left(\frac{2^k\alpha\beta_k}{n}\right)^{-\frac{4}{2^{k+1}-3}}-1\right].$$

Moreover, for $x, c > 0$ the inequality $c^x \leq 1 + 2x\log(c)$ holds whenever $x\log(c) \leq 1$. Together with the assumption $k \geq 2$ we finally obtain

$$q_k \leq 2^k \left[\left(\frac{2^k \alpha \beta_k}{n} \right)^{-\frac{4}{2^{k+1}-3}} - 1 \right] \leq 4 \frac{2^{k+1}}{2^{k+1}-3} \log\left(\frac{n}{2^k \alpha \beta_k}\right) \leq 8 \log\left(\frac{n}{2^k \alpha \beta_k}\right)$$

if

$$2^{-k} \log\left(\frac{n}{2^k \alpha \beta_k}\right) \leq \frac{1}{2} \frac{2^{k+1}-3}{2^{k+1}} \leq \frac{1}{2}.$$

□

C.3. **Exponential deviation bounds.** For the subsequent proofs we need a bound for the distribution function of a single test statistic T_i^j (1.5) which is in our setting a bound for the c.d.f. of a non-central F-distribution.

Lemma C.3. *Let $Y_1, \ldots, Y_n$ be i.i.d. gaussian random variables with expectation $m \in \mathbb{R}$ and variance $s^2 > 0$. Let $x_+ := \max(x, 0)$. Then, for any $\delta \neq 0$, $q > 0$*

$$\begin{aligned} &\mathbb{P}\left(T_1^n(Y, m+\delta) \leq q\right) \\ &\leq \min_{z \geq 0} \left\{ \exp\left(-\frac{1}{2} \left(\frac{\Delta\sqrt{n}}{2} - \frac{q(1+z)}{\Delta\sqrt{n}} \right)_+^2 \right) + \exp\left(-(n-1) \frac{z - \log(1+z)}{2} \right) \right\}, \end{aligned} \tag{C.5}$$

where $\Delta := |\delta|/s$.

Proof. Let $\widetilde{T}_i^j(Y, m) := (j - i + 1)\left(\overline{Y}_{ij} - m\right)^2 / s^2$. Then,

$$T_1^n(Y, m+\delta) = \frac{\widetilde{T}_1^n(Y, m+\delta)}{\hat{s}_{1n}^2 / s^2}.$$

The statistics $\hat{s}_{1n}^2/s^2$ and $\widetilde{T}_1^n(Y, m+\delta)$ are independent, since $\widetilde{T}_1^n(Y, m+\delta)$ depends only on the mean $\overline{Y}_{1n}$. Hence, for all $z \geq 0$

$$\begin{aligned} \mathbb{P}\left(T_1^n(Y, m+\delta) \leq q\right) =& \mathbb{P}\left(\widetilde{T}_1^n(Y, m+\delta) \leq q \frac{\hat{s}_{1n}^2}{s^2} \right) \\ =& \mathbb{P}\left(\widetilde{T}_1^n(Y, m+\delta) \leq q \frac{\hat{s}_{1n}^2}{s^2} \,\middle|\, \frac{\hat{s}_{1n}^2}{s^2} \leq 1+z \right) \mathbb{P}\left(\frac{\hat{s}_{1n}^2}{s^2} \leq 1+z \right) \\ &+ \mathbb{P}\left(\widetilde{T}_1^n(Y, m+\delta) \leq q \frac{\hat{s}_{1n}^2}{s^2} \,\middle|\, \frac{\hat{s}_{1n}^2}{s^2} > 1+z \right) \mathbb{P}\left(\frac{\hat{s}_{1n}^2}{s^2} > 1+z \right) \\ \leq& \mathbb{P}\left(\widetilde{T}_1^n(Y, m+\delta) \leq q(1+z) \right) + \mathbb{P}\left(\frac{\hat{s}_{1n}^2}{s^2} > 1+z \right) \\ \leq& \exp\left(-\frac{1}{2} \left(\frac{\Delta\sqrt{n}}{2} - \frac{q(1+z)}{\Delta\sqrt{n}} \right)_+^2 \right) + \exp\left(-(n-1) \frac{z - \log(1+z)}{2} \right). \end{aligned}$$

The first term of the last inequality follows from (Frick et al., 2014, Lemma 7.3 and the proof) and the second from (Spokoiny and Zhilova, 2013, Theorem 2.1), since $(n-1)\hat{s}_{1n}^2/s^2 \sim \chi_{n-1}^2$.

It remains to show that the minimum in (C.5) is attained for some $z \geq 0$. The function $(\Delta\sqrt{n}/2 - q(1+z)/(\Delta\sqrt{n}))_+^2$ is strictly monotonically decreasing for $z > 0$ until the function value zero is attained for some finite z. The function $(n-1)(z - \log(1+z))$

is zero for $z = 0$ and strictly monotonically increasing on $[0, \infty)$. Therefore, the two continuous functions intersect and the minimum is attained for some $z \geq 0$. $\square$

The minimum in the last lemma cannot be determined analytically, but it can be computed numerically. In Lemma C.4 we estimate the right hand side further to obtain an explicit exponential bound.

Lemma C.4. *Let $Y_1, \ldots, Y_n$, $n \geq 4$, be i.i.d. gaussian random variables with expectation $m \in \mathbb{R}$ and variance $s^2 > 0$, then we have for all $q > 0$ with*

$$\frac{q}{n} \leq \frac{1}{8} \tag{C.6}$$

as well as for all $\delta \neq 0$ and $\Delta := |\delta|/s$ the bound

$$\mathbb{P}\left(T_1^n(Y, m+\delta) \leq q\right) \leq 2 \exp\left(-\frac{1}{48}\left(\sqrt{n}\Delta - \sqrt{2q}\right)_+^2\right). \tag{C.7}$$

Proof. Let $z > 0$ be arbitrary, but fixed. Then, it follows from Lemma C.3 that

$$\begin{aligned}\mathbb{P}\left(T_1^n(Y, m+\delta) \leq q\right) \leq& \exp\left(-\frac{1}{2}\left(\frac{\Delta\sqrt{n}}{2} - \frac{q(1+z)}{\Delta\sqrt{n}}\right)_+^2\right) + \exp\left(-(n-1)\frac{z - \log(1+z)}{2}\right) \\ \leq& 2\exp\left(-\min\left[\frac{1}{2}\left(\frac{\Delta\sqrt{n}}{2} - \frac{q(1+z)}{\Delta\sqrt{n}}\right)_+^2, (n-1)\frac{z-\log(1+z)}{2}\right]\right).\end{aligned}$$

The inequality

$$z - \log(1+z) \geq \frac{1}{2}\frac{z^2}{1+z} \geq \frac{1}{4}\min\left(z^2, z\right)$$

yields

$$\begin{aligned}&\mathbb{P}\left(T_1^n(Y, m+\delta) \leq q\right) \\ \leq& 2\exp\left(-\min\left[\frac{1}{8}n\left(\Delta - \frac{2q(1+z)}{\Delta n}\right)_+^2, \frac{1}{8}(n-1)\min\left(z^2, z\right)\right]\right) \\ \leq& 2\exp\left(-\frac{1}{8}(n-1)\min\left[\min\left[\left(\Delta - \frac{2q(1+z)}{\Delta n}\right)_+^2, z^2\right], \min\left[\left(\Delta - \frac{2q(1+z)}{\Delta n}\right)_+^2, z\right]\right]\right).\end{aligned}$$

Now, we minimize the r.h.s in $z \geq 0$. The functions z and z^2 are both increasing, the function $(\Delta - 2q(1+z)/(\Delta n))_+^2$ in contrast is decreasing in z. Therefore, both inner minima are attained and by solving the corresponding quadratic equations (note that we have to take the solution with $\Delta - 2q(1+z)/(\Delta n) \geq 0$) we get

$$\begin{aligned}&\mathbb{P}\left(T_1^n(Y, m+\delta) \leq q\right) \\ \leq& 2\exp\left(-\frac{1}{8}(n-1)\min\left[\left(\frac{\Delta - \frac{2q}{\Delta n}}{1 + \frac{2q}{\Delta n}}\right)_+^2, \frac{1 + 2\frac{2q}{\Delta n}\left(\Delta - \frac{2q}{\Delta n}\right)_+ - \sqrt{1 + 4\frac{2q}{\Delta n}\left(\Delta - \frac{2q}{\Delta n}\right)_+}}{2\left(\frac{2q}{\Delta n}\right)^2}\right]\right).\end{aligned}$$

Using the inequality $\sqrt{1+4x} \leq 1+2x-2x^2+4x^3$ for all $x > -1/4$ with $x = 2q/(\Delta n)(\Delta - 2q/(\Delta n))_+$ we find

$$\begin{aligned}
&\mathbb{P}\left(T_1^n(Y, m+\delta) \leq q\right) \\
\leq &2\exp\left(-\frac{1}{8}\frac{n-1}{n} n \min\left[\left(\frac{\Delta - \frac{2q}{\Delta n}}{1+\frac{2q}{\Delta n}}\right)_+^2, \left(\Delta - \frac{2q}{\Delta n}\right)_+^2\left[1 - 2\frac{2q}{\Delta n}\left(\Delta - \frac{2q}{\Delta n}\right)_+\right]\right]\right).
\end{aligned}$$

Next, we consider the two terms in the minimum separately. We assume w.l.o.g. that $\sqrt{2q}/(\Delta\sqrt{n}) \leq 1$, since otherwise the r.h.s. in (C.7) is two. For the first term we distinguish the cases $2q > \Delta n$ and $2q \leq \Delta n$. If $2q \leq \Delta n$ is satisfied, then

$$n\left(\frac{\Delta - \frac{2q}{\Delta n}}{1+\frac{2q}{\Delta n}}\right)_+^2 \geq \frac{1}{4}n\left(\Delta - \frac{2q}{\Delta n}\right)_+^2 = \frac{1}{4}\left(\sqrt{n}\Delta - \frac{2q}{\Delta\sqrt{n}}\right)_+^2 \geq \frac{1}{4}\left(\sqrt{n}\Delta - \sqrt{2q}\right)_+^2.$$

For the other case, when $2q > \Delta n$ holds, we obtain with $q/n \leq 1/8$

$$\begin{aligned}
n\left(\frac{\Delta - \frac{2q}{\Delta n}}{1+\frac{2q}{\Delta n}}\right)_+^2 \geq &\frac{1}{4}n\left(\frac{\Delta - \frac{2q}{\Delta n}}{\frac{2q}{\Delta n}}\right)_+^2 = \frac{1}{4}n\left(\frac{n\Delta^2}{2q} - 1\right)_+^2 \\
= &\frac{1}{4}\frac{n}{2q}\left(\frac{(\sqrt{n}\Delta)^2}{\sqrt{2q}} - \sqrt{2q}\right)_+^2 \geq \left(\sqrt{n}\Delta - \sqrt{2q}\right)_+^2.
\end{aligned}$$

For the second term it follows with $q/n \leq 1/8$ that

$$\begin{aligned}
n\left(\Delta - \frac{2q}{\Delta n}\right)_+^2\left[1 - 2\frac{2q}{\Delta n}\left(\Delta - \frac{2q}{\Delta n}\right)_+\right] &= \left(\sqrt{n}\Delta - \frac{2q}{\Delta\sqrt{n}}\right)_+^2\left[1 - 4\frac{q}{n}\left(1 - \frac{2q}{\Delta^2 n}\right)_+\right] \\
&\geq \left(\sqrt{n}\Delta - \sqrt{2q}\right)_+^2 \frac{1}{2}.
\end{aligned}$$

This yields

$$\begin{aligned}
\mathbb{P}\left(T_1^n(Y, m+\delta) \leq q\right) \leq &2\exp\left(-\frac{1}{32}\frac{n-1}{n}\left(\sqrt{n}\Delta - \sqrt{2q}\right)_+^2\right) \\
\leq &2\exp\left(-\frac{1}{48}\left(\sqrt{n}\Delta - \sqrt{2q}\right)_+^2\right).
\end{aligned}$$

□

C.4. **Proofs of Section 3.**

Proof of Theorem 3.3. The estimated number of change-points $\hat{K}$ is by its definition in (3.3) equal to the minimal number of change-points of all feasible functions. Therefore, all functions with the true number of change-points (or less change-points) have to be infeasible, if the number of change-points is overestimated. Hence, by (2.3)

$$\begin{aligned}
\sup_{(\mu,\sigma^2)\in\mathcal{S}} \mathbb{P}_{(\mu,\sigma^2)}\left(\hat{K} > K\right) &\leq \sup_{(\mu,\sigma^2)\in\mathcal{S}} \mathbb{P}_{(\mu,\sigma^2)}\left(\max_{\left[\frac{i}{n},\frac{j}{n}\right]\in\mathcal{D}(\mu)} \left[T_i^j\left(Y, \mu([i/n, j/n])\right) - q_{ij}\right] > 0\right) \\
&\leq \mathbb{P}_{(0,1)}\left(\max_{\left[\frac{i}{n},\frac{j}{n}\right]\in\mathcal{D}} \left[T_i^j\left(Y, 0)\right) - q_{ij}\right] > 0\right) = \alpha,
\end{aligned}$$

where the last inequality follows from $\mathcal{D}(\mu) \subset \mathcal{D}$ and the fact that the distribution of $T_i^j(Y, \mu([i/n, j/n]))$ does not depend on $\mu(\cdot)$ and $\sigma(\cdot)$, as these are constant on intervals in $\mathcal{D}(\mu)$. $\square$

Proof of Theorem 3.4. First of all, we show that it is enough to prove the result for $\mu \equiv 0$ and $\sigma^2 \equiv 1$ and hence $K = 0$. We have

$$\begin{aligned}
&\sup_{(\mu,\sigma^2)\in\mathcal{S}} \mathbb{P}_{(\mu,\sigma^2)}\left(\hat{K} > K + 2k\right) \\
=&\sup_{(\mu,\sigma^2)\in\mathcal{S}} \mathbb{P}_{(\mu,\sigma^2)}\left(\max_{\left[\frac{i}{n},\frac{j}{n}\right]\in\mathcal{D}(\tilde{\mu})} \left[T_i^j(Y, \tilde{\mu}([i/n, j/n])) - q_{ij}\right] > 0 \ \forall\, \tilde{\mu} \in \mathcal{M}\ s.t.\ |\mathcal{I}(\tilde{\mu})| \le K + 2k\right) \\
\le&\sup_{(\mu,\sigma^2)\in\mathcal{S}} \mathbb{P}_{(\mu,\sigma^2)}\Bigg(\max_{\left[\frac{i}{n},\frac{j}{n}\right]\in\mathcal{D}(\tilde{\mu})} \left[T_i^j(Y, \tilde{\mu}([i/n, j/n])) - q_{ij}\right] > 0 \\
&\qquad\qquad \forall\, \tilde{\mu} \in \mathcal{M}\ s.t.\ \mathcal{I}(\mu) \subset \mathcal{I}(\tilde{\mu}), |\mathcal{I}(\tilde{\mu})| \le K + 2k\Bigg) \\
\le&\mathbb{P}_{(0,1)}\left(\max_{\left[\frac{i}{n},\frac{j}{n}\right]\in\mathcal{D}(\tilde{\mu})} \left[T_i^j(Y, \tilde{\mu}([i/n, j/n])) - q_{ij}\right] > 0 \ \forall\, \tilde{\mu} \in \mathcal{M}\ s.t.\ |\mathcal{I}(\tilde{\mu})| \le 2k\right) \\
=&\mathbb{P}_{(0,1)}\left(\hat{K} > 2k\right),
\end{aligned}$$

where the last inequality follows from the same argument as in the proof of Theorem 3.3. Now, we define $R_0 := 0$ and iteratively

$$R_{k+1} := \min\{t > R_k : \exists\ s\ s.t.\ R_k < s < t \text{ and } [s/n, t/n] \in \mathcal{D}, T_s^t(Y, 0) > q_{\log_2(t-s+1)}\},$$

with the convention $\min\emptyset = \infty$. Then,

$$\mathbb{P}_{0,1}(R_{k+1} \le n | R_1 = t) \le \mathbb{P}_{0,1}(R_k \le n) \text{ for all } t \in \{1, \ldots, n\},$$

since for the l.h.s. the remaining k rejections $R_2, \ldots, R_{k+1}$ have to be in $\{t+1, \ldots, n\}$ instead of $\{1, \ldots, n\}$. It follows

$$\begin{aligned}
\mathbb{P}_{0,1}(\hat{K} > 2k) &\le \mathbb{P}_{0,1}(R_{k+1} \le n) = \sum_{t=1}^{n} \mathbb{P}_{0,1}(R_{k+1} \le n | R_1 = t)\mathbb{P}_{0,1}(R_1 = t) \\
\le&\mathbb{P}_{0,1}(R_1 \le n)\mathbb{P}_{0,1}(R_k \le n) \le \cdots \le \mathbb{P}_{0,1}(R_1 \le n)^{k+1} \le \alpha^{k+1},
\end{aligned}$$

where the last inequality is given by Theorem 3.3. It follows

$$\begin{aligned}
\sup_{(\mu,\sigma^2)\in\mathcal{S}} \mathbb{E}_{(\mu,\sigma^2)}\left[(\hat{K} - K)_+\right] &= \sup_{(\mu,\sigma^2)\in\mathcal{S}} \sum_{k=0}^{\infty} \mathbb{P}_{(\mu,\sigma^2)}\left(\hat{K} - K > k\right) \\
&\le \sup_{(\mu,\sigma^2)\in\mathcal{S}} 2\sum_{k=0}^{\infty} \mathbb{P}_{(\mu,\sigma^2)}\left(\hat{K} - K > 2k\right) \le 2\sum_{k=0}^{\infty} \alpha^{k+1} = \frac{2\alpha}{1-\alpha}.
\end{aligned}$$

$\square$

The following theorem is sharper version of 3.5 that shows different probabilities for the detection of the change-points.

Theorem C.5 (Underestimation control II)**.** *Let* $\lambda_j := \tau_{j+1} - \tau_j$ *and* $k_{n,j} := \lfloor \log_2(n\lambda_j/4) \rfloor$, $j = 0, \ldots, K$, *as well as* $\delta_j := |m_j - m_{j-1}|$ *and*

$$\eta_j := \left[1 - 3\exp\left(-\frac{1}{48}\left(\sqrt{\frac{n\lambda_{j-1}\delta_j^2}{32\sigma_{j-1}^2}} - \sqrt{16\log\left(\frac{8}{\lambda_j \alpha \beta_{k_{n,j-1}}}\right)}\right)_+^2\right)\right]_+ \times$$
$$\left[1 - 3\exp\left(-\frac{1}{48}\left(\sqrt{\frac{n\lambda_j\delta_j^2}{32\sigma_j^2}} - \sqrt{16\log\left(\frac{8}{\lambda_j \alpha \beta_{k_{n,j}}}\right)}\right)_+^2\right)\right]_+ ,$$

$j = 1, \ldots, K$. *Under the assumptions of Theorem 3.3 and if* $n\lambda_j \geq 32$ *and*

$$(n\lambda_j)^{-1}\log\left(\frac{8}{\lambda_j \alpha \beta_{k_{n,j}}}\right) \leq \frac{1}{512}$$

are satisfied for all $j = 1, \ldots, K$, *then*

$$\mathbb{P}_{(\mu,\sigma^2)}\left(\hat{K} < K\right) \leq 1 - \prod_{j=1}^{K} \eta_j \;\; \textit{and} \;\; \mathbb{E}_{(\mu,\sigma^2)}\left[\left(K - \hat{K}\right)_+\right] \leq \sum_{j=1}^{K} (1 - \eta_j)\,.$$

Proof. For each $j = 1, \ldots, K$ we consider the disjoint intervals $I_j := [\tau_j - \lambda_{j-1}/2, \tau_j + \lambda_j/2)$ and split them into disjoint intervals $I_j^+ \cup I_j^- = I_j$ such that $\mu(t) = \mu^+ \; \forall \; t \in I_j^+$ and $\mu(t) = \mu^- \; \forall \; t \in I_j^-$, with $\mu^+ := \max(m_{j-1}, m_j)$ and $\mu^- := \min(m_{j-1}, m_j)$. Without loss of generality we assume $\mu^+ = m_{j-1}$ and $\mu^- = m_j$ in the following. Then, there exists subintervals $J_j^+ \subset I_j^+$ and $J_j^- \subset I_j^-$ with $J_j^+, J_j^- \in \mathcal{D}$ that have length $\lambda_{j-1}^* := n^{-1}2^{\lfloor \log_2(n\lambda_{j-1}/4)\rfloor} = n^{-1}2^{k_{n,j-1}} \geq \lambda_{j-1}/8$, since $n|I_j^+| = n\lambda_{j-1}/2 \geq 3$, and $\lambda_j^* := n^{-1}2^{\lfloor \log_2(n\lambda_j/4)\rfloor} = n^{-1}2^{k_{n,j}} \geq \lambda_j/8$, since $n|I_j^-| = n\lambda_j/2 \geq 3$, respectively. It follows

$$\begin{aligned}
&\mathbb{P}_{(\mu,\sigma^2)}\left(\hat{K} < K\right) = 1 - \mathbb{P}_{(\mu,\sigma^2)}\left(\hat{K} \geq K\right) \\
&\leq 1 - \mathbb{P}_{(\mu,\sigma^2)}\left(\nexists\; \hat{\mu} \in C(\mathbf{q}),\; j \in \{1, \ldots, K\} : \hat{\mu} \text{ is constant on } I_j\right) \\
&\leq 1 - \mathbb{P}_{(\mu,\sigma^2)}\Big(\forall\; j \in \{1, \ldots, K\} : \quad \nexists\; \hat{m} \leq (m_{j-1} + m_j)/2 : T_{J_j^+}(Y, \hat{m}) \leq q_{k_{n,j-1}} \text{ and} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad \nexists\; \hat{m} \geq (m_{j-1} + m_j)/2 : T_{J_j^-}(Y, \hat{m}) \leq q_{k_{n,j}}\Big) \\
&\leq 1 - \prod_{j=1}^{K} \mathbb{P}_{(\mu,\sigma^2)}\Big(\nexists\; \hat{m} \leq (m_{j-1} + m_j)/2 : T_{J_j^+}(Y, \hat{m}) \leq q_{k_{n,j-1}} \text{ and} \\
&\qquad\qquad\qquad \nexists\; \hat{m} \geq (m_{j-1} + m_j)/2 : T_{J_j^-}(Y, \hat{m}) \leq q_{k_{n,j}}\Big),
\end{aligned}$$

where we used in the last inequality that the events are independent, since all intervals are disjoint. We denote by $Z_1, \ldots, Z_n$ i.i.d. standard normally distributed random variables. It follows from once again from the independence due to disjoint intervals and from the

Lemmas 7.1 in (Frick et al., 2014), C.4 and 3.1 that

$$\begin{aligned}
&\mathbb{P}_{(\mu,\sigma^2)}\Big(\nexists\ \hat{m} \le (m_{j-1}+m_j)/2 : T_{J_j^+}(Y,\hat{m}) \le q_{k_{n,j-1}} \text{ and} \\
&\qquad\qquad \nexists\ \hat{m} \ge (m_{j-1}+m_j)/2 : T_{J_j^-}(Y,\hat{m}) \le q_{k_{n,j}}\Big) \\
&\ge \Big[1-\mathbb{P}_{(\mu,\sigma^2)}\Big(\exists\ \hat{m} \le (m_{j-1}+m_j)/2 : T_{J_j^+}(Y,\hat{m}) \le q_{k_{n,j-1}}\Big)\Big] \times \\
&\quad \Big[1-\mathbb{P}_{(\mu,\sigma^2)}\Big(\exists\ \hat{m} \ge (m_{j-1}+m_j)/2 : T_{J_j^-}(Y,\hat{m}) \le q_{k_{n,j}}\Big)\Big] \ge \eta_j,
\end{aligned}$$

since

$$\begin{aligned}
&\mathbb{P}_{(\mu,\sigma^2)}\Big(\exists\ \hat{m} \le (m_{j-1}+m_j)/2 : T_{J_j^+}(Y,\hat{m}) \le q_{k_{n,j-1}}\Big) \\
&\le \mathbb{P}_{(\mu,\sigma^2)}\Big(\overline{Y}_{J_j^+} \le (m_{j-1}+m_j)/2 \text{ or } T_{J_j^+}(Y,(m_{j-1}+m_j)/2) \le q_{k_{n,j-1}}\Big) \\
&\le \mathbb{P}_{(\mu,\sigma^2)}\Big(\overline{Y}_{J_j^+} \le (m_{j-1}+m_j)/2\Big) + \mathbb{P}_{(\mu,\sigma^2)}\Big(T_{J_j^+}(Y,(m_{j-1}+m_j)/2) \le q_{k_{n,j-1}}\Big) \\
&\le \mathbb{P}\Big(\overline{Z}_{[0,\lambda_{j-1}^*]} \ge \frac{\delta_j}{2\sigma_{j-1}}\Big) + \mathbb{P}\Big(T_{[0,\lambda_{j-1}^*]}\left(Z,\frac{\delta_j}{2\sigma_{j-1}}\right) \le q_{k_{n,j-1}}\Big) \\
&\le \exp\left(-\frac{1}{64}\frac{n\lambda_{j-1}\delta_j^2}{\sigma_{j-1}^2}\right) + 2\exp\left(-\frac{1}{48}\left(\sqrt{\frac{n\lambda_{j-1}\delta_j^2}{32\sigma_{j-1}^2}} - \sqrt{2q_{k_{n,j-1}}}\right)_+^2\right) \\
&\le 3\exp\left(-\frac{1}{48}\left(\sqrt{\frac{n\lambda_{j-1}\delta_j^2}{32\sigma_{j-1}}} - \sqrt{16\log\left(\frac{8}{\lambda_{j-1}\alpha\beta_{k_{n,j-1}}}\right)}\right)_+^2\right)
\end{aligned}$$

and the second term by symmetry arguments. Moreover, it follows

$$\begin{aligned}
&\mathbb{E}_{(\mu,\sigma^2)}\left[\left(K-\hat{K}\right)_+\right] \\
\le&\mathbb{E}_{(\mu,\sigma^2)}\left[\sum_{j=1}^{K} \mathbb{1}_{\exists\,\hat{m}\le(m_{j-1}+m_j)/2:\,T_{J_j^+}(Y,\hat{m})\le q_{k_{n,j-1}} \text{ or } \exists\,\hat{m}\ge(m_{j-1}+m_j)/2:\,T_{J_j^-}(Y,\hat{m})\le q_{k_{n,j}}}\right] \\
\le&\sum_{j=1}^{K}(1-\eta_j)\,.
\end{aligned}$$

□

Proof of Theorem 3.5. The proof is analogue to the proof of Theorem C.5, but with $I_j = [\tau_j - \lambda/2, \tau_j + \lambda/2)$. □

Proof of Theorem 3.7. We prove the theorem with the Borel-Cantelli lemma. It follows from Theorems 3.3 and 3.5 that

$$\begin{aligned}
&\sup_{(\mu,\sigma^2)\in\mathcal{S}_{\Delta,\lambda}} \mathbb{P}_{(\mu,\sigma^2)}\left(\hat{K}_n \neq K\right) \\
=&\sup_{(\mu,\sigma^2)\in\mathcal{S}_{\Delta,\lambda}} \mathbb{P}_{(\mu,\sigma^2)}\left(\hat{K}_n > K\right) + \sup_{(\mu,\sigma^2)\in\mathcal{S}_{\Delta,\lambda}} \mathbb{P}_{(\mu,\sigma^2)}\left(\hat{K}_n < K\right) \\
\leq&\alpha_n + 1 - \left[1 - 3\exp\left(-\frac{1}{48}\left(\sqrt{\frac{n\lambda\Delta^2}{32}} - \sqrt{16\log\left(\frac{8}{\lambda\alpha_n\beta_{k_n,n}}\right)}\right)_+^2\right)\right]_+^{2K} \\
\leq&\alpha_n + 6K\exp\left(-\frac{1}{48}\left(\sqrt{\frac{n\lambda\Delta^2}{32}} - \sqrt{16\log\left(\frac{8}{\lambda\alpha_n\beta_{k_n,n}}\right)}\right)_+^2\right),
\end{aligned}$$

since under the given assumptions the conditions of Theorem 3.5 are satisfied. The upper bounds for the error probabilities are summable if (3.6) is satisfied. ☐

Lemma C.6 (Confidence set). *Assume the setting and assumptions of Theorem 3.5 and let $C(\boldsymbol{q})$ be as in (3.7) with significance level α and weights $\beta_1, \ldots, \beta_{d_n}$. Let $\mathcal{S}_{\Delta,\lambda}$ be as in (3.4) with $\Delta, \lambda > 0$ arbitrary, but fixed, and $k_n := \lfloor \log_2(n\lambda/4)\rfloor$. If $n\lambda \geq 32$ and*

$$\frac{\log\left(\frac{8}{\lambda\alpha_n\beta_{k_n}}\right)}{n\lambda} \leq \frac{1}{512}$$

hold, then uniformly in $\mathcal{S}_{\Delta,\lambda}$

$$\mathbb{P}_{(\mu,\sigma^2)}\left(\mu \in C\left(\boldsymbol{q}\right)\right) \geq 1 - \alpha - (1 - \eta^K),$$

with η like in Theorem 3.5.

Proof. It follows from the definition of $C(\mathbf{q})$ in (3.7) as well as from Theorems 3.3 and 3.5 that

$$\begin{aligned}
&\inf_{(\mu,\sigma^2)\in\mathcal{S}_{\Delta,\lambda}} \mathbb{P}_{(\mu,\sigma^2)}\left(\mu \in C\left(\mathbf{q}\right)\right) \\
=&\inf_{(\mu,\sigma^2)\in\mathcal{S}_{\Delta,\lambda}} \mathbb{P}_{(\mu,\sigma^2)}\left(\max_{[\frac{i}{n},\frac{j}{n}]\in\mathcal{D}(\mu)} \left[T_i^j\big(Y, \mu([i/n, j/n])\big) - q_{ij}\right] \leq 0, \hat{K} = K\right) \\
=&\inf_{(\mu,\sigma^2)\in\mathcal{S}_{\Delta,\lambda}} \mathbb{P}_{(\mu,\sigma^2)}\left(\max_{[\frac{i}{n},\frac{j}{n}]\in\mathcal{D}(\mu)} \left[T_i^j\big(Y, \mu([i/n, j/n])\big) - q_{ij}\right] \leq 0, \hat{K} \geq K\right) \\
\geq&\inf_{(\mu,\sigma^2)\in\mathcal{S}_{\Delta,\lambda}} \mathbb{P}_{(\mu,\sigma^2)}\left(\max_{[\frac{i}{n},\frac{j}{n}]\in\mathcal{D}(\mu)} \left[T_i^j\big(Y, \mu([i/n, j/n])\big) - q_{ij}\right] \leq 0\right) - \sup_{(\mu,\sigma^2)\in\mathcal{S}_{\Delta,\lambda}} \mathbb{P}_{(\mu,\sigma^2)}\left(\hat{K} < K\right) \\
\geq&1 - \alpha - (1 - \eta^K).
\end{aligned}$$

☐

Proof of Theorem 3.8. The statement is a direct consequence of Lemma C.6. ☐

Lemma C.7 (Change-point locations)**.** *Assume the setting of Lemma C.6. If c_n is a sequence with $0 < c_n \leq \lambda/2$ and $k_n := \lfloor \log_2(nc_n/2) \rfloor$ such that $nc_n \geq 16$ and*

$$\frac{\log\left(\frac{4}{c_n \alpha \beta_{k_n}}\right)}{nc_n} \leq \frac{1}{256} \tag{C.8}$$

hold, then uniformly in $\mathcal{S}_{\Delta,\lambda}$

$$\mathbb{P}_{(\mu,\sigma^2)}\left(\sup_{\hat{\mu}\in C(\boldsymbol{q}_n)} \max_{\tau\in\mathcal{I}(\mu)} \min_{\hat{\tau}\in\mathcal{I}(\hat{\mu})} |\hat{\tau}-\tau| > c_n\right)$$
$$\leq 1 - \left[1 - 3\exp\left(-\frac{1}{48}\left(\sqrt{\frac{nc_n\Delta^2}{16}} - \sqrt{16\log\left(\frac{4}{c_n\alpha\beta_{k_n}}\right)}\right)_+^2\right)\right]_+^{2K}.$$

Proof. Analogously to the proof of Theorem C.5 we have

$$\sup_{(\mu,\sigma^2)\in\mathcal{S}_{\Delta,\lambda}} \mathbb{P}_{(\mu,\sigma^2)}\left(\sup_{\hat{\mu}\in C(\mathbf{q}_n)} \max_{\tau\in\mathcal{I}(\mu)} \min_{\hat{\tau}\in\mathcal{I}(\hat{\mu})} |\hat{\tau}-\tau| > c_n\right)$$
$$\leq \sup_{(\mu,\sigma^2)\in\mathcal{S}_{\Delta,\lambda}} \mathbb{P}_{(\mu,\sigma^2)}\big(\exists\, j\in\{1,\ldots,K\} \text{ and } \hat{\mu}\in C(\mathbf{q}_n) : \hat{\mu} \text{ is constant on } [\tau_j - c_n, \tau_j + c_n)\big)$$
$$\leq 1 - \left[1 - 3\exp\left(-\frac{1}{48}\left(\sqrt{\frac{nc_n\Delta^2}{16}} - \sqrt{16\log\left(\frac{4}{c_n\alpha\beta_{k_n}}\right)}\right)_+^2\right)\right]_+^{2K}.$$

□

Proof of Theorem 3.9. For n large enough such that (3.9) guarantees the assumption of Lemma C.7 it follows

$$\mathbb{P}_{(\mu,\sigma^2)}\left(\sup_{\hat{\mu}\in C(\mathbf{q}_n)} \max_{j=1,\ldots,K} c_n^{-1}|\tau_j - \hat{\tau}_j| > 1\right)$$
$$\leq \mathbb{P}_{(\mu,\sigma^2)}\left(\hat{K} > K \text{ or } \exists\, \hat{\mu}\in C(\mathbf{q}_n),\ j\in\{1,\ldots,K\} : \hat{\mu} \text{ is constant on } [\tau_j - c_n, \tau_j + c_n]\right)$$
$$\leq \mathbb{P}_{(\mu,\sigma^2)}\left(\hat{K} > K\right) + \mathbb{P}_{(\mu,\sigma^2)}\left(\exists\, \hat{\mu}\in C(\mathbf{q}_n),\ j\in\{1,\ldots,K\} : \hat{\mu} \text{ is constant on } [\tau_j - c_n, \tau_j + c_n]\right)$$
$$\leq \alpha_n + \left(1 - \left[1 - 3\exp\left(-\frac{1}{48}\left(\sqrt{\frac{nc_n\Delta^2}{16}} - \sqrt{16\log\left(\frac{4}{c_n\alpha_n\beta_{k_n,n}}\right)}\right)_+^2\right)\right]_+^{2K}\right).$$

The assertion follows from $\alpha_n \to 0$ and

$$\lim_{n\to\infty} \sqrt{\frac{nc_n\Delta^2}{16}} - \sqrt{16\log\left(\frac{4}{c_n\alpha_n\beta_{k_n,n}}\right)} = \infty,$$

whereby latter one is direct consequence of (3.9). □

The following theorem deals with the detection of single vanishing bump against a noisy background.

Theorem C.8 (Single vanishing bump)**.** *Assume the heterogeneous gaussian change-point model* (1.2) *with sequences of bump signals $\mu_n(t) := m_0 + \delta_n \mathbb{1}_{I_n}(t)$ and $\sigma_n(t) :=$*

$\mathbb{1}_{I_n^C}(t) + s_n \mathbb{1}_{I_n}(t)$, where $\delta_n \neq 0$ is a sequence of change-point sizes, $s_n > 0$ a sequence of standard deviations on $I_n \in \mathcal{D}$, which is a sequence of intervals with $|I_n| \to 0$. Let $k_n := \lfloor \log_2(n|I_n|) \rfloor$ and $\Delta_n := |\delta_n|/s_n$ be the sequence of the signal to noise ratios. Let $(\hat{K}_n)_n$, α_n and $\beta_{1,n}, \ldots, \beta_{d_n,n}$ be as in Theorem 3.10. We further assume

$$\sqrt{n|I_n|}\Delta_n \geq (4 + \epsilon_n)\sqrt{-\log(|I_n|)}, \tag{C.9}$$

with possibly $\epsilon_n \to 0$, but such that $\epsilon_n \sqrt{-\log(|I_n|)} \to \infty$ and

$$\limsup_{n\to\infty} \frac{\sqrt{-\log\left(\alpha_n \beta_{k_n,n}\right)}}{\epsilon_n \sqrt{-\log(|I_n|)}} < \frac{1}{4},$$

$$\liminf_{n\to\infty} \frac{n|I_n|}{\log(n)} > 64 \text{ and } \lim_{n\to\infty} \frac{\log\left(\alpha_n \beta_{k_n,n}\right)}{n|I_n|} = 0, \tag{C.10}$$

$$\lim_{n\to\infty} s_n \frac{\sqrt{|I_n^C|}}{\sqrt{|I_n|}} = \infty \text{ and} \tag{C.11}$$

$$\liminf_{n\to\infty} \frac{\log(\beta_{k_n,n})}{\log(\beta_{\min,n})} > 0, \text{ with } \beta_{\min,n} := \min\{\beta_{1,n}, \ldots, \beta_{d_n,n}\}. \tag{C.12}$$

Then,

$$\lim_{n\to\infty} \mathbb{P}_{(\mu_n,\sigma_n^2)}\left(\hat{K}_n > 0\right) = 1. \tag{C.13}$$

Conditions (C.9) and (C.10) are the main assumptions of the theorem to detect the vanishing signal on I_n. We discussed them together with the conditions of Theorem 3.10 in Section 3.3. We also need the weak technical conditions (C.11) and (C.12) on the size of $|I_n^C|$ and the minimal weight $\beta_{\min,n}$ to ensure that the detection power on the complement I_n^C is large enough, too. Condition (C.12) is for instance fulfilled by uniform weights $\beta_{1,n} = \cdots = \beta_{d_n,n} = 1/d_n$, but many other choices are possible, too. We further assumed $I_n \in \mathcal{D}$, otherwise we have to replace I_n by the largest subinterval which is an element of the dyadic partition. Such an interval exists always and has at least length $n^{-1}2^{\lfloor \log_2(n|I_n|/2)\rfloor} > |I_n|/4$. Therefore, omitting the condition $I_n \in \mathcal{D}$ would not change the rate. It is possible to strengthen (C.13) further to $\lim_{n\to\infty} \mathbb{P}_{(\mu_n,\sigma_n^2)}(\hat{K}_n \geq K) = 1$ if we increase all constants a little bit.

Proof of Theorem C.8. We denote by J_n the longest subinterval $J_n \subset I_n^C$ which is part of the dyadic partition. Such an interval exists (at least for n large enough) always, since $|I_n| \to 0$, and has at least length $|I_n^C|/8$. Moreover, let $k_n := \log_2(n|I_n|)$ and $l_n := \log_2(n|J_n|)$. Then, the Lemmas 7.1 in (Frick et al., 2014) and C.4 yield for any

$\theta_n > 0$

$$
\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}_{(\mu_n,\sigma_n^2)}\left(\hat{K}_n > 0\right) \\
&= \lim_{n\to\infty} 1 - \mathbb{P}_{(\mu_n,\sigma_n^2)}\left(\hat{\mu} \text{ is constant}\right) \\
&\geq \lim_{n\to\infty} 1 - \mathbb{P}_{(\mu_n,\sigma_n^2)}\left(\exists\, \hat{m} \leq m_0 + \theta_n : T_{I_n}(Y,\hat{m}) \leq q_{k_n} \text{ or } \exists\, \hat{m} \geq m_0 + \theta_n : T_{J_n}(Y,\hat{m}) \leq q_{l_n}\right) \\
&\geq \lim_{n\to\infty} 1 - \mathbb{P}_{(\mu_n,\sigma_n^2)}\left(\exists\, \hat{m} \leq m_0 + \theta_n : T_{I_n}(Y,\hat{m}) \leq q_{k_n}\right) \\
&\qquad\qquad - \mathbb{P}_{(\mu_n,\sigma_n^2)}\left(\exists\, \hat{m} \geq m_0 + \theta_n : T_{J_n}(Y,\hat{m}) \leq q_{l_n}\right) \\
&\geq \lim_{n\to\infty} 1 - \mathbb{P}_{(\mu_n,\sigma_n^2)}\left(\overline{Y}_{I_n} \leq m_0 + \theta_n\right) - \mathbb{P}_{(\mu_n,\sigma_n^2)}\left(T_{I_n}(Y, m_0+\theta_n) \leq q_{k_n}\right) \\
&\qquad\qquad - \mathbb{P}_{(\mu_n,\sigma_n^2)}\left(\overline{Y}_{J_n} \geq m_0 + \theta_n\right) - \mathbb{P}_{(\mu_n,\sigma_n^2)}\left(T_{J_n}(Y, m_0+\theta_n) \leq q_{l_n}\right) \\
&\geq \lim_{n\to\infty} 1 - 2\mathbb{P}_{(\mu_n,\sigma_n^2)}\left(T_{I_n}(Y, m_0+\theta_n) \leq q_{k_n}\right) - 2\mathbb{P}_{(\mu_n,\sigma_n^2)}\left(T_{J_n}(Y, m_0+\theta_n) \leq q_{l_n}\right) \\
&\geq \lim_{n\to\infty} 1 - 4\exp\left(-\frac{1}{48}\left(\Gamma_{I_n}\right)_+^2\right) - 4\exp\left(-\frac{1}{48}\left(\Gamma_{J_n}\right)_+^2\right) = 1,
\end{aligned}
$$

if

$$\Gamma_{I_n} := \sqrt{n|I_n|}\frac{\delta_n - \theta_n}{s_n} - \sqrt{2q_{k_n}} \to \infty \text{ and } \Gamma_{J_n} := \sqrt{n|J_n|}\theta_n - \sqrt{2q_{l_n}} \to \infty,$$

and if the conditions of Lemma C.4 are satisfied. This is the case, since $n|I_n| \to \infty$ and $n|J_n| \to \infty$, because of (C.10) and $|I_n| \to 0$, as well as $q_{k_n}/(n|I_n|) \leq 1/8$ and $q_{l_n}/(n|J_n|) \leq 1/8$ hold at least for n large enough: The first one is a direct consequence of Lemma 3.1 and (C.10)

$$\frac{q_{k_n}}{n|I_n|} \leq \frac{8\log\left(\frac{1}{|I_n|\alpha_n\beta_{k_n,n}}\right)}{n|I_n|} \leq \frac{1}{8},$$

since then the assumptions of Lemma 3.1 are also fulfilled. The second inequality follows from Lemma 3.1, (C.10) and (C.12) as well as the fact that $|I_n|/|J_n| \to 0$

$$\lim_{n\to\infty} \frac{q_{l_n}}{n|J_n|} \leq \lim_{n\to\infty} \frac{8\log\left(\frac{1}{|J_n|\alpha_n\beta_{l_n,n}}\right)}{n|J_n|} \leq \lim_{n\to\infty} \frac{8\log\left(\frac{1}{|J_n|\alpha_n\beta_{l_n,n}}\right)}{8\log\left(\frac{1}{|I_n|\alpha_n\beta_{k_n,n}}\right)} \frac{|I_n|}{|J_n|} \frac{8\log\left(\frac{1}{|I_n|\alpha_n\beta_{k_n,n}}\right)}{n|I_n|} \to 0,$$

since then the assumptions of Lemma 3.1 are also fulfilled.
We define now $\theta_n = \sqrt{\gamma_n/n}$ via the equation

$$\sqrt{\frac{\gamma_n|I_n|}{s_n^2}} = c\epsilon_n\sqrt{\log\left(\frac{1}{|I_n|}\right)}$$

for $0 < c < 1$. Then, it follows from Lemma 3.1 and from $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ for $x, y > 0$ together with the assumptions of the theorem that

$$\begin{aligned}
\Gamma_{I_n} =& \sqrt{n|I_n|}\frac{\delta_n - \theta_n}{s_n} - \sqrt{2q_{k_n}} \\
=& \sqrt{n|I_n|\delta_n^2/s_n^2} - \sqrt{\gamma_n|I_n|/s_n^2} - \sqrt{2q_{k_n}} \\
\geq& \sqrt{n|I_n|\Delta_n^2} - \sqrt{\gamma_n|I_n|/s_n^2} - \sqrt{16\log\left(\frac{1}{|I_n|\alpha_n\beta_{k_n,n}}\right)} \\
\geq& (4+\epsilon_n)\sqrt{\log\left(\frac{1}{|I_n|}\right)} - c\epsilon_n\sqrt{\log\left(\frac{1}{|I_n|}\right)} - 4\sqrt{\log\left(\frac{1}{|I_n|}\right)} - 4\sqrt{\log\left(\frac{1}{\alpha_n\beta_{k_n,n}}\right)} \\
\geq& (1-c)\epsilon_n\sqrt{\log\left(\frac{1}{|I_n|}\right)} - 4\sqrt{\log\left(\frac{1}{\alpha_n\beta_{k_n,n}}\right)} \to \infty,
\end{aligned}$$

since the conditions of Lemma 3.1 are satisfied, as shown above.
Moreover, we have $\Gamma_{J_n} := \sqrt{n|J_n|}\theta_n - \sqrt{2q_{l_n}} = \sqrt{|J_n|\gamma_n} - \sqrt{2q_{l_n}} \to \infty$ if

$$\sqrt{\frac{|J_n|\gamma_n}{2q_{l_n}}} \geq \sqrt{\frac{|J_n|\gamma_n}{16\log\left(\frac{1}{|J_n|\alpha_n\beta_{l_n,n}}\right)}} \to \infty,$$

where we used Lemma 3.1 again. Finally, it follows from the assumptions of the theorem that $\liminf_{n\to\infty}|J_n| \geq \liminf_{n\to\infty}|I_n^C|/8 > 0$ and thus

$$\begin{aligned}
\sqrt{\frac{|J_n|\gamma_n}{\log\left(\frac{1}{\alpha_n\beta_{l_n,n}}\right)}} =& \frac{\sqrt{|I_n|\gamma_n}}{s_n\sqrt{\log\left(\frac{1}{\alpha_n\beta_{k_n,n}}\right)}}\frac{s_n\sqrt{|J_n|}}{\sqrt{|I_n|}}\frac{\sqrt{\log\left(\frac{1}{\alpha_n\beta_{k_n,n}}\right)}}{\sqrt{\log\left(\frac{1}{\alpha_n\beta_{l_n,n}}\right)}} \\
\geq& \frac{c\epsilon_n\sqrt{\log\left(\frac{1}{|I_n|}\right)}}{\sqrt{\log\left(\frac{1}{\alpha_n\beta_{k_n,n}}\right)}}\frac{s_n\sqrt{|I_n^C|/8}}{\sqrt{|I_n|}}\frac{\sqrt{\log\left(\frac{1}{\alpha_n\beta_{k_n,n}}\right)}}{\sqrt{\log\left(\frac{1}{\alpha_n\beta_{\min,n}}\right)}} \to \infty.
\end{aligned}$$

□

Proof of Theorem 3.10. It follows from Theorem 3.5 that

$$\mathbb{P}_{(\mu_n,\sigma_n^2)}\left(\hat{K}_n < K_n\right) \leq 1 - \left[1 - 3\exp\left(-\frac{1}{48}\left(\Gamma_n\right)_+^2\right)\right]_+^{2K_n} \leq 6K_n\exp\left(-\frac{1}{48}\left(\Gamma_n\right)_+^2\right),$$

with

$$\Gamma_n := \sqrt{\frac{n\lambda_n\Delta_n^2}{32}} - \sqrt{16\log\left(\frac{8}{\lambda_n\alpha_n\beta_{k_n,n}}\right)},$$

since the assumptions of Theorem 3.5 are satisfied by (3.11).
In case (1) it is enough to show $\Gamma_n \to \infty$, because K_n is bounded. Finally, $\Gamma_n \to \infty$ follows from

$$\frac{n\lambda_n\Delta_n^2}{\log\left(\frac{8}{\lambda_n\alpha_n\beta_{k_n,n}}\right)} \to \infty.$$

In case (2) for bounded K_n, $\Gamma_n \to \infty$ follows from

$$\begin{aligned}
\Gamma_n =& \sqrt{\frac{n\lambda_n\Delta_n^2}{32}} - \sqrt{16\log\left(\frac{8}{\lambda_n\alpha_n\beta_{k_n,n}}\right)} \\
\geq& \left(\frac{\sqrt{512}}{\sqrt{32}} + \frac{\epsilon_n}{\sqrt{32}}\right)\sqrt{\log\left(\frac{1}{\lambda_n}\right)} - \sqrt{16}\sqrt{\log\left(\frac{1}{\lambda_n}\right)} - \sqrt{16}\sqrt{\log\left(\frac{8}{\alpha_n\beta_{k_n,n}}\right)} \\
=& \frac{1}{\sqrt{32}}\left(\epsilon_n\sqrt{\log\left(\frac{1}{\lambda_n}\right)} - \sqrt{512}\sqrt{\log\left(\frac{8}{\alpha_n\beta_{k_n,n}}\right)}\right) \to \infty.
\end{aligned}$$

For unbounded K_n we have $K_n \leq 1/\lambda_n$. It follows

$$\begin{aligned}
& K_n \exp\left(-\frac{1}{48}\left(\Gamma_n\right)_+^2\right) \\
\leq& \exp\left(\log\left(\frac{1}{\lambda_n}\right) - \frac{1}{48}\left(\frac{C}{\sqrt{32}}\sqrt{\log\left(\frac{1}{\lambda_n}\right)} + \frac{1}{\sqrt{32}}\epsilon_n\sqrt{\log\left(\frac{1}{\lambda_n}\right)} - \sqrt{16}\sqrt{\log\left(\frac{8}{\alpha_n\beta_{k_n,n}}\right)}\right)_+^2\right) \\
\leq& \exp\left(-\frac{1}{48}\left(\frac{1}{\sqrt{32}}\epsilon_n\sqrt{\log\left(\frac{1}{\lambda_n}\right)} - \sqrt{16}\sqrt{\log\left(\frac{8}{\alpha_n\beta_{k_n,n}}\right)}\right)^2\right) \to 0.
\end{aligned}$$

□

C.5. **Proofs of Section A.**

Proof of Theorem A.1. We prove the assertion with (van der Vaart, 2007, Theorem 5.9) which states three conditions for the convergence of a Z-estimator. Note that the convergence in probability can be replaced by almost sure convergence, if the assumptions hold almost surely. We define

$$\Psi\left(\theta\right) := \left|F\left(\theta\right) - (1-\alpha)\right| + \sum_{k=2}^{d_n}\left|\frac{1-F_1\left(\theta_1\right)}{\beta_1} - \frac{1-F_k\left(\theta_k\right)}{\beta_k}\right|$$

and

$$\Psi_M\left(\theta\right) := \left|F_M\left(\theta\right) - (1-\alpha)\right| + \sum_{k=2}^{d_n}\left|\frac{1-F_{M,1}\left(\theta_1\right)}{\beta_1} - \frac{1-F_{M,k}\left(\theta_k\right)}{\beta_k}\right|$$

as well as $\Theta := [0,\infty)^{d_n}$, $\theta_0 := \mathbf{q}$ and $\hat{\theta}_M := \widehat{\mathbf{q}}_M$. Now, (A.2) and (A.3) yield

$$\Psi_M\left(\widehat{\mathbf{q}}_M\right) \leq \frac{1}{M}\left(1 + \frac{d_n - 1}{\min\{\beta_1,\ldots,\beta_{d_n}\}}\right) = \mathrm{o}(1)$$

almost surely. In addition, Lemma 2.1 shows that the vector of critical values $\mathbf{q}$ is unique. Moreover, $\sup_{\theta\in[0,\infty)^{d_n}}\|F_M(\theta) - F(\theta)\|$ and $\sup_{\theta_k\geq 0}\|F_{M,k}(\theta_k) - F_k(\theta_k)\|$ for all $k \in \{1,\ldots,d_n\}$ converge to zero almost surely. Thus, all assumptions of (van der Vaart, 2007, Theorem 5.9) are satisfied and the assertion follows. □

Proof of Lemma A.2. The computation time for the bounds $\underline{b}_{i,j}$ and $\overline{b}_{i,j}$ is $\mathcal{O}(1)$ for every fixed interval $[i/n, j/n] \in \mathcal{D}$, since they depend only on the sums $\sum_{l=i}^{j} Y_l$ as well as $\sum_{l=i}^{j} Y_l^2$ and these can be obtained from (precomputed) cumulative sums. The computation time for the intersected bounds $\underline{B}_{i,j}$ and $\overline{B}_{i,j}$ are also $\mathcal{O}(1)$ for a fixed interval $[i/n, j/n]$, since they can be computed iteratively. Therefore, the total time to compute the bounds is $\mathcal{O}(n)$, since the dyadic partition contains less than n intervals.

It follows from its iterative definition that the left limits $L_1, \ldots, L_{\hat{K}}$ can be computed in $\mathcal{O}(n)$. Therefore, the dynamic programming algorithm has cost $\mathcal{O}(\sum_{k=1}^{\hat{K}-1}(R_k - L_k + 1)$ $(R_{k+1} - L_{k+1} + 1))$ besides some linear costs, since for each point in the interval $[L_{k+1}, R_{k+1}]$ the optimal change-point in the interval $[L_k, R_k]$ has to be determined by computing the cost functional for each of these points. But, for a single interval the computation time for the restricted maximum likelihood estimator and for the cost functional is $\mathcal{O}(1)$ if the constraints $\underline{B}_{i,j}$ and $\overline{B}_{i,j}$ are given, since the restricted maximum likelihood estimator and the cost functional depend besides these constraints again only on the sums $\sum_{l=i}^{j} Y_l$ and $\sum_{l=i}^{j} Y_l^2$. This proves the assertion. $\square$

References

K. Frick, A. Munk, and H. Sieling. Multiscale change-point inference. With discussion and rejoinder by the authors. *J. Roy. Statist. Soc. Ser. B*, 76(3):495–580, 2014.

Y. Fujikoshi and S. Mukaihata. Approximations for the quantiles of Student's t and F distributions and their error bounds. *Hiroshima Math. J.*, 23(3):557–564, 1993.

N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions*, volume 1. Wiley, 2. edition, 1994.

R. Killick, P. Fearnhead, and I. A. Eckley. Optimal detection of changepoints with a linear computational cost. *J. Amer. Statist. Assoc.*, 107(500):1590–1598, 2012.

R. Maidstone and B. Pickering. Discussion of "Multiscale change-point inference" by K. Frick, A. Munk and H. Sieling. *J. Roy. Statist. Soc. Ser. B*, 76(3):560–561, 2014.

H. Sieling. *Statistical Multiscale Segmentation: Inference, Algorithms and Applications.* PhD thesis, Georg-August University Göttingen, 2013.

V. Spokoiny and M. Zhilova. Sharp deviation bounds for quadratic forms. *Math. Methods Statist.*, 22(2):100–113, 2013.

A. W. van der Vaart. *Asymptotic Statistics.* Cambridge Series in Statistical and Probabilistic Mathematics, 8th printing edition, 2007.

E. S. Venkatraman, A. B. Olshen, R. Lucito, and M. Wigler. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics*, 5(4):557–572, 2004.