Online Supplement to: A likelihood ratio approach to sequential change point detection for a general class of parameters

Holger Dette, Josua Gösmann

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum

Germany

May 23, 2019

A   Technical details

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

$$m^{-3/2}\tilde{\mathbb{U}}(\lfloor mr \rfloor, \lfloor ms \rfloor, \lfloor mt \rfloor) = m^{-3/2}(\lfloor mt \rfloor - \lfloor ms \rfloor)(\lfloor ms \rfloor - \lfloor mr \rfloor)\left(\hat{\theta}_{\lfloor mr+1 \rfloor}^{\lfloor ms \rfloor} - \hat{\theta}_{\lfloor ms \rfloor +1}^{\lfloor mt \rfloor}\right)$$

$$= \frac{\lfloor mt \rfloor - \lfloor ms \rfloor}{m^{3/2}} \sum_{i=\lfloor mr \rfloor +1}^{\lfloor ms \rfloor} \mathcal{IF}(X_i, F, \theta) - \frac{\lfloor ms \rfloor - \lfloor mr \rfloor}{m^{3/2}} \sum_{t=\lfloor ms \rfloor +1}^{\lfloor mt \rfloor} \mathcal{IF}(X_i, F, \theta)$$

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

By Assumption 3.1 we have

$$
\left\{ \frac{\lfloor mt \rfloor - \lfloor ms \rfloor}{m^{3/2}} \sum_{i=\lfloor mr \rfloor+1}^{\lfloor ms \rfloor} \mathcal{IF}(X_i, F, \theta) - \frac{\lfloor ms \rfloor - \lfloor mr \rfloor}{m^{3/2}} \sum_{i=\lfloor ms \rfloor+1}^{\lfloor mt \rfloor} \mathcal{IF}(X_i, F, \theta) \right\}_{(r,s,t)\in\Delta_3}
$$

$$
\overset{\mathcal{D}}{\Longrightarrow} \Sigma_F^{1/2} \left\{ (t-s)\big(W(s) - W(r)\big) - (s-r)\big(W(t) - W(s)\big) \right\}_{(r,s,t)\in\Delta_3}
$$

$$
= \Sigma_F^{1/2} \left\{ B(s,t) + B(r,s) - B(r,t) \right\}_{(r,s,t)\in\Delta_3} \ ,
$$

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

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

Finally, Assumption 3.2 yields

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

uniformly with respect to $(r,s,t) \in \Delta_3$ so that the proof of Theorem 3.3 is finished by Slutsky's Theorem. □

Proof of Corollary 3.5. Define

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

Using the fact, that the detection scheme $\{D_m(\lfloor mt \rfloor)\}_{t\in[0,T]}$ is piecewise constant (with respect to t) and the monotonicity of the threshold function we obtain the representation

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

By Remark 3.4 and the continuous mapping theorem we have

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

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

$$
\left| \max_{k=1}^{Tm} \frac{D_m(k)}{w(k/m)} - \max_{k=1}^{Tm} \frac{\hat{D}_m(k)}{w(k/m)} \right| \leq \max_{k=1}^{Tm} \left| D_m(k) - \hat{D}_m(k) \right|
$$

$$
\leq \|\hat{\Sigma}_m^{-1} - \Sigma_F^{-1}\|_{op} \sup_{t \in [1, T+1]} \sup_{s \in [1, t]} |m^{-3/2} \mathbb{U}(\lfloor ms \rfloor, \lfloor mt \rfloor)|^2
$$

$$
= o_{\mathbb{P}}(1) \ .
$$

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

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

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

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

$$
m^{-3/2} \mathbb{U}(\lfloor mc \rfloor, m(T+1)) = \frac{\lfloor mc \rfloor (m(T+1) - \lfloor mc \rfloor)}{m^{3/2}} \left(\hat{\theta}_1^{\lfloor mc \rfloor} - \hat{\theta}_{\lfloor mc \rfloor+1}^{m(T+1)} \right)
$$

$$
= \frac{m(T+1) - \lfloor mc \rfloor}{m^{3/2}} \sum_{i=1}^{\lfloor mc \rfloor} \mathcal{IF}(X_i, F^{(1)}, \theta_{F^{(1)}})
$$

$$
- \frac{\lfloor mc \rfloor}{m^{3/2}} \sum_{i=\lfloor mc \rfloor+1}^{m(T+1)} \mathcal{IF}(X_i, F^{(2)}, \theta_{F^{(2)}})
$$

$$
+ \frac{\lfloor mc \rfloor (m(T+1) - \lfloor mc \rfloor)}{m^{3/2}} \left(\theta_{F^{(1)}} - \theta_{F^{(2)}} + R_{1, \lfloor mc \rfloor}^{(F^{(1)})} - R_{\lfloor mc \rfloor+1, m(T+1)}^{(F^{(2)})} \right) \ ,
$$

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

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

3

and

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

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

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

$$
\begin{aligned}
m^{-4}\mathbb{V}(\lfloor ms \rfloor, \lfloor mt \rfloor) &= m^{-4} \sum_{j=1}^{\lfloor ms \rfloor} j^2 (\lfloor ms \rfloor - j)^2 \left(\hat{\theta}_1^j - \hat{\theta}_{j+1}^{\lfloor ms \rfloor} \right) \left(\hat{\theta}_1^j - \hat{\theta}_{j+1}^{\lfloor ms \rfloor} \right)^\top \\
&\quad + m^{-4} \sum_{j=\lfloor ms \rfloor+1}^{\lfloor mt \rfloor} (\lfloor mt \rfloor - j)^2 (j - \lfloor ms \rfloor)^2 \left(\hat{\theta}_{\lfloor ms \rfloor+1}^j - \hat{\theta}_{j+1}^{\lfloor mt \rfloor} \right) \left(\hat{\theta}_{\lfloor ms \rfloor+1}^j - \hat{\theta}_{j+1}^{\lfloor mt \rfloor} \right)^\top \\
&= m^{-4} \sum_{j=1}^{\lfloor ms \rfloor} \mathbb{U}(j, \lfloor ms \rfloor) \mathbb{U}^\top(j, \lfloor ms \rfloor) \\
&\quad + m^{-4} \sum_{j=\lfloor ms \rfloor+1}^{\lfloor mt \rfloor} \tilde{\mathbb{U}}(\lfloor ms \rfloor, j, \lfloor mt \rfloor) \tilde{\mathbb{U}}^\top(\lfloor ms \rfloor, j, \lfloor mt \rfloor) \\
&= m^{-3} \int_0^s \mathbb{U}(\lfloor mr \rfloor, \lfloor ms \rfloor) \mathbb{U}^\top(\lfloor mr \rfloor, \lfloor ms \rfloor) dr \\
&\quad + m^{-3} \int_s^t \tilde{\mathbb{U}}(\lfloor ms \rfloor, \lfloor mr \rfloor, \lfloor mt \rfloor) \tilde{\mathbb{U}}^\top(\lfloor ms \rfloor, \lfloor mr \rfloor, \lfloor mt \rfloor) dr \ .
\end{aligned}
$$

By Theorem 3.3 we have

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

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

4

Continuous Mapping Theorem yields in the space $\ell^\infty(\Delta_2 \cap [1, T+1]^2, \mathbb{R}^p \times \mathbb{R}^{p \times p})$

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

(A.4)

where N_1, N_2 are defined in (3.22). Now the assertion of Theorem 3.10 follows by a further application of the Continuous Mapping Theorem using that $\Sigma_F^{1/2}(N_1(s) + N_2(s,t))\Sigma_F^{1/2}$ is positive definite for $s > 0$ with probability one. $\qquad\square$

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

$$\max_{k=0}^{Tm} \frac{\hat{D}_m^{\mathrm{SN}}(k)}{w(k/m)} \geq m \cdot \frac{\left| \mathbb{U}^\top(\lfloor mc \rfloor, m(T+1)) \mathbb{V}^{-1}(\lfloor mc \rfloor, m(T+1)) \mathbb{U}(\lfloor mc \rfloor, m(T+1)) \right|}{w(T)} ,$$

(A.5)

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

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

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

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

(A.6)

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

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

where we use the extra definition

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

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

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

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

Proof of Proposition 4.2. For the sake of readability, we will give the proof only for the case $d = 2$. The arguments presented here can be easily extended to higher dimension. In view of the representation in (4.12), we may also assume without loss of generality that $\mu = \mathbb{E}[X_t] = 0$.

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

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

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

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

where g_i denotes the i-th component of the function g in (4.1). In view of definition (4.2) introduce the notation

$$X_t' = g(\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_1, \varepsilon_0', \varepsilon_{-1}, \dots) .$$

The corresponding physical dependence coefficients $\delta_{t,2}^Z$ in (4.2) are then given by

$$\delta_{t,2}^Z = \left\| \sqrt{(X_{t,1}^2 - (X_{t,1}')^2)^2 + (X_{t,2}^2 - (X_{t,2}')^2)^2 + (X_{t,1}X_{t,2} - X_{t,1}'X_{t,2}')^2} \right\|_2$$

$$\leq \|X_{t,1}^2 - (X_{t,1}')^2\|_2 + \|X_{t,2}^2 - (X_{t,2}')^2\|_2 + \|X_{t,1}X_{t,2} - X_{t,1}'X_{t,2}'\|_2$$

$$\leq 3 \cdot \max\left\{ \|X_{t,1}^2 - (X_{t,1}')^2\|_2 , \|X_{t,2}^2 - (X_{t,2}')^2\|_2 , \|X_{t,1}X_{t,2} - X_{t,1}'X_{t,2}'\|_2 \right\} ,$$

where we used the inequality $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for $a, b > 0$. Now Hölder's inequality yields for an appropriate constant C

$$\|X_{t,1}^2 - (X_{t,1}')^2\|_2 \leq \|X_{t,1} + X_{t,1}'\|_4 \|X_{t,1} - X_{t,1}'\|_4 \leq C \cdot \delta_{t,4} ,$$

$$\|X_{t,1}X_{t,2} - X_{t,1}'X_{t,2}'\|_2 \leq \left\| X_{t,1}(X_{t,2} - X_{t,2}') \right\|_2 + \left\| X_{t,2}'(X_{t,1} - X_{t,1}') \right\|_2$$

$$\leq \left\| X_{t,1} \right\|_4 \left\| X_{t,2} - X_{t,2}' \right\|_4 + \left\| X_{t,2}' \right\|_4 \left\| X_{t,1} - X_{t,1}' \right\|_4 \leq C \cdot \delta_{t,4}^{(1)} .$$

Combining these results gives $\sum_{t=1}^{\infty} \delta_{t,2}^Z \leq C \cdot \Theta_4^{(1)} < \infty$ and Theorem 3 from Wu (2005) implies the weak convergence

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

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

7

To finish part (a) it remains to show that Assumption 3.2 holds. Due to (4.14) this is a consequence of

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

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

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

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

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

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

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

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

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

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

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

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

8

and so (A.10) is shown. To establish (A.11) we will use Corollary 1 from Wu and Zhou (2011), which implies, that (on a richer probability space) there exists a process $\{\check{S}_i\}_{i=1}^n$ and a Gaussian process $\{\check{G}_i\}_{i=1}^n$, such that

$$(\check{S}_1, \ldots, \check{S}_n) \overset{\mathcal{D}}{=} (S_1, \ldots, S_n) \qquad \text{and} \qquad \max_{1 \leq i \leq n} |\check{S}_i - \check{G}_i| = \mathcal{O}_{\mathbb{P}}\big(n^{1/4}(\log n)^{3/2}\big) \,.$$

Additionally, (again on a richer probability space) there exists another Gaussian process $\{\hat{G}_i\}_{i=1}^n$ such that

$$(\check{G}_1, \ldots, \check{G}_n) \overset{\mathcal{D}}{=} (\hat{G}_1, \ldots, \hat{G}_n) \qquad \text{and} \qquad \max_{1 \leq i \leq n} |\hat{G}_i - G_i| = \mathcal{O}_{\mathbb{P}}\big(n^{1/4}(\log n)^{3/2}\big) \,,$$

where the process G is given by

$$\{G_i\}_{i=1}^n = \Big\{ \sum_{t=1}^i Y_t \Big\}_{i=1}^n \,,$$

with i.i.d. Gaussian distributed random variables $Y_1, \ldots, Y_n \sim \mathcal{N}\big(0, (\Gamma(g))_{1,1}\big)$ with $\Gamma(g)$ defined in (4.4). Therefore we obtain

$$\left| \max_{1 \leq j \leq n} \max_{\lfloor \log^2(n) \rfloor \leq k \leq n-j} \frac{\check{S}_{j+k} - \check{S}_j}{\sqrt{k} n^{1/4}} - \max_{1 \leq j \leq n} \max_{\lfloor \log^2(n) \rfloor \leq k \leq n-j} \frac{\check{G}_{j+k} - \check{G}_j}{\sqrt{k} n^{1/4}} \right|$$
$$\leq \max_{1 \leq j \leq n} \max_{\lfloor \log^2(n) \rfloor \leq k \leq n-j} \left| \frac{\check{S}_{j+k} - \check{S}_j - \check{G}_{j+k} + \check{G}_j}{\sqrt{k} n^{1/4}} \right| \leq 2 \max_{1 \leq j \leq n} \frac{|\check{S}_j - \check{G}_j|}{\log^2(n) n^{1/4}} = o_{\mathbb{P}}(1)$$

and by the same arguments

$$\left| \max_{1 \leq j \leq n} \max_{\lfloor \log^2(n) \rfloor \leq k \leq n-j} \frac{\hat{G}_{j+k} - \hat{G}_j}{\sqrt{k} n^{1/4}} - \max_{1 \leq j \leq n} \max_{\lfloor \log^2(n) \rfloor \leq k \leq n-j} \frac{G_{j+k} - G_j}{\sqrt{k} n^{1/4}} \right| = o_{\mathbb{P}}(1) \,.$$

Now Theorem 1 in Shao (1995) gives

$$\lim_{n \to \infty} \max_{1 \leq j \leq n} \max_{\lfloor \log^2(n) \rfloor \leq k \leq n-j} \frac{G_{j+k} - G_j}{\sqrt{k} n^{1/4}} = 0$$

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

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

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

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

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

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

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

Further, using (4.9) one observes

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

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

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

Similar arguments give

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

Now consider the mapping

$$
\Phi_A :
\begin{cases}
\ell^\infty([0, T+1], \mathbb{R}^3) \to \ell^\infty([0, c], \mathbb{R}^3) \times \ell^\infty([c, T+1], \mathbb{R}^3) \,, \\[2mm]
\{f(s)\}_{s \in [0,T+1]} \mapsto
\begin{pmatrix}
\{f(s)\}_{s \in [0,c]} \\
\left\{ A^{(v)}(f(s) - f(c)) \right\}_{s \in [c,T+1]}
\end{pmatrix} \,,
\end{cases}
$$

10

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

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

$$
\overset{\mathcal{D}}{=}
\begin{pmatrix}
\left\{\sqrt{\Sigma_{F^{(1)}}}W(s)\right\}_{s\in[0,c]} \\
\left\{\sqrt{\Sigma_{F^{(2)}}}\big(W(s)-W(c)\big)\right\}_{s\in[c,T+1]}
\end{pmatrix} ,
$$

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

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

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

Proof of Theorem 4.4. For a proof of Theorem 4.4 we will require six Lemmas, that are stated below. Lemmas A.1, A.2, A.3, A.5, A.8 are partially adapted from Lemma 2 and the proof of Theorem 4 in Wu (2005b) but extended to hold uniformly in sample size. Lemma A.7 controls the error of the quantile estimators in case of small samples, where the tail assumptions on the distribution function comes into play.

Lemma A.1 *Under the assumptions of Theorem 4.4 for all $0<r<1$ and $\vartheta>1$, there exists a constant $C_{r,\vartheta}$, such that*

$$
\mathbb{P}\Bigg(\max_{\substack{1\le i<j\le n \\ |j-i|\ge n^r}}\sup_{x\in\mathbb{R}}|\hat{F}_i^j(x)-F(x)|>C_{r,\vartheta}\frac{\sqrt{r\log(n)}}{n^{r/2}}\Bigg)\lesssim n^{-\vartheta}\ .
$$

11

Proof. We have the following upper bounds

$$\mathbb{P}\left(\max_{\substack{1 \le i < j \le n \\ |j-i| \ge n^r}} \sup_{x \in \mathbb{R}} |\hat{F}_i^j(x) - F(x)| > C_{r,\vartheta} \frac{\sqrt{r \log(n)}}{n^{r/2}}\right)$$

$$\le \sum_{\substack{1 \le i < j \le n \\ |j-i| \ge n^r}} \mathbb{P}\left(\sup_{x \in \mathbb{R}} |\hat{F}_i^j(x) - F(x)| > C_{r,\vartheta} \frac{\sqrt{r \log(n)}}{n^{r/2}}\right)$$

$$\le \sum_{\substack{1 \le i < j \le n \\ |j-i| \ge n^r}} \mathbb{P}\left(\sup_{x \in \mathbb{R}} |\hat{F}_i^j(x) - F(x)| > C_{r,\vartheta} \frac{\sqrt{\log(j - i + 1)}}{\sqrt{j - i + 1}}\right) .$$

Now choose $\tau > 0$ sufficiently large to fulfill $2 - \tau r < -\vartheta$. Applying Lemma 2 from Wu (2005b), we obtain that $C_{r,\vartheta}$ can be chosen, such that the last term is (up to a constant) bounded by

$$\sum_{\substack{1 \le i < j \le n \\ |j-i| \ge n^r}} |j - i + 1|^{-\tau} \le \sum_{\substack{1 \le i < j \le n \\ |j-i| \ge n^r}} n^{-r\tau} \le n^{2-r\tau} \le n^{-\vartheta} .$$

$\square$

The following inequality is a (direct) consequence of inequality 14.0.9 from Shorack and Wellner (1986).

Lemma A.2 *Under the assumptions of Theorem 4.4, let $L_F := \sup f(x) > 0$. It holds for all $0 < a \le \dfrac{1}{2L_F}$, $s > 0$, $n \in \mathbb{N}$ that*

$$\mathbb{P}\left(\sup_{|x-y| \le a} \left|\hat{F}_1^n(x) - F(x) - \left(\hat{F}_1^n(y) - F(y)\right)\right| \ge \frac{s\sqrt{L_F a}}{\sqrt{n}}\right) \le \frac{c_1}{a} \exp\left(-c_2 s^2 \psi\left(\frac{s}{\sqrt{n L_F a}}\right)\right) ,$$

where c_1 and c_2 are positive constants (only depending on F) and ψ is defined by

$$\psi(x) = 2\frac{(x + 1)\log(x + 1) - x}{x^2} \qquad \text{for } x > 0 .$$

Proof. Denote by $U_1, \ldots, U_n$ a sample of i.i.d. $\sim \mathcal{U}([0, 1])$ random variables and note that by Lipschitz continuity $|x - y| \le a$ implies $|F(x) - F(y)| \le L_F \cdot a$. Using also that F is

12

surjective and continuous by assumption, we obtain by quantile transformation

$$
\sup_{|x-y|\leq a} \left| \hat{F}_1^n(x) - F(x) - \left(\hat{F}_1^n(y) - F(y) \right) \right|
$$

$$
= \sup_{|x-y|\leq a} \frac{1}{n} \left| \sum_{i=1}^n I\{X_i \leq x\} - F(x) - I\{X_i \leq y\} + F(y) \right|
$$

$$
\overset{\mathcal{D}}{=} \sup_{|x-y|\leq a} \frac{1}{n} \left| \sum_{i=1}^n I\{F^-(U_i) \leq x\} - F(x) - I\{F^-(U_i) \leq y\} + F(y) \right|
$$

$$
= \sup_{|x-y|\leq a} \frac{1}{n} \left| \sum_{i=1}^n I\{U_i \leq F(x)\} - F(x) - I\{U_i \leq F(y)\} + F(y) \right|
$$

$$
\leq \sup_{|F(x)-F(y)|\leq L_F \cdot a} \frac{1}{n} \left| \sum_{i=1}^n I\{U_i \leq F(x)\} - F(x) - I\{U_i \leq F(y)\} + F(y) \right|
$$

$$
= \sup_{\substack{x,y\in[0,1] \\ |x-y|\leq L_F \cdot a}} \frac{1}{n} \left| \sum_{i=1}^n I\{U_i \leq x\} - x - I\{U_i \leq y\} + y \right| .
$$

The claim now follows from inequality 14.0.9 in Shorack and Wellner (1986) for the uniform empirical process. □

Lemma A.3 *Under the assumptions of Theorem 4.4 for all $0 < r < 1$ and $\vartheta > 1$, there exists a constant $C_{r,\vartheta}$, such that for all positive sequences $\{a_n\}_{n\in\mathbb{N}}$ with*

$$
\frac{r \log n}{n^r a_n} = o(1) \quad and \quad a_n = o(1) \tag{A.15}
$$

it holds that

$$
\mathbb{P}\left(\max_{\substack{1\leq i<j\leq n \\ |j-i|\geq n^r}} \sup_{|x-y|\leq a_n} |\hat{F}_i^j(x) - F(x) - (\hat{F}_i^j(y) - F(y))| > C_{r,\vartheta}\sqrt{\frac{a_n r \log(n)}{n^r}} \right) \lesssim n^{-\vartheta}
$$

provided that n is sufficiently large.

Proof. First consider $m = m_n \geq n^r$. For n sufficiently large we have $a_n \leq 1/(2L_F)$ and so

choosing $a = a_n$ and $s = C_{r,\vartheta}\sqrt{\log(m)/L_F}$ in Lemma A.2 we obtain that

$$\mathbb{P}\left(\sup_{|x-y|\leq a_n} \left| \hat{F}_1^m(x) - F(x) - \left(\hat{F}_1^m(y) - F(y) \right) \right| \geq C_{r,\vartheta} \frac{\sqrt{a_n \log(m)}}{\sqrt{m}} \right)$$

$$\leq \frac{c_1}{a_n} \exp\left(- c_2 C_{r,\vartheta}^2 \log(m) \psi\left(\frac{C_{r,\vartheta}}{L_F} \sqrt{\frac{\log(m)}{ma_n}} \right) \right) .$$

Using that ψ is non-increasing the last expression can be bounded by

$$\frac{c_1}{a_n} \exp\left(- c_2 \log(n^{C_{r,\vartheta}^2 r}) \psi\left(\frac{C_{r,\vartheta}}{L_F} \sqrt{\frac{r\log(n)}{n^r a_n}} \right) \right) . \tag{A.16}$$

Next note that by the assumption on a_n we obtain

$$\lim_{n\to\infty} \psi\left(\frac{C_{r,2}}{L_F} \sqrt{\frac{r\log(n)}{n^r a_n}} \right) = 1 .$$

Thus for n sufficiently large and with an adapted constant $\tilde{c}_2$ the term in (A.16) is bounded by

$$\frac{c_1}{a_n} \exp\left(- \tilde{c}_2 \log(n^{C_{r,\vartheta}^2 r}) \right) = \frac{c_1 n^{-\tilde{c}_2 C_{r,\vartheta}^2 r}}{a_n} = \frac{c_1 n^{-\tilde{c}_2 C_{r,\vartheta}^2 r + r}}{n^r a_n} \lesssim n^{-\vartheta - 2} ,$$

where we chose $C_{r,\vartheta}$ sufficiently large in the last estimate and used that $(n^r a_n)^{-1} = o(1)$ by assumption (A.15). Since the sequence $\{X_t\}_{t\in\mathbb{Z}}$ is i.i.d. we can now finish the proof

$$\mathbb{P}\left(\max_{\substack{1\leq i < j \leq n \\ |j-i|\geq n^r}} \sup_{|x-y|\leq a_n} |\hat{F}_i^j(x) - F(x) - (\hat{F}_i^j(y) - F(y))| > C_{r,\vartheta} \frac{\sqrt{a_n r \log(n)}}{n^{r/2}} \right)$$

$$\leq \sum_{\substack{1\leq i<j\leq n \\ |j-i|\geq n^r}} \mathbb{P}\left(\sup_{|x-y|\leq a_n} |\hat{F}_i^j(x) - F(x) - (\hat{F}_i^j(y) - F(y))| > C_{r,\vartheta} \frac{\sqrt{a_n r \log(n)}}{n^{r/2}} \right)$$

$$\leq \sum_{\substack{1\leq i<j\leq n \\ |j-i|\geq n^r}} \mathbb{P}\left(\sup_{|x-y|\leq a_n} |\hat{F}_i^j(x) - F(x) - (\hat{F}_i^j(y) - F(y))| > C_{r,\vartheta} \frac{\sqrt{a_n \log(j-i+1)}}{\sqrt{j-i+1}} \right)$$

$$\lesssim \sum_{\substack{1\leq i<j\leq n \\ |j-i|\geq n^r}} n^{-\vartheta-2} \leq n^{-\vartheta} .$$

$\square$

14

Remark A.4 For the remainder of the chapter we can choose fixed $\vartheta > 1$ and denote by $C_{r,1}$ and $C_{r,2}$ the corresponding constants from Lemma A.1 and A.3, respectively. Further define the sequence

$$b_{n,r} = C_{r,3}\sqrt{r\log(n)}/n^{r/2} \; , \tag{A.17}$$

where $C_{r,3}$ is a constant such that $C_{r,3} > 2(C_{r,1} + 1)/f(q_\beta)$. Now let 'i.o.' be a shortcut for 'infinitely often' and note that Lemma A.1 and A.3 together with the Borel-Cantelli Lemma imply

$$\mathbb{P}\left(\max_{\substack{1\leq i<j\leq n \\ |j-i|\geq n^r}} \sup_{x\in\mathbb{R}} |\hat{F}_i^j(x) - F(x)| > C_{r,1}\frac{\sqrt{r\log(n)}}{n^{r/2}} \;\; \text{i.o.}\right) = 0$$

and

$$\mathbb{P}\left(\max_{\substack{1\leq i<j\leq n \\ |j-i|\geq n^r}} \sup_{|x-y|\leq b_{n,r}} |\hat{F}_i^j(x) - F(x) - \left(\hat{F}_i^j(y) - F(y)\right)| > C_{r,2}\frac{\sqrt{b_{n,r}r\log(n)}}{n^{r/2}} \;\; \text{i.o.}\right) = 0 \; ,$$

which we require for the proof of the next Lemma.

Lemma A.5 *Under the assumptions of Theorem 4.4 it holds for all $r \in (0,1)$ that*

$$\limsup_{n\to\infty} b_{n,r}^{-1} \max_{\substack{1\leq i<j\leq n \\ |j-i|\geq n^r}} |[\hat{q}_\beta]_i^j - q_\beta| \leq 1$$

with probability one.

Proof. The claim is equivalent to

$$\mathbb{P}\left(\max_{\substack{1\leq i<j\leq n \\ |j-i|\geq n^r}} |[\hat{q}_\beta]_i^j - q_\beta| > b_{n,r} \;\; \text{i.o.}\right) = 0 \; . \tag{A.18}$$

By definition of the empirical quantile $\max_{\substack{1\leq i<j\leq n \\ |j-i|\geq n^r}} |[\hat{q}_\beta]_i^j - q_\beta| > b_{n,r}$ means that at least one of the considered e.d.f.s first exceeds the level β outside of the interval $[q_\beta - b_{n,r}, q_\beta + b_{n,r}]$. Thus using also monotonicity of the e.d.f.s statement (A.18) follows if we can establish

15

(i) $\mathbb{P}\left(\min\limits_{\substack{1\leq i<j\leq n \\ |j-i|>n^r}} \hat{F}_i^j(q_\beta + b_{n,r}) - \beta < 0 \ \text{ i.o.}\right) = 0$,

(ii) $\mathbb{P}\left(\max\limits_{\substack{1\leq i<j\leq n \\ |j-i|>n^r}} \hat{F}_i^j(q_\beta - b_{n,r}) - \beta > 0 \ \text{ i.o.}\right) = 0$.

Let us start with (i). By a Taylor expansion we obtain

$$\min\limits_{\substack{1\leq i<j\leq n \\ |j-i|\geq n^r}} \hat{F}_i^j(q_\beta + b_{n,r}) - \beta$$

$$= \min\limits_{\substack{1\leq i<j\leq n \\ |j-i|\geq n^r}} \left[F(q_\beta + b_{n,r}) - \beta - \hat{F}_i^j(q_\beta) + \beta + \hat{F}_i^j(q_\beta + b_{n,r}) - F(q_\beta + b_{n,r}) + \hat{F}_i^j(q_\beta) - F(q_\beta) \right]$$

$$\geq F(q_\beta + b_{n,r}) - \beta - \max\limits_{\substack{1\leq i<j\leq n \\ |j-i|\geq n^r}} |\hat{F}_i^j(q_\beta) - \beta| - \max\limits_{\substack{1\leq i<j\leq n \\ |j-i|>n^r}} \sup\limits_{|x-y|\leq b_{n,r}} \left|\hat{F}_i^j(x) - F(x) - \left(\hat{F}_i^j(y) - F(y)\right)\right|$$

$$\geq b_{n,r} f(q_\beta) + \frac{b_{n,r}^2}{2} \inf\limits_{x\in\mathbb{R}} f'(x) - \max\limits_{\substack{1\leq i<j\leq n \\ |j-i|\geq n^r}} |\hat{F}_i^j(q_\beta) - \beta|$$

$$- \max\limits_{\substack{1\leq i<j\leq n \\ |j-i|\geq n^r}} \sup\limits_{|x-y|\leq b_{n,r}} \left|\hat{F}_i^j(x) - F(x) - \left(\hat{F}_i^j(y) - F(y)\right)\right| .$$

This yields

$$\mathbb{P}\left(\min\limits_{\substack{1\leq i<j\leq n \\ |j-i|\geq n^r}} \hat{F}_i^j(q_\beta + b_{n,r}) - \beta \leq 0 \ \text{ i.o.}\right)$$

$$\leq \mathbb{P}\left(\max\limits_{\substack{1\leq i<j\leq n \\ |j-i|\geq n^r}} \sup\limits_{|x-y|\leq b_{n,r}} \left|\hat{F}_i^j(x) - F(x) - \left(\hat{F}_i^j(y) - F(y)\right)\right| + \max\limits_{\substack{1\leq i<j\leq n \\ |j-i|\geq n^r}} |\hat{F}_i^j(q_\beta) - \beta|\right.$$

$$\left. \geq b_{n,r} f(q_\beta) + \frac{b_{n,r}^2}{2} \inf\limits_{x\in\mathbb{R}} f'(x) \ \text{ i.o.}\right)$$

$$\leq \mathbb{P}\left(\max\limits_{\substack{1\leq i<j\leq n \\ |j-i|\geq n^r}} |\hat{F}_i^j(q_\beta) - \beta| \geq \frac{b_{n,r}}{2} f(q_\beta) + \frac{b_{n,r}^2}{4} \inf\limits_{x\in\mathbb{R}} f'(x) \ \text{ i.o.}\right)$$

$$+ \mathbb{P}\left(\max\limits_{\substack{1\leq i<j\leq n \\ |j-i|\geq n^r}} \sup\limits_{|x-y|\leq b_{n,r}} \left|\hat{F}_i^j(x) - F(x) - \left(\hat{F}_i^j(y) - F(y)\right)\right| \geq \frac{b_{n,r}}{2} f(q_\beta) + \frac{b_{n,r}^2}{4} \inf\limits_{x\in\mathbb{R}} f'(x) \ \text{ i.o.}\right) .$$

By definition of $b_{n,r}$ in (A.17) we have $b_{n,r} f(q_\beta)/2 > (C_{r,1}+1)\sqrt{r \log(n)}/n^{r/2}$ and so Remark A.4 yields that the last two probabilities are zero. To achieve (ii), we proceed similar and

16

obtain

$$\max_{\substack{1 \le i < j \le n \\ |j-i| \ge n^r}} \hat{F}_i^j(q_\beta - b_{n,r}) - \beta$$

$$= \max_{\substack{1 \le i < j \le n \\ |j-i| \ge n^r}} F(q_\beta - b_{n,r}) - \beta - \hat{F}_i^j(q_\beta) + \beta + \hat{F}_i^j(q_\beta - b_{n,r}) - F(q_\beta - b_{n,r}) + \hat{F}_i^j(q_\beta) - F(q_\beta)$$

$$\le F(q_\beta - b_{n,r}) - \beta + \max_{\substack{1 \le i < j \le n \\ |j-i| \ge n^r}} |\hat{F}_i^j(q_\beta) - \beta|$$

$$+ \max_{\substack{1 \le i < j \le n \\ |j-i| \ge n^r}} \sup_{|x-y| \le b_{n,r}} |\hat{F}_i^j(x) - F(x) - \left(\hat{F}_i^j(y) - F(y) \right)|$$

$$\le -f(q_\beta) b_{n,r} + \sup_{x \in \mathbb{R}} f'(x) \frac{b_{n,r}^2}{2} + \max_{\substack{1 \le i < j \le n \\ |j-i| > n^r}} |\hat{F}_i^j(q_\beta) - \beta|$$

$$+ \max_{\substack{1 \le i < j \le n \\ |j-i| \ge n^r}} \sup_{|x-y| \le b_{n,r}} |\hat{F}_i^j(x) - F(x) - \left(\hat{F}_i^j(y) - F(y) \right)|$$

This leads to

$$\mathbb{P}\left(\max_{\substack{1 \le i < j \le n \\ |j-i| \ge n^r}} \hat{F}_i^j(q_\beta - b_{n,r}) - \beta \ge 0 \ \text{i.o.} \right)$$

$$\le \mathbb{P}\left(-f(q_\beta) b_{n,r} + \sup_{x \in \mathbb{R}} f'(x) \frac{b_{n,r}^2}{2} + \max_{\substack{1 \le i < j \le n \\ |j-i| \ge n^r}} |\hat{F}_i^j(q_\beta) - \beta| \right.$$

$$\left. + \max_{\substack{1 \le i < j \le n \\ |j-i| \ge n^r}} \sup_{|x-y| \le b_{n,r}} |\hat{F}_i^j(x) - F(x) - \left(\hat{F}_i^j(y) - F(y) \right)| \ge 0 \ \text{i.o.} \right)$$

$$\le \mathbb{P}\left(\max_{\substack{1 \le i < j \le n \\ |j-i| \ge n^r}} |\hat{F}_i^j(q_\beta) - \beta| \ge f(q_\beta) b_{n,r} - \sup_{x \in \mathbb{R}} f'(x) \frac{b_{n,r}^2}{2} \ \text{i.o.} \right)$$

$$+ \mathbb{P}\left(\max_{\substack{1 \le i < j \le n \\ |j-i| \ge n^r}} \sup_{|x-y| \le b_{n,r}} |\hat{F}_i^j(x) - F(x) - \left(\hat{F}_i^j(y) - F(y) \right)| \ge f(q_\beta) b_{n,r} - \sup_{x \in \mathbb{R}} f'(x) \frac{b_{n,r}^2}{2} \ \text{i.o.} \right).$$

Using again the definition of $b_{n,r}$ and Remark A.4 the two probabilities are zero, which finishes the proof of Lemma A.5. $\qquad\square$

17

Remark A.6 Note that Lemma A.5 in particular implies that (for all $0 < r < 1$) and $c_0 < 1$

$$n^{c_0 r/2} \max_{\substack{1 \leq i < j \leq n \\ |j-i| \geq n^r}} \left| [\hat{q}_\beta]_i^j - q_\beta \right| = o_{\mathbb{P}}(1) ,$$

which we require later on.

Lemma A.7 *Under the assumptions of Theorem 4.4 it holds for all $0 < r < 1/2 - 1/\lambda$*

$$\frac{1}{\sqrt{n}} \max_{\substack{1 \leq i < j \leq n \\ |j-i| < n^r}} (j - i + 1) \left| [\hat{q}_\beta]_i^j - q_\beta \right| = o_{\mathbb{P}}(1) . \tag{A.19}$$

Proof. Due to $\min_{t=1}^n X_t \leq [\hat{q}_\beta]_i^j \leq \max_{t=1}^n X_t$ for all $i, j \in \{1, \ldots, n\}$, we observe that the term on the left-hand side of (A.19) is bounded by

$$n^{r-1/2} \left(\left| \max_{t=1}^n X_t - q_\beta \right| + \left| \min_{t=1}^n X_t - q_\beta \right| \right) \leq 2n^{r-1/2} \max_{t=1}^n |X_t| + o(1) .$$

Now for $\varepsilon > 0$ we can employ the independence and obtain

$$\mathbb{P}\left(n^{r-1/2} \max_{i=1}^n |X_t| > \varepsilon \right) = 1 - \mathbb{P}\left(n^{r-1/2} \max_{i=1}^n |X_t| \leq \varepsilon \right) = 1 - \mathbb{P}\left(|X_1| \leq n^{1/2-r}\varepsilon \right)^n$$

$$= 1 - \left(1 - \mathbb{P}\left(|X_1| > n^{1/2-r}\varepsilon \right) \right)^n .$$

By the assumption (4.16) on the tails of the distribution of $|X_1|$ this is now bounded by

$$1 - \left(1 - \varepsilon^{-\lambda} n^{-\lambda(1/2-r)} \right)^n = o(1) ,$$

where we used that by assumption $\lambda(1/2 - r) > 1$. $\qquad\square$

Lemma A.8 *Under the assumptions of Theorem 4.4 it holds for all $2/9 < r < 1$*

$$\frac{1}{\sqrt{n}} \max_{\substack{1 \leq i < j \leq n \\ |j-i| > n^r}} (j - i + 1) \left| \hat{F}_i^j([\hat{q}_\beta]_i^j) - \hat{F}_i^j(q_\beta) - (F([\hat{q}_\beta]_i^j) - F(q_\beta)) \right| = o_p(1) .$$

Proof. Fix $\varepsilon > 0$ and choose δ such that $2/3 < \delta < \frac{3}{4}r + 1/2$, then it holds that

$$\mathbb{P}\left(\frac{1}{\sqrt{n}} \max_{\substack{1 \le i < j \le n \\ |j-i| \ge n^r}} (j - i + 1)\left|\hat{F}_i^j([\hat{q}_\beta]_i^j) - \hat{F}_i^j(q_\beta) - \left(F([\hat{q}_\beta]_i^j) - F(q_\beta)\right)\right| > \varepsilon\right)$$

$$\le \mathbb{P}\left(\frac{1}{\sqrt{n}} \max_{\substack{1 \le i < j \le n \\ n^\delta > |j-i| \ge n^r}} (j - i + 1)\left|\hat{F}_i^j([\hat{q}_\beta]_i^j) - \hat{F}_i^j(q_\beta) - \left(F([\hat{q}_\beta]_i^j) - F(q_\beta)\right)\right| > \varepsilon\right) \quad \text{(A.20)}$$

$$+ \mathbb{P}\left(\frac{1}{\sqrt{n}} \max_{\substack{1 \le i < j \le n \\ |j-i| \ge n^\delta}} (j - i + 1)\left|\hat{F}_i^j([\hat{q}_\beta]_i^j) - \hat{F}_i^j(q_\beta) - \left(F([\hat{q}_\beta]_i^j) - F(q_\beta)\right)\right| > \varepsilon\right).$$

We will treat the two summands on the right-hand side separately.

First summand of (A.20): Using $\delta - 1/2 < 3/4r$, we can choose a constant $0 < c_0 < 1$ sufficiently large, such that $\delta - 1/2 < (c_0/4 + 1/2)r$. Further choose $a_n = n^{-c_0 r/2}$. The first summand of (A.20) is then bounded by

$$\mathbb{P}\left(\max_{\substack{1 \le i < j \le n \\ |j-i| \ge n^r}} n^{\delta-1/2}\left|\hat{F}_i^j([\hat{q}_\beta]_i^j) - \hat{F}_i^j(q_\beta) - \left(F([\hat{q}_\beta]_i^j) - F(q_\beta)\right)\right| > \varepsilon\right)$$

$$\le \mathbb{P}\left(\max_{\substack{1 \le i < j \le n \\ |j-i| \ge n^r}} \sup_{|x-y| \le a_{n,r}} \left|\hat{F}_i^j(x) - F(x) - \hat{F}_i^j(y) + F(y)\right| > \frac{\varepsilon}{n^{\delta-1/2}}\right) \quad \text{(A.21)}$$

$$+ \mathbb{P}\left(\max_{\substack{1 \le i < j \le n \\ |j-i| \ge n^r}} |[\hat{q}_\beta]_i^j - q_\beta| > a_{n,r}\right).$$

By Remark A.6 the second summand of the right-hand side of (A.21) converges to zero. For the first summand of (A.21) note that using $\delta - 1/2 < (c_0/4 + 1/2)r$, we obtain (provided that n is sufficiently large)

$$\frac{\varepsilon}{n^{\delta-1/2}} \ge C_{r,2} \frac{\sqrt{r \log(n)}}{n^{c_0 r/4 + r/2}} = C_{r,2} \frac{\sqrt{a_{n,r} r \log(n)}}{n^{r/2}}$$

and so the first summand of (A.21) converges to zero by Lemma A.3.

Second summand of (A.20): Due to $\delta > 2/3$, we can choose a constant $0 < c_0 < 1$ sufficiently large, such that $1/2 < \delta/2 + c_0\delta/4$. Next define $a_{n,\delta} = n^{-c_0\delta/2}$ and obtain the

19

bound

$$\mathbb{P}\left(\frac{1}{\sqrt{n}} \max_{\substack{1 \leq i < j \leq n \\ |j-i| \geq n^\delta}} (j-i+1) \left| \hat{F}_i^j([\hat{q}_\beta]_i^j) - \hat{F}_i^j(q_\beta) - \left(F([\hat{q}_\beta]_i^j) - F(q_\beta) \right) \right| > \varepsilon \right)$$

$$\leq \mathbb{P}\left(\max_{\substack{1 \leq i < j \leq n \\ |j-i| \geq n^\delta}} \sup_{|x-y| \leq a_{n,\delta}} \left| \hat{F}_i^j(x) - F(x) - \hat{F}_i^j(y) + F(y) \right| > \frac{\varepsilon}{n^{1/2}} \right)$$

$$+ \mathbb{P}\left(\max_{\substack{1 \leq i < j \leq n \\ |j-i| \geq n^\delta}} |[\hat{q}_\beta]_i^j - q_\beta| > a_{n,\delta} \right) . \tag{A.22}$$

Employing again Remark (A.6) the second summand of (A.22) converges to zero. For the first summand of (A.22), note that we have (for sufficiently large n)

$$\frac{\varepsilon}{n^{1/2}} \geq C_{\delta,2} \frac{\sqrt{\delta \log(n)}}{n^{c_0 \delta/4 + \delta/2}} = C_{\delta,2} \frac{\sqrt{a_{n,\delta} \delta \log(n)}}{n^{\delta/2}}$$

and so Lemma A.3 finishes the proof. $\qquad \square$

Now we are able to proceed to the actual proof of Theorem 4.4.

Proof of Theorem 4.4:
Since β is fixed, it is easy to see that the claim is equivalent to

$$\frac{1}{\sqrt{n}} \max_{1 \leq i < j \leq n} (j-i+1) \left| f(q_\beta)\left([\hat{q}_\beta]_i^j - q_\beta\right) - \beta + \hat{F}_i^j(q_\beta) \right| = o_\mathbb{P}(1) . \tag{A.23}$$

Further note that (since F is continuous) $|\hat{F}_i^j([\hat{q}_\beta]_i^j) - \beta| \leq (j-i+1)^{-1}$ almost surely, which yields

$$\frac{1}{\sqrt{n}} \max_{1 \leq i < j \leq n} (j-i+1) |\hat{F}_i^j([\hat{q}_\beta]_i^j) - \beta| = o_\mathbb{P}(1)$$

and so it remains to prove

$$\frac{1}{\sqrt{n}} \max_{1 \leq i < j \leq n} (j-i+1) \left| f(q_\beta)\left([\hat{q}_\beta]_i^j - q_\beta\right) - \hat{F}_i^j([\hat{q}_\beta]_i^j) + \hat{F}_i^j(q_\beta) \right| = o_\mathbb{P}(1) . \tag{A.24}$$

20

Now due to $\lambda > 18/5$, we can choose r_1 with $2/9 < r_1 < 1/2 - 1/\lambda < 1/2$. By Lemma A.7 and $\hat{F}_i^j([\hat{q}_\beta]_i^j), \hat{F}_i^j(q_\beta) \in [0,1]$ we only have to verify

$$\frac{1}{\sqrt{n}} \max_{\substack{1 \leq i < j \leq n \\ |j-i| \geq n^{r_1}}} (j - i + 1) \left| f(q_\beta)([\hat{q}_\beta]_i^j - q_\beta) - \hat{F}_i^j([\hat{q}_\beta]_i^j) + \hat{F}_i^j(q_\beta) \right| = o_\mathbb{P}(1) .$$

Now employing Lemma A.8, the statement above follows if we can establish

$$\frac{1}{\sqrt{n}} \max_{\substack{1 \leq i < j \leq n \\ |j-i| \geq n^{r_1}}} (j - i + 1) \left| f(q_\beta)([\hat{q}_\beta]_i^j - q_\beta) - F([\hat{q}_\beta]_i^j) + F(q_\beta) \right| = o_\mathbb{P}(1) . \tag{A.25}$$

By means of a Taylor expansion the term on the left-hand side is (up to a constant almost surely) bounded by

$$\frac{1}{\sqrt{n}} \max_{\substack{1 \leq i < j \leq n \\ |j-i| \geq n^{r_1}}} (j - i + 1) \sup_{x \in \mathbb{R}} |f'(x)| \left([\hat{q}_\beta]_i^j - q_\beta\right)^2 ,$$

where the factor $\sup_{x \in \mathbb{R}} |f'(x)|$ is bounded by assumption. Now since $2/9 < r_1 < 1/2$, it is easy to see, that we can choose $0 < c_0 < 1$ (sufficiently large), such that

$$\frac{1}{2c_0} \leq r_1 c_0 + 1/2 .$$

Thus we can select δ that fulfills

$$r_1 < \frac{1}{2c_0} \leq \delta \leq r_1 c_0 + 1/2 .$$

We consider the cases $n^{r_1} \leq |i - j| \leq n^\delta$ and $n^\delta \leq |i - j|$ separately. For the first one we obtain

$$\frac{1}{\sqrt{n}} \max_{\substack{1 \leq i < j \leq n \\ n^{r_1} \leq |j-i| < n^\delta}} (j - i + 1) \left| \left([\hat{q}_\beta]_i^j - q_\beta\right)^2 \leq \max_{\substack{1 \leq i < j \leq n \\ n^{r_1} \leq |j-i| < n^\delta}} n^{\delta - 1/2} \left([\hat{q}_\beta]_i^j - q_\beta\right)^2 \right.$$

$$\leq \max_{\substack{1 \leq i < j \leq n \\ n^{r_1} \leq |j-i| < n}} n^{c_0 r_1} \left([\hat{q}_\beta]_i^j - q_\beta\right)^2 = \max_{\substack{1 \leq i < j \leq n \\ n^{r_1} \leq |j-i| < n}} \left(n^{c_0 r_1/2}([\hat{q}_\beta]_i^j - q_\beta)\right)^2 = o_\mathbb{P}(1) ,$$

where we used Remark A.6 for the last estimate. For the other case we obtain similarly

$$\frac{1}{\sqrt{n}} \max_{\substack{1 \leq i < j \leq n \\ n^\delta \leq |j-i| < n}} (j - i + 1) |\big([\hat{q}_\beta]_i^j - q_\beta\big)^2 \leq \max_{\substack{1 \leq i < j \leq n \\ n^\delta \leq |j-i| < n}} n^{1/2} \big([\hat{q}_\beta]_i^j - q_\beta\big)^2$$

$$\leq \max_{\substack{1 \leq i < j \leq n \\ n^\delta \leq |j-i| < n}} n^{c_0 \delta} \big([\hat{q}_\beta]_i^j - q_\beta\big)^2 = \max_{\substack{1 \leq i < j \leq n \\ n^\delta \leq |j-i| < n}} \big(n^{c_0 \delta/2}([\hat{q}_\beta]_i^j - q_\beta)\big)^2 = o_{\mathbb{P}}(1) \ ,$$

where we again employed Remark A.6. $\qquad\qquad\square$

B Additional simulation results

In this section we provide some additional simulation results to allow a more detailed analysis of the presented detection schemes. We will focus on changes in the mean as presented in Section 5.1 and study the following aspects:

Section B.1: The influence of the actual change point locations on the power.

Section B.2: Other choices of the factor T, that controls the monitoring window length.

B.1 Influence of change point locations

In this section we report simulation results for the situation considered in Figure 2 except for the change point locations, for which we consider rather early and late locations. Figure 8 displays the power of the non self-normalized procedures for the different choices of the model and the threshold considered in Section 5.1, where the change occurs already at observation X_{120} and a historical training data ending at X_{100}. This can be considered as a situation of an early change and the displayed plots can be explained as follows. In all combinations, the detection scheme based on $\hat{D}$ still has a slightly larger power compared to the methods based on $\hat{P}$ and $\hat{Q}$, while $\hat{P}$ slightly outperforms $\hat{Q}$. Compared to Figure 2 the differences with respect to the different schemes are considerably smaller. These observations may be explained by the different constructions of the detection schemes,

22

that are described at the end of Section 2. In particular the performance of the monitoring schemes based on $\hat{Q}$ and $\hat{P}$ improves if the change occurs closer to the monitoring start, see also the discussion at the end of Section 2.

In Figure 9 we report the power for a change located close to the end of the monitoring period. Here the break occurs at observation X_{180}, while the monitoring window ends with observation X_{200}. Concerning the small number of 20 observations after the change, such an event is certainly harder to detect. Consequently, all schemes perform inferior compared to the situations considered in Figure 2 and 8. However, the power superiority of the methods based on the statistics $\hat{D}$ over $\hat{P}$ and $\hat{Q}$ is even more significant now. These results support our initial conjecture: While all schemes behave more or less equivalent for changes close to the start, $\hat{D}$ offers better characteristics, if changes are located closer to the end.
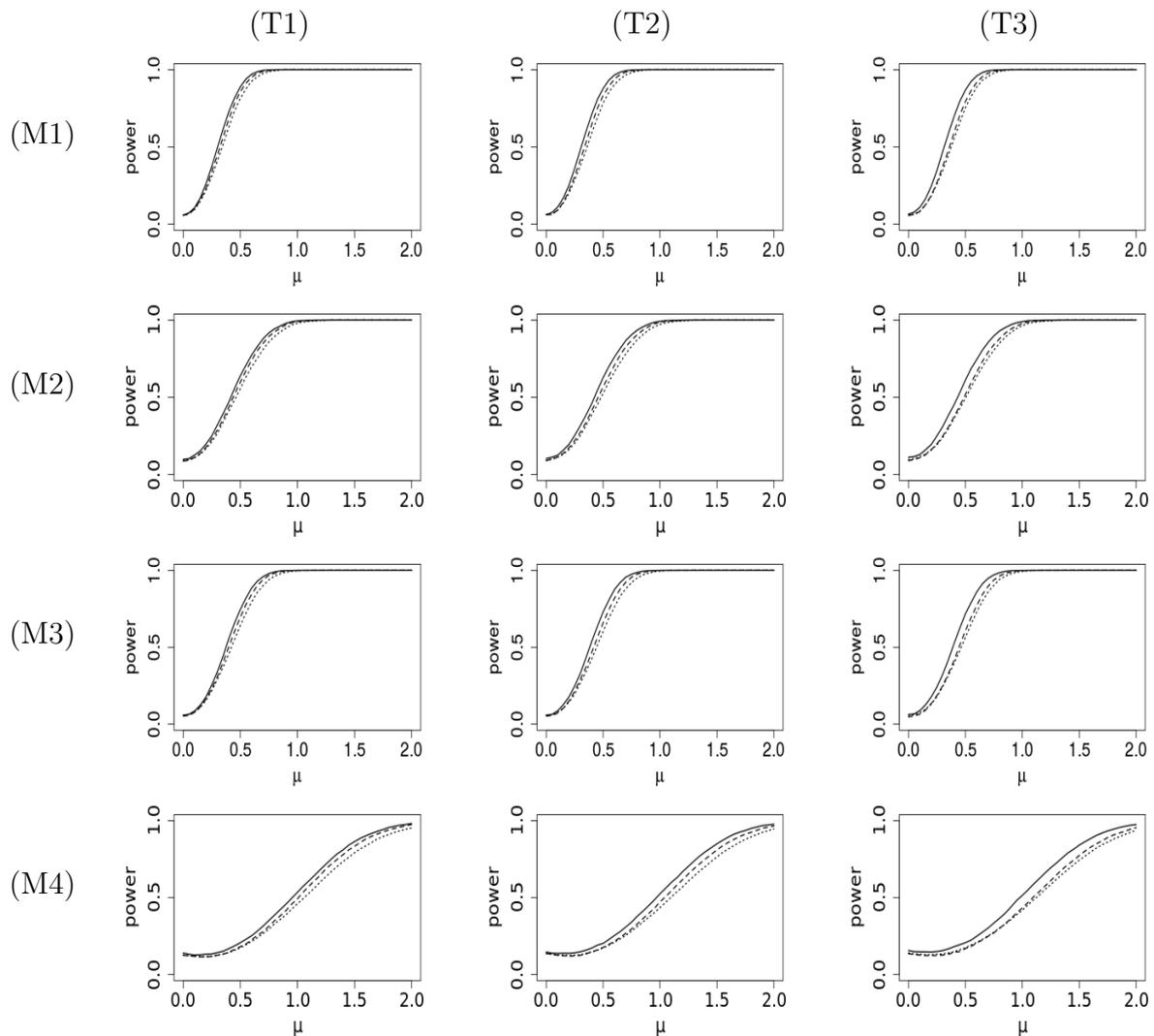
Figure 8: *Empirical rejection probabilities of the sequential tests for a change in the mean based on the statistics $\hat{D}$ (solid line), $\hat{P}$ (dashed line), $\hat{Q}$ (dotted line). The initial and total sample size are $m = 100$ and $m(T + 1) = 200$, respectively, and the change occurs at observation 120. The level is $\alpha = 0.05$. Different rows correspond to different threshold functions, while different columns correspond to different models.*
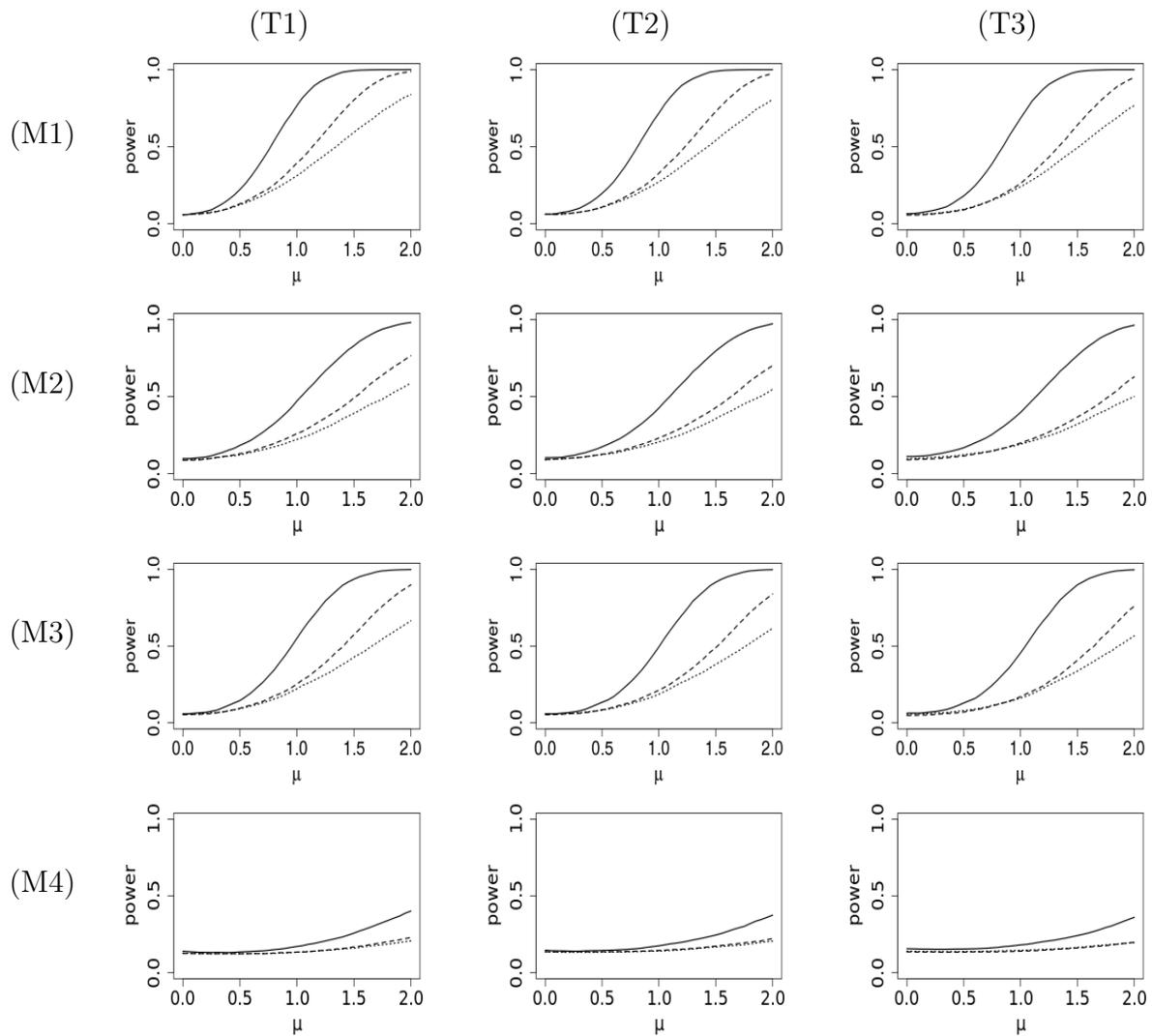
Figure 9: *Empirical rejection probabilities of the sequential tests for a change in the mean based on the statistics $\hat{D}$ (solid line), $\hat{P}$ (dashed line), $\hat{Q}$ (dotted line). The initial and total sample size are $m = 100$ and $m(T + 1) = 200$, respectively, and the change occurs at observation 180. The level is $\alpha = 0.05$. Different rows correspond to different models, while different columns correspond to different threshold functions.*

B.2   Larger monitoring windows

In this section we report simulations with the same settings as in Figure 2 but with a larger monitoring window. More precisely, we operate again with a set of $m = 100$ stable observations, while the factor T is set to 2 and 3 for the simulations in Figure 10 and 11, respectively. The change point is again located at the middle of the monitoring period. The obtained results are similar to those for the case $T = 1$ given in Section 5.1 and for this reason we omit a detailed discussion here.
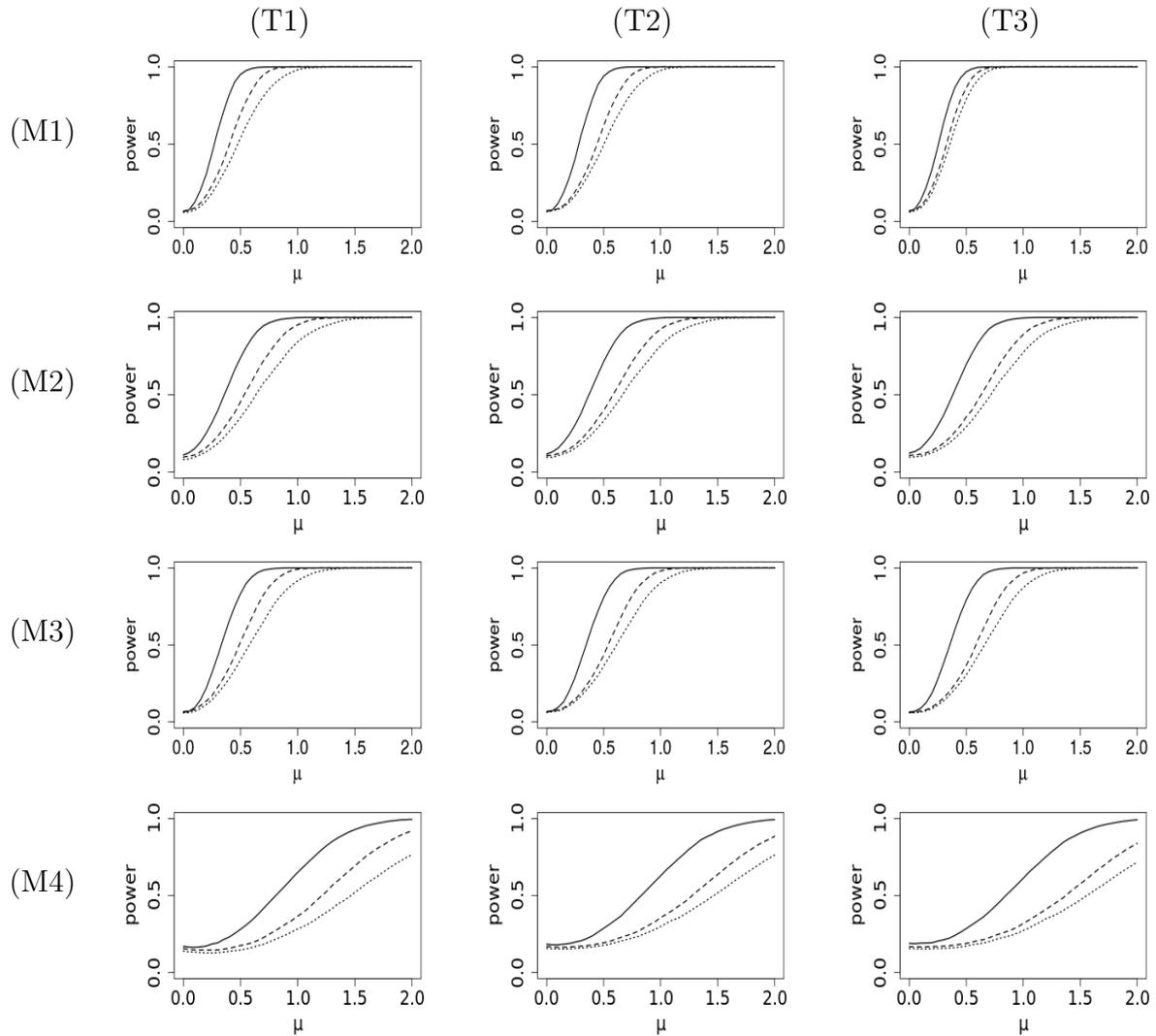
Figure 10: *Empirical rejection probabilities of the sequential tests for a change in the mean based on the statistics $\hat{D}$ (solid line), $\hat{P}$ (dashed line), $\hat{Q}$ (dotted line). The initial and total sample size are $m = 100$ and $m(T + 1) = 300$, respectively, and the change occurs at observation 200. The level is $\alpha = 0.05$. Different rows correspond to different models, while different columns correspond to different threshold functions.*
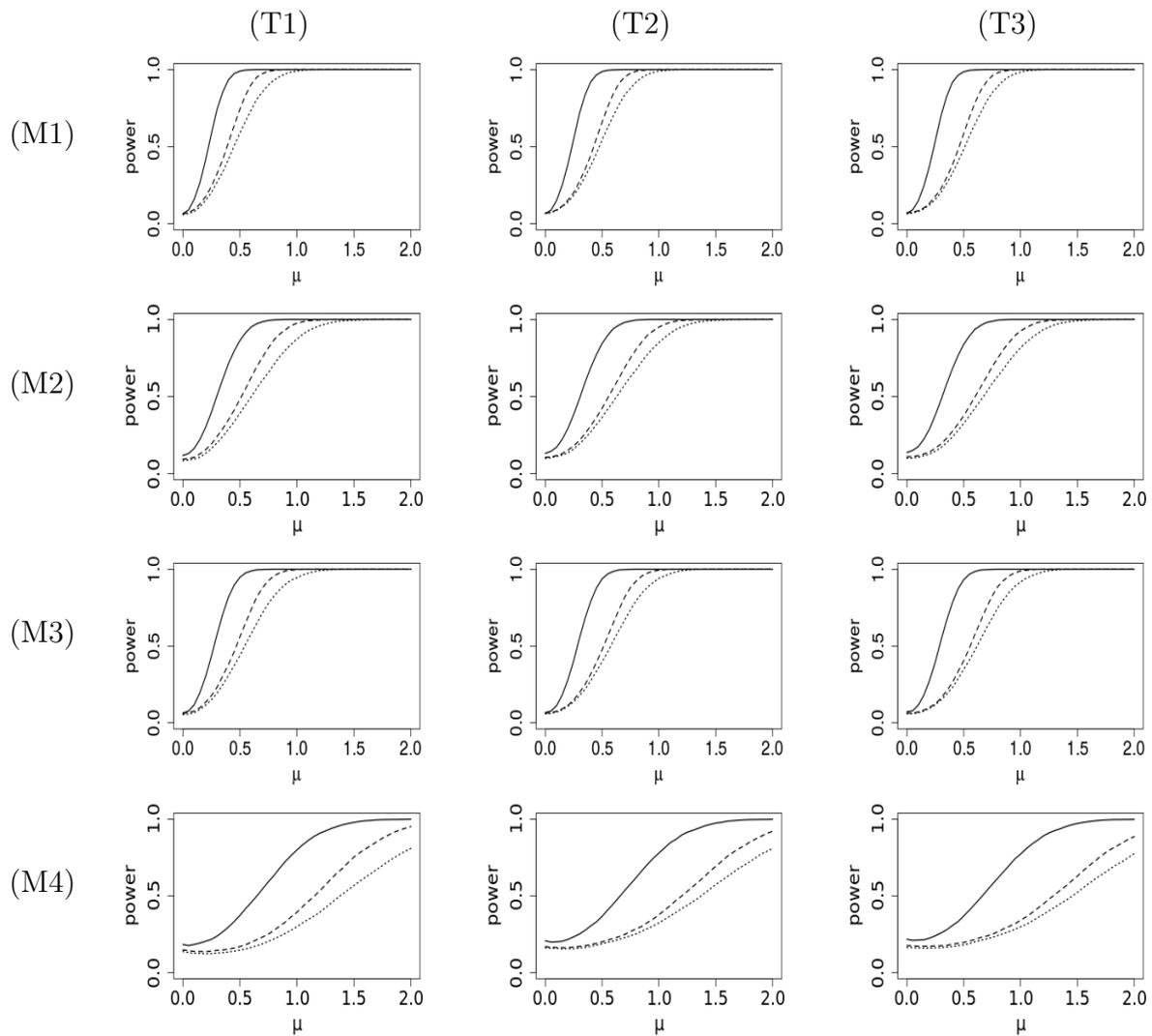
Figure 11: *Empirical rejection probabilities of the sequential tests for a change in the mean based on the statistics $\hat{D}$ (solid line), $\hat{P}$ (dashed line), $\hat{Q}$ (dotted line). The initial and total sample size are $m = 100$ and $m(T + 1) = 400$, respectively, and the change occurs at observation 250. The level is $\alpha = 0.05$. Different rows correspond to different models, while different columns correspond to different threshold functions.*

References

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

Shorack, G. and Wellner, J. (1986). *Empirical processes with applications to statistics.* Wiley series in probability and mathematical statistics: Probability and mathematical statistics. Wiley.

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

Wu, W. B. (2005b). On the Bahadur representation of sample quantiles for dependent sequences. *The Annals of Statistics*, 33(4):1934–1963.

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