Inference for Change Points in High Dimensional Mean Shift Models

Abhishek Kaul[a] and George Michailidis[b]

[a] *Washington State University,* [b] *University of Florida*

S1 Proofs of results in Section 3

To present the arguments of this section, we require some additional notation. In all to follow define $\hat{\eta}_{(j)} = \hat{\theta}_{(j)} - \hat{\theta}_{(j+1)}$, $j = 1, ..., N$. For any non-negative sequences $0 \le v_T \le u_T \le 1$ define the following collection,

$$\mathcal{G}_j(u_T, v_T) \;\; = \;\; \Big\{\tau_j \in \{1, 2, ..., (T-1)\};\; Tv_T \le |\tau_j - \tau_j^0| \le Tu_T\Big\} \qquad \text{(S1.1)}$$

Finally, for any vectors $\theta_{(j)} \in \mathbb{R}^p$ and any $(\tau_j, \tau_{-j}^T)^T \in \{1, ..., (T-1)\}^N$, define,

$$\mathcal{U}_j(\tau_j, \tau_{-j}, \theta) = Q_j(\tau_j, \tau_{-j}, \theta) - Q_j(\tau_j^0, \tau_{-j}, \theta), \quad j = 1, ..., N. \qquad \text{(S1.2)}$$

Recall that $Q_j(\cdot\,, \cdot\,, \cdot\,)$ is the squared loss defined in (2.3).

Lemma S1.1. *Suppose conditions A, B and C hold and let* $0 \le v_T \le u_T \le 1$, *be any non-negative sequences. Then, for any* $0 < a < 1$, *choosing* $c_a \ge \sqrt{(1/a)}$ *and for any given* $j = 1, ..., N$, *we have the following uniform lower bound,*

$$\inf_{\tau_j \in \mathcal{G}_j(u_T, v_T)} \mathcal{U}_j(\tau_j, \hat{\tau}_{-j}, \hat{\theta}) \;\; \ge \;\; \frac{\xi_j^2}{2}\Big[v_T - \frac{c_u c_a \sigma}{\xi_j}\Big(\frac{u_T}{T}\Big)^{\frac{1}{2}}\Big],$$

with probability at least $1-2a-o(1)-\pi_T$.

Proof of Lemma S1.1. We begin with a few observations that shall be required to obtain the desired lower bound of this lemma. Define, $\mathcal{A}=\{\text{event where Condition C holds}\}$, then we have by assumption $pr(\mathcal{A})\geq 1-\pi_T$. First, begin by noting that from Condition C(i) we have, $\max_{1\leq j\leq N}|\hat\tau_j-\tau_j^0|\leq c_{u1}T\underline{\ell}$, for a suitably small constant $c_{u1}>0$. Since by Condition B(ii), $T\underline{\ell}$ is the least separation between consecutive change points, consequently, τ_j^0 must lie between $\hat\tau_{j-1}$ and $\hat\tau_{j+1}$ on the event under consideration. Moreover, the same assumption also provides that there can only be atmost the immediate neighboring change points τ_{j-1}^0 and τ_{j+1}^0 in the interval $\hat\tau_{j-1}$ and $\hat\tau_{j+1}$ on the event $\mathcal{A}$, and no further change points can be contained in this interval on this event. Second, from Condition C(ii) we have the following relations on the event $\mathcal{A}$, for all $j=1,...,N$,

$$\begin{aligned}
\|\hat\eta_{(j)}-\eta_{(j)}^0\|_2 &\leq \|\hat\theta_{(j)}-\theta_{(j)}^0\|_2+\|\hat\theta_{(j+1)}-\theta_{(j+1)}^0\|_2\leq\frac{2c_{u1}\underline{\xi}}{(Ns)^{1/2}\log(p\vee T)}\quad\text{and,}\\
\|\hat\eta_{(j)}-\eta_{(j)}^0\|_1 &\leq 4\sqrt{Ns}\|\hat\theta_1-\theta_1^0\|_2+4\sqrt{Ns}\|\hat\theta_2-\theta_2^0\|_2\leq\frac{8c_{u1}\underline{\xi}}{\log(p\vee T)}
\end{aligned}\tag{S1.3}$$

The third inequality follows from assumption $\|(\hat\theta_{(j)})_{S_j^c}\|_1\leq 3\|(\hat\theta_{(j)}-\theta_{(j)}^0)_{S_j}\|_1$ of Condition C(iia) which implies that $\|\hat\theta_{(j)}-\theta_{(j)}^0\|_1\leq 4\sqrt{(Ns)}\|\hat\theta_{(j)}-\theta_{(j)}^0\|_2$. Next, consider

$$\begin{aligned}
\|\hat\eta_{(j)}\|_2 &\leq \|\eta_{(j)}^0\|_2+\|\hat\eta_{(j)}-\eta_{(j)}^0\|_2\leq\xi_j\big\{1+\frac{2c_{u1}}{(Ns)^{1/2}\log(p\vee T)}\big\}\leq\frac{3}{2}\xi_j,\quad\text{and,}\\
\|\hat\eta_{(j)}\|_1 &\leq \|\eta^0\|_1+\|\hat\eta-\eta^0\|_1\\
&\leq \sqrt{(Ns)}\xi_j\big\{1+\frac{2c_{u1}}{(Ns)^{1/2}\log(p\vee T)}\big\}\leq\frac{3}{2}\sqrt{(Ns)}\xi_j,
\end{aligned}\tag{S1.4}$$

The second inequality follows from (S1.3). The ℓ_1 bound follows analogously. In the following consider any $\tau_j \in \mathcal{G}_j(u_T, v_T)$, and assume the ordering $\tau_j \geq \tau^0_{j+1} > \tau^0_j$. The remaining permutations of the ordering of τ_j with respect to τ^0_{j-1}, τ^0_j and τ^0_{j+1}, possible on the set $\mathcal{A}$ are $\tau^0_{j+1} > \tau_j \geq \tau^0_j$, $\tau_j \leq \tau^0_{j-1} < \tau^0_j$, and $\tau^0_{j-1} \leq \tau_j < \tau^0_j$. These ordering permutations can be handled with analogous arguments as below, all yielding the same uniform lower bounds stated in the lemma. Another observation here is that under the assumed ordering, for any $\tau_j \in \mathcal{G}_j(u_T, v_T)$ and on the set $\mathcal{A}$ and applicable in the construction of the squared loss Q_j, we have the following relation,

$$\frac{(\tau_j - \tau^0_{j+1})}{(\tau^0_{j+1} - \tau^0_j)} \leq \frac{(\hat{\tau}_{j+1} - \tau^0_{j+1})}{(\tau^0_{j+1} - \tau^0_j)} \leq \frac{c_{u1} T \underline{\ell}}{T \ell_j} \leq c_{u1}, \tag{S1.5}$$

for a suitably small constant $c_{u1} > 0$. Here the first inequality follows from the construction of the refitted squared loss Q_j, i.e, the search space is restricted to $\{\hat{\tau}_{j-1}, ..., \hat{\tau}_{j+1} - 1\}$. The second inequality follows from Condition C(i) (on numerator) and the definition of ℓ_j (on denominator). The final inequality follows from definition of $\underline{\ell}$. The relation (S1.5) directly implies the following,

$$\frac{(\tau_j - \tau^0_{j+1})}{(\tau_j - \tau^0_j)} = \frac{(\tau_j - \tau^0_{j+1})}{(\tau_j - \tau^0_j) + (\tau^0_{j+1} - \tau^0_j)} \leq c_{u1}, \tag{S1.6}$$

on the event $\mathcal{A}$ and for all $\tau_j \in \mathcal{G}_j(u_T, v_T)$ under consideration in context of the squared loss Q_j. As a final observation, consider the expression,

$$\begin{aligned} &\|\hat{\eta}_{(j)}\|_2^2 + 2(\hat{\theta}_{(j+1)} - \theta^0_{(j+1)})^T \hat{\eta}_{(j)} - 2\frac{(\tau_j - \tau^0_{j+1})}{(\tau_j - \tau^0_j)}(\theta^0_{(j+2)} - \theta^0_{(j+1)})^T \hat{\eta}_{(j)} \\ &\quad= \|\eta^0_{(j)} + (\hat{\eta}_{(j)} - \eta^0_{(j)})\|_2^2 + 2(\hat{\theta}_{(j+1)} - \theta^0_{(j+1)})^T \hat{\eta}_{(j)} \\ &\quad\quad -2\frac{(\tau_j - \tau^0_{j+1})}{(\tau_j - \tau^0_j)}(\theta^0_{(j+2)} - \theta^0_{(j+1)})^T \hat{\eta}_{(j)} \end{aligned} \tag{S1.7}$$

$$\begin{aligned}
&\geq \|\eta^0_{(j)}\|_2^2 + 2(\hat{\eta}_{(j)} - \eta^0_{(j)})^T\eta^0_{(j)} + 2(\hat{\theta}_{(j+1)} - \theta^0_{(j+1)})^T\hat{\eta}_{(j)} \\
&\quad -2\frac{(\tau_j - \tau^0_{j+1})}{(\tau_j - \tau^0_j)}(\theta^0_{(j+2)} - \theta^0_{(j+1)})^T\hat{\eta}_{(j)} \\
&\geq \|\eta^0_{(j)}\|_2^2 - 2\|\hat{\eta}_{(j)} - \eta^0_{(j)}\|_2\|\eta^0_{(j)}\|_2 - 2\|\hat{\theta}_{(j+1)} - \theta^0_{(j+1)}\|_2\|\hat{\eta}_{(j)}\|_2 \\
&\quad -2\frac{(\tau_j - \tau^0_{j+1})}{(\tau_j - \tau^0_j)}\|\theta^0_{(j+2)} - \theta^0_{(j+1)}\|_2\|\hat{\eta}_{(j)}\|_2 \\
&\geq \xi_j^2\Big(1 - \frac{4c_{u1}}{(Ns)^{1/2}\log(p\vee T)} - \frac{3c_{u1}}{(Ns)^{1/2}\log(p\vee T)} - 3c_{u1}c_u\Big) \geq \frac{\xi_j^2}{2},
\end{aligned}$$

which holds on the set $\mathcal{A}$ with probability at least $1-\pi_T$. Note that the first inequality is simply an algebraic manipulation, the second follows from the Cauchy-Schwartz inequality and the third follows from Condition C.1, (S1.3) and (S1.4) together with Condition B(iv) and (S1.6).

We can now proceed to the main proof of the uniform bound of this lemma. Consider the following decomposition under the ordering $\tau_j \geq \tau^0_{j+1} > \tau^0_j$,

$$\begin{aligned}
\mathcal{U}_j(\tau_j, \hat{\tau}, \hat{\theta}) &= Q_j(\tau_j, \hat{\tau}_{-j}, \hat{\theta}) - Q_j(\tau^0_j, \hat{\tau}_{-j}, \hat{\theta}) \\
&= \sum_{t=\hat{\tau}_{j-1}}^{\tau_j}\|x_t - \hat{\theta}_{(j)}\|^2 + \sum_{t=\tau_j+1}^{\hat{\tau}_{j+1}}\|x_t - \hat{\theta}_{(j+1)}\|^2 \\
&\quad - \sum_{t=\hat{\tau}_{j-1}}^{\tau^0_j}\|x_t - \hat{\theta}_{(j)}\|^2 + \sum_{t=\tau^0_j+1}^{\hat{\tau}_{j+1}}\|x_t - \hat{\theta}_{(j+1)}\|^2 \\
&= \sum_{t=\tau^0_j+1}^{\tau_j}\|x_t - \hat{\theta}_{(j)}\|^2 - \sum_{t=\tau^0_j+1}^{\tau_j}\|x_t - \hat{\theta}_{(j+1)}\|^2 \\
&= (\tau_j - \tau^0_j)\Big\{\|\hat{\eta}_{(j)}\|_2^2 + 2(\hat{\theta}_{(j+1)} - \theta^0_{(j+1)})^T\hat{\eta}_{(j)} \\
&\quad -2\frac{(\tau_j - \tau^0_{j+1})}{(\tau_j - \tau^0_j)}(\theta^0_{(j+2)} - \theta^0_{(j+1)})^T\hat{\eta}_{(j)}\Big\} - 2\sum_{t=\tau^0_j+1}^{\tau_j}\varepsilon_t^{*T}\hat{\eta}_{(j)}
\end{aligned}$$

$$
\begin{aligned}
&\geq \frac{Tv_T\xi_j^2}{2} - 2\sum_{t=\tau_j^0+1}^{\tau_j} \varepsilon_t^{*T}\hat{\eta}_{(j)} \\
&= \frac{Tv_T\xi_j^2}{2} - 2\sum_{t=\tau_j^0+1}^{\tau_j} \varepsilon_t^{*T}\eta_{(j)}^0 - 2\sum_{t=\tau_j^0+1}^{\tau_j} \varepsilon_t^{*T}(\hat{\eta}_{(j)} - \eta_{(j)}^0),
\end{aligned}
\tag{S1.8}
$$

on the event $\mathcal{A}$ which holds with probability at least $1 - \pi_T$. Here the first inequality follows from (S1.7) and construction of the set $\mathcal{G}_j(u_T, v_T)$. The noise variables ε_t^* arise from the reparametrized model (2.1). We now consider uniform upper bounds for each of the stochastic terms in the expression (S1.8). First, applying Lemma S3.4 for any $0 < a < 1$, with $c_a \geq \sqrt{(1/a)}$, we have,

$$
\sup_{\substack{\tau_j \in \mathcal{G}_j(u_T, v_T);\\ \tau_j \geq \tau_j^0}} 2\Big|\sum_{t=\tau^0+1}^{\tau} \varepsilon_t^{*T}\eta_{(j)}^0\Big| \leq c_u c_a \sigma \xi_j \big(Tu_T\big)^{\frac{1}{2}} \tag{S1.9}
$$

w.p. at least $1 - 2a$. The second stochastic term in (S1.8) can be bounded above as,

$$
2\sum_{t=\tau_j^0+1}^{\tau_j} \varepsilon_t^{*T}(\hat{\eta}_{(j)} - \eta_{(j)}^0) \leq 2\Big\|\sum_{t=\tau_j^0+1}^{\tau_j} \varepsilon_t^*\Big\|_\infty \big\|\hat{\eta}_{(j)} - \eta_{(j)}^0\big\|_1 \leq c_u \sigma \underline{\xi}\big(Tu_T\big)^{\frac{1}{2}}, \tag{S1.10}
$$

w.p. at least $1 - o(1) - \pi_T$. Here the second inequality follows using the deviation bounds in Lemma S3.2 together with the ℓ_1 error bound of (S1.3). Substituting (S1.9) and (S1.10) in (S1.8), we obtain,

$$
\begin{aligned}
\inf_{\substack{\tau_j \in \mathcal{G}_j(u_T, v_T);\\ \tau_j \geq \tau_j^0}} \mathcal{U}_j(\tau_j, \hat{\tau}_{-j}, \hat{\theta}) &\geq \frac{Tv_T\xi_j^2}{2} - c_u c_a \sigma \xi_j \big(Tu_T\big)^{\frac{1}{2}} - c_u \sigma \underline{\xi}\big(Tu_T\big)^{\frac{1}{2}} \\
&\geq \frac{T\xi_j^2}{2}\Big[v_T - \frac{c_u c_a \sigma}{\xi_j}\Big(\frac{u_T}{T}\Big)^{\frac{1}{2}}\Big]
\end{aligned}
$$

w.p. at least $1 - 2a - o(1) - \pi_T$. The remaining permutations of $\tau_{j+1}^0 > \tau_j \geq \tau_j^0$, $\tau_j \leq \tau_{j-1}^0 < \tau_j^0$, and $\tau_{j-1}^0 \leq \tau_j < \tau_j^0$, can be handled with analogous arguments. This completes the proof of the lemma. $\square$

Lemma S1.2. *Suppose conditions A, B and C hold and let* $0 \le v_T \le u_T \le 1$, *be any non-negative sequences. Then, we have the following uniform lower bound*

$$\min_{1\le j\le N}\inf_{\tau_j\in\mathcal{G}_j(u_T,v_T)}\mathcal{U}_j(\tau_j,\hat\tau_{-j},\hat\theta)\;\ge\;\frac{\underline{\xi}^2}{2}\Big[v_T-\frac{c_u\sigma}{\underline{\xi}}\Big(\frac{u_T\log^2 T}{T}\Big)^{\frac12}\Big],$$

with probability at least $1-o(1)-\pi_T$.

Proof of Lemma S1.2. The structure of this proof is similar to that of Lemma S1.1, with an additional uniformity required over $j=1,...,N$, which in turn requires utilizing stochastic bounds with this additional uniformity (Lemma S3.3). Proceeding identically as in Lemma S1.1, under the ordering $\tau_j\ge\tau_{j+1}^0>\tau_j^0$, we have (S1.8), i.e.,

$$\mathcal{U}_j(\tau_j,\hat\tau,\hat\theta)\ge\frac{Tv_T\xi_j^2}{2}-2\sum_{t=\tau_j^0+1}^{\tau_j}\varepsilon_t^{*T}\eta_{(j)}^0-2\sum_{t=\tau_j^0+1}^{\tau_j}\varepsilon_t^{*T}(\hat\eta_{(j)}-\eta_{(j)}^0),\tag{S1.11}$$

on the event $\mathcal{A}$ which holds with probability at least $1-\pi_T$. Now consider each of the stochastic terms in (S1.11) and apply the bounds of Lemma S3.3 which possess the required additional uniformity over $j=1,...,N$. First, from Part (ii) of Lemma S3.3,

$$\max_{1\le j\le N}\sup_{\substack{\tau_j\in\mathcal{G}_j(u_T,v_T);\\ \tau_j\ge\tau_j^0}}\Big|\sum_{t=\tau^0+1}^{\tau}\varepsilon_t^{*T}\eta_{(j)}^0\Big|\le c_u\bar\xi\sigma\{Tu_T\log^2 T\}^{\frac12}\tag{S1.12}$$

w.p. at least $1-o(1)$. Second,

$$\begin{aligned}&\max_{1\le j\le N}\sup_{\substack{\tau_j\in\mathcal{G}_j(u_T,v_T);\\ \tau_j\ge\tau_j^0}}\sum_{t=\tau_j^0+1}^{\tau_j}\varepsilon_t^{*T}(\hat\eta_{(j)}-\eta_{(j)}^0)\\ &\le\max_{1\le j\le N}\sup_{\substack{\tau_j\in\mathcal{G}_j(u_T,v_T);\\ \tau_j\ge\tau_j^0}}\Big\|\sum_{t=\tau_j^0+1}^{\tau_j}\varepsilon_t^{*}\Big\|_\infty\big\|\hat\eta_{(j)}-\eta_{(j)}^0\big\|_1\le c_u\sigma\underline{\xi}\big(Tu_T\big)^{\frac12},\end{aligned}\tag{S1.13}$$

w.p. at least $1-o(1)-\pi_T$, where the second inequality follows from Part (i) of Lemma S3.3 together with the ℓ_1 error bound of (S1.3). Substituting (S1.12) and (S1.13) in

(S1.11) yields,

$$\min_{1\le j\le N}\inf_{\substack{\tau_j\in\mathcal{G}_j(u_T,v_T);\\ \tau_j\ge\tau_j^0}}\mathcal{U}_j(\tau_j,\hat\tau,\hat\theta) \ge \frac{Tv_T\underline{\xi}^2}{2} - c_u\sigma\overline{\xi}\big(Tu_T\log^2 T\big)^{\frac{1}{2}} - c_u\sigma\underline{\xi}\big(Tu_T\big)^{\frac{1}{2}}$$

$$\ge \frac{T\underline{\xi}^2}{2}\Big[v_T - \frac{c_u\sigma}{\underline{\xi}}\Big(\frac{u_T\log^2 T}{T}\Big)^{\frac{1}{2}}\Big],$$

w.p. at least $1-o(1)-\pi_T$. Note that we have also utilized $\overline{\xi}\le c_u\underline{\xi}$ of Condition B(iv). The same bound can be obtained for all other permutations of the ordering of τ_js w.r.t τ_j^0s via analogous arguments. This completes the proof of this lemma. □

Proof of Theorem 1. The proof of this result relies on a recursive argument on Lemma S1.1, where the desired rate of convergence is obtained by a series of recursions, with this rate being sharpened at each step. We begin by considering any given $j=1,...,N$, and any $v_T>0$ and applying Lemma S1.1 on the set $\mathcal{G}_j(1,v_T)$ to obtain,

$$\inf_{\tau_j\in\mathcal{G}_j(1,v_T)}\mathcal{U}_j(\tau_j,\hat\tau_{-j},\hat\theta)\ge\frac{\xi_j^2}{2}\Big[v_T-\frac{c_uc_a\sigma}{\xi_j}\Big(\frac{1}{T}\Big)^{\frac{1}{2}}\Big].$$

with probability at least $1-2a-o(1)-\pi_T$. Now upon choosing any,

$$v_T>v_T^*=\frac{c_uc_a\sigma}{\xi_j}\Big(\frac{1}{T}\Big)^{\frac{1}{2}},$$

we obtain $\inf_{\tau_j\in\mathcal{G}_j(1,v_T)}\mathcal{U}_j(\tau_j,\hat\tau_{-j},\hat\theta)>0$, thus implying that $\tilde\tau_j\notin\mathcal{G}_j(1,v_T^*)$, i.e., $|\tilde\tau_j-\tau_j^0|\le Tv_T^*$, with probability at least $1-2a-o(1)-\pi_T$.[1] Now reset $u_T=v_T^*$ and reapply Lemma S1.1 for any $v_T>0$ to obtain,

$$\inf_{\tau_j\in\mathcal{G}_j(u_T,v_T)}\mathcal{U}_j(\tau_j,\hat\tau_{-j},\hat\theta)\ge\frac{\xi_j^2}{2}\Big[v_T-\Big(\frac{c_uc_a\sigma}{\xi_j}\Big)^{1+\frac{1}{2}}\Big(\frac{1}{T}\Big)^{\frac{1}{2}+\frac{1}{4}}\Big],$$

[1]Since by construction of $\tilde\tau_j$, we have $\mathcal{U}_j(\tau_j,\hat\tau_{-j},\hat\theta)\le 0$.

with probability at least $1-2a-o(1)-\pi_T$.. Again choosing any,

$$v_T > v_T^* = \Big(\frac{c_u c_a \sigma}{\xi_j}\Big)^{1+\frac{1}{2}}\Big(\frac{1}{T}\Big)^{\frac{1}{2}+\frac{1}{4}},$$

we obtain $\inf_{\tau_j\in\mathcal{G}_j(u_T,v_T)}\mathcal{U}_j(\tau_j,\hat\tau_{-j},\hat\theta)>0$, thus yielding $\tilde\tau_j\notin\mathcal{G}_j(u_T,v_T^*)$, i.e.,

$$|\tilde\tau_j-\tau_j^0|\le T\Big(\frac{c_u c_a \sigma}{\xi_j}\Big)^{a_2}\Big(\frac{1}{T}\Big)^{b_2}, \tag{S1.14}$$

with probability at least $1-2a-o(1)-\pi_T$. Where,

$$a_2 = 1+\frac{1}{2}=\sum_{j=0}^{1}\frac{1}{2^j}, \quad \text{and} \quad b_2=\frac{1}{2}+\frac{1}{4}=\sum_{j=1}^{2}\frac{1}{2^j}.$$

Note that the rate of convergence of $\tilde\tau$ has been sharpened at the second recursion in comparison to the first. Continuing these recursions by resetting u_T to the bound of the previous recursion, and applying Lemma S1.1, we obtain for the m^{th} recursion,

$$|\tilde\tau_j-\tau_j^0|\le T\Big(\frac{c_u c_a \sigma}{\xi_j}\Big)^{a_m}\Big(\frac{1}{T}\Big)^{b_m}, \tag{S1.15}$$

with probability at least $1-2a-o(1)-\pi_T$. Repeating these recursions an infinite number of times and noting that $a_\infty=\sum_{j=0}^{\infty}(1/2^j)=2$, and $b_\infty=\sum_{j=1}^{\infty}(1/2^j)=1$ we obtain,

$$|\tilde\tau_j-\tau_j^0|\le T\Big(\frac{c_u c_a \sigma}{\xi_j}\Big)^{2}\frac{1}{T}=c_u^2 c_a^2\sigma^2\xi_j^{-2}$$

with probability at least $1-2a-o(1)-\pi_T$. Finally, note that despite the recursions in the above argument, the probability of the bound after every recursion is maintained to be at least $1-2a-o(1)-\pi_T$. This follows since the probability statement of Lemma S1.1 arises from the stochastic upper bounds in Lemma sS3.2 and S3.4 applied recursively with a tighter bound at each recursion. This yields a sequence of events such that the event at each recursion is a proper subset of the event at the previous recursion. This completes the proof of this theorem. □

Proof of Theorem 2. The argument to follow is largely similar to that of the proof of Theorem 1. It is a recursive argument applied on Lemma S1.2 which possesses uniformity over $j = 1, ..., N$, in comparison to Lemma S1.1 which does not. Recall that this uniformity is gained in exchange for a weaker bound in comparison to Lemma S1.1. Consider any $v_T > 0$ and apply Lemma S1.2 on the sets $\mathcal{G}_j(1, v_T)$, $j = 1, ..., N$, to obtain,

$$\min_{1\le j\le N} \inf_{\tau_j\in\mathcal{G}_j(1,v_T)} \mathcal{U}_j(\tau_j, \hat{\tau}_{-j}, \hat{\theta}) \ge \frac{\underline{\xi}^2}{2}\Big[v_T - \frac{c_u\sigma}{\underline{\xi}}\Big(\frac{\log^2 T}{T}\Big)^{\frac{1}{2}}\Big].$$

with probability at least $1 - o(1) - \pi_T$. Upon choosing any,

$$v_T > v_T^* = \frac{c_u\sigma}{\xi_j}\Big(\frac{\log^2 T}{T}\Big)^{\frac{1}{2}},$$

we obtain $\min_{1\le j\le N} \inf_{\tau_j\in\mathcal{G}_j(1,v_T)} \mathcal{U}_j(\tau_j, \hat{\tau}_{-j}, \hat{\theta}) > 0$, thus implying that $\tilde{\tau}_j \notin \mathcal{G}_j(1, v_T^*)$, $\forall j = 1, ..., N$, with probability at least $1-o(1)-\pi_T$. Note that this implies $\max_{1\le j\le N} |\tilde{\tau}_j - \tau_j^0| \le T v_T^*$, with the same probability. Now reset $u_T = v_T^*$ and reapply Lemma S1.2 for any $v_T > 0$ to obtain,

$$\min_{1\le j\le N} \inf_{\tau_j\in\mathcal{G}_j(u_T,v_T)} \mathcal{U}_j(\tau_j, \hat{\tau}_{-j}, \hat{\theta}) \ge \frac{\underline{\xi}_j}{2}\Big[v_T - \Big(\frac{c_u\sigma}{\underline{\xi}}\Big)^{1+\frac{1}{2}}\Big(\frac{\log^2 T}{T}\Big)^{\frac{1}{2}+\frac{1}{4}}\Big],$$

with probability at least $1 - o(1) - \pi_T$.. Again choosing any,

$$v_T > v_T^* = \Big(\frac{c_u\sigma}{\underline{\xi}}\Big)^{1+\frac{1}{2}}\Big(\frac{\log^2 T}{T}\Big)^{\frac{1}{2}+\frac{1}{4}},$$

we obtain $\min_{1\le j\le N}\inf_{\tau_j\in\mathcal{G}_j(u_T,v_T)}\mathcal{U}_j(\tau_j,\hat{\tau}_{-j},\hat{\theta})>0$, thus yielding $\tilde{\tau}_j\notin\mathcal{G}_j(u_T,v_T^*)$, $\forall j$, i.e.,

$$\max_{1\le j\le N}|\tilde{\tau}_j-\tau_j^0|\le T\Big(\frac{c_u\sigma}{\underline{\xi}}\Big)^{a_2}\Big(\frac{\log^2 T}{T}\Big)^{b_2}, \tag{S1.16}$$

with probability at least $1-o(1)-\pi_T$. Where,

$$a_2=1+\frac{1}{2}=\sum_{j=0}^{1}\frac{1}{2^j},\quad\text{and}\quad b_2=\frac{1}{2}+\frac{1}{4}=\sum_{j=1}^{2}\frac{1}{2^j}.$$

Continuing these recursions by resetting u_T to the bound of the previous recursion, and applying Lemma S1.2, we obtain for the m^{th} recursion,

$$\max_{1\le j\le N}|\tilde{\tau}_j-\tau_j^0|\le T\Big(\frac{c_u\sigma}{\underline{\xi}}\Big)^{a_m}\Big(\frac{\log^2 T}{T}\Big)^{b_m}, \tag{S1.17}$$

with probability at least $1-o(1)-\pi_T$. Repeating these recursions an infinite number of times and noting that $a_\infty=\sum_{j=0}^{\infty}(1/2^j)=2$, and $b_\infty=\sum_{j=1}^{\infty}(1/2^j)=1$ we obtain,

$$\max_{1\le j\le N}|\tilde{\tau}_j-\tau_j^0|\le T\Big(\frac{c_u\sigma}{\underline{\xi}}\Big)^{2}\frac{\log^2 T}{T}=c_u^2\sigma^2\underline{\xi}^{-2}\log^2 T$$

with probability at least $1-o(1)-\pi_T$. As in the proof of Theorem 1, despite the recursions in the above argument, the probability of the bound after every recursion is maintained to be at least $1-o(1)-\pi_T$, for the same reason as described there. Thus completing the proof of this theorem. $\square$

As the reader may have observed, a change of notation has been carried out for the results of Theorems 3 and 4. These results are presented in more conventional *argmax* notation instead of the *argmin* notation of the problem setup in Section 1. This is purely a notational change and all results can equivalently be stated in the *argmin* language. Accordingly we define the following versions. Let $\mathcal{U}_j(\tau_j, \tau_{-j}, \theta)$ $j = 1, ..., N$, be as defined in (S1.2) and consider,

$$\mathcal{C}_j(\tau, \tau_{-j}, \theta) = -\mathcal{U}_j(\tau, \tau_{-j}, \theta), \quad j = 1, ..., N. \tag{S1.18}$$

Then, we can re-express the change point estimators $\tilde{\tau}_j(\hat{\tau}_{-j}, \hat{\theta})$ as,

$$\tilde{\tau}_j(\hat{\tau}_{-j}, \hat{\theta}) = \underset{\hat{\tau}_{j-1} \le \tau_j < \hat{\tau}_{j+1}}{\arg\max}\ \mathcal{C}_j(\tau_j, \hat{\tau}_{-j}, \hat{\theta}), \quad j = 1, ..., N.$$

The proofs of Theorem 3, Theorem 4 and Theorem 5 below are applications of the Argmax Theorem (reproduced as Theorem S4.2 in Appendix S4). The arguments here are largely an exercise in verification of requirements of this theorem.

Proof of Theorem 3. Consider any fixed $j = 1, ..., N$, then we begin by noting that although $(\tilde{\tau}_j - \tau_j^0)$ is discrete, however the sequence whose limiting distribution is being examined is $\xi_j^2(\tilde{\tau}_j - \tau_j^0)$, consequently, the underlying indexing metric space here is $\mathbb{R}$. Now consider the two cases of known and unknown plug-in parameters.

Case I $\big(\tau_{-j}^0$ **and** θ^0 **known**$\big)$**:** The following requirements of the Argmax theorem need to be verified for this case (see, page 288 of Vaart and Wellner (1996)).

1. The sequence $\xi_j^2(\tilde{\tau}_j^* - \tau_j^0)$ is uniformly tight (see, Definition S4.3 in Appendix S4).

2. $\big\{2\sigma_{(\infty,j)}W_j(\zeta) - |\zeta|\big\}$ satisfies suitable regularity conditions[2].

3. For any $\zeta \in [-c_u, c_u]$ we have

$$\mathcal{C}_j(\tau_j^0 + \zeta\xi_j^{-2}, \tau_{-j}^0, \theta^0) \Rightarrow \big\{2\sigma_{(\infty,j)}W_j(\zeta) - |\zeta|\big\}.$$

Note that by setting $\hat{\theta} = \theta^0$, and $\hat{\tau}_{-j} = \tau^0_{-j}$, the requirements of Condition C are trivially satisfied. Now using Theorem 1 we have that $\xi_j^2(\tilde{\tau}_j^* - \tau_j^0) = O_p(1)$. This yields requirement (1). The second requirement follows from well known properties of Brownian motion. The only remaining requirement is (3), which is provided below.

$$\zeta \leftarrow (\zeta - \xi_j^2) \le \xi_j^2 \lfloor \zeta \xi_j^{-2} \rfloor \le (\zeta + \xi_j^2) \rightarrow \zeta$$

Hence, w.l.o.g. we may assume $\zeta\xi_j^{-2}$ is integer valued. Now for any $\zeta \in (0, c_u]$, consider

$$\begin{aligned}
\mathcal{C}_j(\tau_j^0 + \zeta\xi_j^{-2}, \tau^0_{-j}, \theta^0) &= -\sum_{t=(\tau_j^0+1)}^{\tau_j^0+\zeta\xi_j^{-2}} \left\{ \|x_t - \theta^0_{(j)}\|_2^2 - \|x_t - \theta^0_{(j+1)}\|_2^2 \right\} \\
&= 2\sum_{t=(\tau_j^0+1)}^{\tau_j^0+\zeta\xi_j^{-2}} \varepsilon_t^{*T}\eta^0_{(j)} - \zeta \\
&= 2\sum_{t=(\tau_j^0+1)}^{\tau_j^0+\zeta\xi_j^{-2}} \varepsilon_t^{T}\eta^0_{(j)} - \zeta - 2\zeta\xi_j^{-2}\bar{\varepsilon}^T\eta^0_{(j)} \\
&= 2\sum_{t=(\tau_j^0+1)}^{\tau_j^0+\zeta\xi_j^{-2}} \varepsilon_t^{T}\eta^0_{(j)} - \zeta - o_p(1) \\
&\Rightarrow 2\sigma_{(\infty,j)}W_{1j}(\zeta) - \zeta,
\end{aligned} \tag{S1.19}$$

where the final equality follows from (S3.8) together with Condition B(iv) and Condition E. The weak convergence follows from the functional central limit theorem. Repeating the same argument with $\zeta \in [-c_u, 0)$, yields $\mathcal{C}(\tau^0 + \zeta\xi_j^{-2}, \tau^0_{-j}, \theta^0) \Rightarrow 2\sigma_{(\infty,j)}W_{2j}(-\zeta) - |\zeta|$. his completes the proof of requirement (3) for the Argmax theorem and consequently an application of its results yields $\xi_j^2(\tilde{\tau}_j^* - \tau_j^0) \Rightarrow \arg\max_{\zeta\in\mathbb{R}} \big\{2\sigma_{(\infty,j)}W_j(\zeta) -$

[2]Almost all sample paths $\zeta \rightarrow \{2\sigma_{(\infty,j)}W_j(\zeta) - |\zeta|\}$ are upper semicontinuous and posses a unique maximum at a (random) point $\arg\max_{\zeta\in\mathbb{R}} \{2\sigma_{(\infty,j)}W_j(\zeta) - |\zeta|\}$, which as a random map in the indexing metric space is tight.

$|\zeta|\}$, which completes the proof of this case.

Case II $\left(\tau^0_{-j}\right.$ and θ^0 unknown$\left.\right)$: In this case, to apply the Argmax theorem requires verifying the following conditions.

1. The sequence $\xi_j^2(\tilde{\tau}_j - \tau_j^0)$ is uniformly tight.

2. $\{2\sigma_{(\infty,j)}W_j(\zeta) - |\zeta|\}$ satisfies suitable regularity conditions.

3. For any $\zeta \in [-c_u, c_u]$ we have

$$\mathcal{C}_j(\tau_j^0 + \zeta\xi_j^{-2}, \hat{\tau}_{-j}, \hat{\theta}) \Rightarrow \{2\sigma_{(\infty,j)}W_j(\zeta) - |\zeta|\}.$$

Part (i) again follows from the result of Theorem 1 under the assumed Condition C on the nuisance estimates $\hat{\tau}_{-j}$ and $\hat{\theta}$. Part (2) is identical to the corresponding requirement of Case I. Finally to prove part (3) note that from Lemma S1.3 we have that,

$$\sup_{\tau_j \in \mathcal{G}_j\left(c_u T^{-1}\xi_j^{-2}, 0\right)} |\mathcal{C}_j(\tau_j, \hat{\tau}_{-j}, \hat{\theta}) - \mathcal{C}_j(\tau_j, \tau^0_{-j}, \theta^0)| = o_p(1). \tag{S1.20}$$

The approximation (S1.20) and Part (3) of Case I together imply Part (3) for this case. This completes the verification of all requirements for this case. The stated limiting distribution now follows by an application of the Argmax theorem. □

Proof of Theorem 4. The proof of this theorem is similar to Theorem 3 in that it is also an application of the Argmax theorem. The distinction here is in the limiting distribution that is induced by the change of regime of the jump size. Consider any given $j = 1, ..., N$ and the discrete sequence $(\tilde{\tau}_w - \tau_w^0)$, consequently the underlying indexing metric space here is $\mathbb{Z}$. Now consider the two cases of known and unknown plug-in parameters.

Case I $\big(\tau^0_{-j}$ and θ^0 known$\big)$: The requirements to be verified here are as follows.

1. The sequence $(\tilde\tau^*_j - \tau^0_j)$ is uniformly tight.

2. $\mathcal{C}_{(\infty,j)}(\zeta)$ satisfies suitable regularity conditions.

3. For any $\zeta \in \{-c_u, -c_u+1, ..., -1, 0, 1, ...c_u\}$, we have

$$\mathcal{C}_j(\tau^0_j+\zeta, \tau^0_{-j}, \theta^0) \Rightarrow \mathcal{C}_{(\infty,j)}(\zeta).$$

As in the proof of Theorem 3, requirement (1) follows directly from the result of Theorem 1. Requirement (2) of regularity of the *argmax* of two sided negative drift random walk $\mathcal{C}_{(\infty,j)}(\zeta)$ has been proved earlier in Lemma A.3 of the Supplement in Kaul et al. (2021). The requirement (3) is verified next. For any $\zeta \in \{1, 2, ..., c_u\}$, consider

$$\begin{aligned}
\mathcal{C}_j(\tau^0_j+\zeta, \tau^0_{-j}, \theta^0) &= -\sum_{t=(\tau^0_j+1)}^{\tau^0_j+\zeta} \big\{\|x_t-\theta^0_{(j)}\|_2^2 - \|x_t-\theta^0_{(j+1)}\|_2^2\big\} \\
&= \sum_{t=(\tau^0_j+1)}^{\tau^0_j+\zeta} \Big(2\varepsilon^{*T}_t\eta^0_{(j)} - \xi^2_j\Big) \\
&= \sum_{t=(\tau^0_j+1)}^{\tau^0_j+\zeta} \Big(2\varepsilon^{T}_t\eta^0_{(j)} - \xi^2_j - o_p(1)\Big) \\
&\Rightarrow \sum_{t=1}^{\zeta} \mathcal{P}\big(-\xi^2_{(\infty,j)}, 4\xi^2_{(\infty,j)}\sigma^2_{(\infty,j)}\big), \qquad \text{(S1.21)}
\end{aligned}$$

The final equality follows similarly to the final equality of (S1.19). The convergence in distribution follows from Condition A$'$, Condition D together with Slutsky's theorem. Repeating the same argument with $\zeta \in \{-c_u, -c_u+1, ..., -1\}$, yields $\mathcal{C}_j(\tau_j^0+\zeta, \tau_{-j}^0, \theta^0) \Rightarrow \sum_{t=1}^{-\zeta} \mathcal{P}\big(-\xi_{(\infty,j)}^2, 4\xi_{(\infty,j)}^2\sigma_{(\infty,j)}^2\big)$. An application the Argmax theorem now yields $(\tilde{\tau}_j^* - \tau_j^0) \Rightarrow \arg\max_{\zeta\in\mathbb{Z}} \mathcal{C}_{(\infty,j)}(\zeta)$, which completes the proof of this case.

Case II $\big(\tau_{-j}^0$ **and** θ^0 **unknown**$\big)$**:** In this case, the applicability of the Argmax Theorem requires verification of the following.

- *(i)* The sequence $(\tilde{\tau}_j - \tau_j^0)$ is uniformly tight.
- *(ii)* $\mathcal{C}_{(\infty,j)}(\zeta)$ satisfies suitable regularity conditions.
- *(iii)* For any $\zeta \in \{-c_u, -c_u+1, ..., -1, 0, 1, ...c_u\}$ we have, $\mathcal{C}(\tau_j^0+\zeta, \hat{\tau}_{-j}, \hat{\theta}) \Rightarrow \mathcal{C}_{(\infty,j)}(\zeta)$.

Part (i) follows from Theorem 1 under the assumed Condition C on the nuisance estimates $\hat{\tau}_{-j}$ and $\hat{\theta}$. Part (ii) is identical to the corresponding requirement of Case I. Finally to prove part (iii) note that from Lemma S1.3 we have that,

$$\sup_{\tau_j \in \mathcal{G}_j(c_u T^{-1}\xi_j^{-2}, 0)} |\mathcal{C}_j(\tau_j, \hat{\tau}_{-j}, \hat{\theta}) - \mathcal{C}_j(\tau_j, \tau_{-j}^0, \theta^0)| = o_p(1). \tag{S1.22}$$

The approximation (S1.22) and Part (iii) of Case I together imply Part (iii) for this case. This completes the verification of all requirements for this case. The statement of the theorem now follows by an application of the Argmax theorem. □

Lemma S1.3. *Suppose Conditions A, B, C and E hold and let* $\mathcal{C}_j(\tau_j, \tau_{-j}, \hat{\theta})$ *be as in (S1.18). Further, assume that* r_T *in Condition C satisfies* $r_T = \{o(1)\underline{\xi}\}/\{(Ns)^{1/2}\log(p \vee T)\}$.

Then, for any given $j = 1, ..., N$ and any $c_u > 0$, we obtain

$$\sup_{\tau_j \in \mathcal{G}_j\left(c_u T^{-1}\xi_j^{-2}, 0\right)} \left|\mathcal{C}_j(\tau_j, \hat{\tau}_{-j}, \hat{\theta}) - \mathcal{C}_j(\tau_j, \tau^0_{-j}, \theta^0)\right| = o_p(1).$$

Proof of Lemma S1.3. Recall that

$$\begin{aligned}
\mathcal{C}_j(\tau_j, \hat{\tau}_{-j}, \hat{\theta}) &= \sum_{t=\hat{\tau}_{j-1}+1}^{\tau_j} \|x_t - \hat{\theta}_{(j)}\|_2^2 + \sum_{t=\tau_j+1}^{\hat{\tau}_{j+1}} \|x_t - \hat{\theta}_{(j+1)}\|_2^2 \\
\mathcal{C}_j(\tau_j, \tau^0_{-j}, \theta^0) &= \sum_{t=\tau^0_{j-1}+1}^{\tau_j} \|x_t - \theta^0_{(j)}\|_2^2 + \sum_{t=\tau_j+1}^{\tau^0_{j+1}} \|x_t - \theta^0_{(j+1)}\|_2^2,
\end{aligned}$$

Further, note that $\mathcal{C}_j(\tau_j, \hat{\tau}_{-j}, \hat{\theta})$ and $\mathcal{C}_j(\tau_j, \tau^0_{-j}, \theta^0)$ are sums over indices whose start and end points may differ. Despite this incoherence, the desired supremum over collection $\mathcal{G}_j\left(c_u T^{-1}\xi_j^{-2}, 0\right)$ is well defined. This is enforced by Condition C together with the rate assumption of Condition F. These ensure that the collection of τ_j's over which the supremum is evaluated remains between $\max\{\hat{\tau}_{j-1}, \tau^0_{j-1}\}$ and $\min\{\hat{\tau}_{j+1}, \tau^0_{j+1}\}$ with probability $1 - \pi_T$. Specifically, from Condition B(ii) we have that $(\tau^0_j - \tau^0_{j-1}) \geq T\underline{\ell}$, $\forall j$ and from Condition C that $\max_j |\hat{\tau}_j - \tau^0_j| \leq c_{u1} T\underline{\ell}$, w.p. $1 - \pi_T$. In addition, the left and right end points of the set $\mathcal{G}_j\left(c_u T^{-1}\xi_j^{-2}, 0\right)$, i.e., $\tau^0_j - c_u\xi_j^{-2}$ and $\tau^0_j + c_u\xi_j^{-2}$, respectively. Consequently, the rate assumption of Condition F forces these end points to be in a sufficiently small neighborhood of τ^0_j, such that all values of τ_j in the collection $\mathcal{G}_j\left(c_u T^{-1}\xi_j^{-2}, 0\right)$, remain away from $\hat{\tau}_{j-1}, \tau^0_{j-1}$ as well as $\hat{\tau}_{j+1}, \tau^0_{j+1}$, w.p. $1 - \pi_T$, thus allowing the desired supremum of interest to be well defined.

The second observation is that by proceeding as in (S1.3), under Condition C with

$r_T = \{o(1)\underline{\xi}\}/\{(Ns)^{1/2}\log(p\vee T)\}$, we get

$$\|\hat{\eta}_{(j)} - \eta^0_{(j)}\|_1 \leq c_{u1}\sqrt{Ns}\|\hat{\eta}_{(j)} - \eta^0_{(j)}\|_2 = \frac{o(1)\underline{\xi}}{\log(p\vee T)} \tag{S1.23}$$

w.p. at least $1-\pi_T$. Next, consider any $\tau_j \geq \tau^0_j$ and define the following:

$$R_1 = 2\sum_{t=\tau^0_j+1}^{\tau_j} \varepsilon_t^{*T}(\hat{\eta}_{(j)} - \eta^0_{(j)}), \quad \text{and} \quad R_2 = (\tau_j - \tau^0_j)\big(\|\hat{\eta}_{(j)}\|_2^2 - \|\eta^0_{(j)}\|_2^2\big).$$

Then, under the orientation $\min\{\hat{\tau}_{j+1}, \tau^0_{j+1}\} > \tau_j \geq \tau^0_j$, we have the following algebraic expansion,

$$\begin{aligned}
&\mathcal{C}_j(\tau_j, \hat{\tau}_{-j}, \hat{\theta}) - \mathcal{C}_j(\tau_j, \tau^0_{-j}, \theta^0) \\
&\qquad = 2\sum_{t=\tau^0_j+1}^{\tau_j} \varepsilon_t^{*T}(\hat{\eta}_{(j)} - \eta^0_{(j)}) - (\tau_j - \tau^0_j)\big(\|\hat{\eta}_{(j)}\|_2^2 - \|\eta^0_{(j)}\|_2^2\big) \\
&\qquad = R_1 - R_2
\end{aligned} \tag{S1.24}$$

Next, we provide uniform bounds for the terms R_1 and R_2 of (S1.24). Consider

$$\begin{aligned}
\sup_{\substack{\tau_j\in\mathcal{G}_j(c_uT^{-1}\xi_j^{-2},0);\\ \tau_j\geq\tau^0_j}} |R_1| &\leq 2\sup_{\substack{\tau_j\in\mathcal{G}_j(c_uT^{-1}\xi_j^{-2},0);\\ \tau_j\geq\tau^0_j}} \Big\|\sum_{t=\tau^0_j+1}^{\tau_j}\varepsilon^*_t\Big\|_\infty \|\hat{\eta}_{(j)} - \eta^0_{(j)}\|_1 \\
&\leq c_u\sigma\xi_j^{-1}\log(p\vee T)\|\hat{\eta}_{(j)} - \eta^0_{(j)}\|_1 = o(1),
\end{aligned} \tag{S1.25}$$

w.p. at least $1-o(1)-\pi_T$. The second inequality follows from Lemma S3.2, while the final equality follows from an application of (S1.23). Next, consider term R_2 of (S1.24)

$$\sup_{\substack{\tau_j\in\mathcal{G}(c_uT^{-1}\xi_j^{-2},0);\\ \tau_j\geq\tau^0_j}} |R_2| \leq c_u\xi_j^{-2}\big|\|\hat{\eta}_{(j)}\|_2^2 - \|\eta^0_{(j)}\|_2^2\big| \tag{S1.26}$$

$$\begin{aligned}
&= c_u\xi_j^{-2}\big|\|\hat{\eta}_{(j)} - \eta^0_{(j)}\|_2^2 + 2(\hat{\eta}_{(j)} - \eta^0_{(j)})^T\eta^0_{(j)}\big| \\
&\leq c_u\xi_j^{-2}\|\hat{\eta}_{(j)} - \eta^0_{(j)}\|_2^2 + 2c_u\xi_j^{-1}\|\hat{\eta}_{(j)} - \eta^0_{(j)}\|_2 = o_p(1),
\end{aligned}$$

wherein the second inequality follows as an application of the Cauchy-Schwarz inequality and the final equality follows from (S1.23). Applying (S1.25) and (S1.26) in the expression (S1.24) yields

$$\sup_{\substack{\tau_j\in\mathcal{G}_j(c_uT^{-1}\xi_j^{-2},0);\\ \tau_j\geq\tau_j^0}}\big|\mathcal{C}_j(\tau_j,\hat{\tau}_{-j},\hat{\theta})-\mathcal{C}_j(\tau_j,\tau_{-j}^0,\theta^0)\big|$$
$$\leq \sup_{\substack{\tau_j\in\mathcal{G}_j(c_uT^{-1}\xi_j^{-2},0);\\ \tau_j\geq\tau_j^0}}|R_1|+\sup_{\substack{\tau_j\in\mathcal{G}_j(c_uT^{-1}\xi_j^{-2},0);\\ \tau_j\geq\tau_j^0}}|R_2|=o_p(1)$$

Per the discussion in the first paragraph, the only other orientation allowed for any τ_j in the set $\mathcal{G}_j(c_uT^{-1}\xi_j^{-2},0)$ is $\max\{\hat{\tau}_{j-1},\tau_{j-1}^0\}\leq\tau_j<\tau_j^0$, w.p. $1-\pi_T$. Hence, the same bound for this mirroring orientation can be obtained via symmetrical arguments. This completes the proof of the lemma. $\square$

The proof of Theorem 5 is also an application of the Argmax Theorem. However, we require preliminary work in order to establish a framework for this problem that can fit into the setup of the theorem. To that end, introduce some additional notation. Let $H\subseteq\{1,...,N\}$ be any finite subset, and let $\hat{\tau},\hat{\theta}$ be the preliminary estimates as discussed in the main article. Define a new estimator

$$\breve{\tau}_H=\operatorname*{arg\,max}_{\substack{\tau_H\in\mathbb{Z}^{|H|};\\ \hat{\tau}_{j-1}<\tau_j<\hat{\tau}_{j+1};\\ \forall j\in H}}\sum_{j\in H}\mathcal{C}_j(\tau_j,\hat{\tau}_{-j},\hat{\theta}),\tag{S1.27}$$

with $\mathcal{C}_j(\tau_j,\hat{\tau}_{-j},\hat{\theta})$ defined in (S1.18). Then, all but the j^{th} summand in (S1.27) are constants in the j^{th} component of the maximizing argument τ_H, and thus this estimator

(S1.27) is the same as the component-wise refitted estimates of (2.4), i.e.,

$$\breve{\tau}_j = \tilde{\tau}_j \quad \forall j \in H. \tag{S1.28}$$

Proof of Theorem 5. Let $H \subseteq \{1, ..., N\}$ be any finite subset. Recall that by the non-vanishing jump size regime assumption, we have, $\xi_j \to \xi_{(\infty,j)}$, $0 < \xi_{(\infty,j)} < \infty$, $\forall j \in H$. The proceeding argument shall apply the Argmax theorem in context of the H dimensional sequence $(\breve{\tau}_H - \tau^0_H)$ of (S1.27), the limiting result of which shall pass over to the proposed $\tilde{\tau}_H$ due to the equality (S1.28). Clearly, the underlying indexing metric space is $\mathbb{Z}^{|H|}$. The reequirements to be verified for the Argmax theorem are:

1. The sequence $(\breve{\tau}_H - \tau^0_H)$ is uniformly tight in $\mathbb{Z}^{|H|}$.

2. The $\mathbb{Z}^{|H|} \to \mathbb{R}$ random field $\sum_{j\in H} \mathcal{C}_{(\infty,j)}(\zeta_j)$ satisfies suitable regularity conditions.

3. For any $\zeta = (\zeta_1, ..., \zeta_{|H|})^T \in \big\{-\mathbf{c_u}, ..., \mathbf{c_u}\big\}_{|H|\times 1}$, with $\mathbf{c_u} \in \mathbb{Z}^{+|H|}$, we have

$$\sum_{j\in H} \mathcal{C}_j(\tau^0_j + \zeta_j, \hat{\tau}_{-j}, \hat{\theta}) \Rightarrow \sum_{j\in H} \mathcal{C}_{(\infty,j)}(\zeta_j). \tag{S1.29}$$

From Theorem 1 we have for each fixed $j \in H$, the sequence $(\tilde{\tau}_j - \tau^0_j)$ is uniformly tight in $\mathbb{Z}$. Then, the equality (S1.28) together with the assumption that $|H|$ is finite, implies Requirement (1). The second requirement is verified in Lemma S1.4 below. To prove requirement (3), first note that from Lemma S1.3 and the finiteness of $|H|$, we obtain

$$\sup_{\tau_H \in \mathcal{G}_H\left(\mathbf{c_u}T^{-1},0\right)} \Big| \sum_{j\in H} \mathcal{C}_j(\tau_j, \hat{\tau}_{-j}, \hat{\theta}) - \sum_{j\in H} \mathcal{C}_j(\tau_j, \tau^0_{-j}, \theta^0) \Big| = o_p(1).$$ [3]

Thus, to complete the proof of requirement (3) it only remains to show that

$$\sum_{j\in H}\mathcal{C}_j(\tau_j^0+\zeta_j,\tau_{-j}^0,\theta^0)\Rightarrow\sum_{j\in H}\mathcal{C}_{(\infty,j)}(\zeta_j),\tag{S1.30}$$

where the increments of $\mathcal{C}_{(\infty,j)}(\zeta_j)$ are additionally independent over all j's. In all arguments to follow, we assume w.l.o.g. $|H|=2$, where $H=\{1,2\}$. Let $\zeta_1>0$, and $\zeta_2<0$. Proceeding analogously as in (S1.21), we get

$$\begin{aligned}
&\mathcal{C}_1(\tau_1^0+\zeta_1,\tau_{-1}^0,\theta^0)+\mathcal{C}_2(\tau_2^0+\zeta_2,\tau_{-2}^0,\theta^0)\\
&\quad=\sum_{t=(\tau_1^0+1)}^{\tau_1^0+\zeta_1}\Big(2\varepsilon_t^{*T}\eta_{(1)}^0-\xi_1^2\Big)+\sum_{t=(\tau_2^0+\zeta_2+1)}^{\tau_2^0}\Big(2\varepsilon_t^{*T}\eta_{(2)}^0-\xi_2^2\Big)\\
&\quad=\sum_{t=(\tau_1^0+1)}^{\tau_1^0+\zeta_1}\Big(2\varepsilon_t^{T}\eta_{(1)}^0-\xi_1^2\Big)+\sum_{t=(\tau_2^0+\zeta_2+1)}^{\tau_2^0}\Big(2\varepsilon_t^{T}\eta_{(2)}^0-\xi_2^2\Big)-o_p(1)\\
&\quad\Rightarrow\sum_{j=1}^{2}\sum_{t=1}^{|\zeta_j|}z_{tj},
\end{aligned}$$

where $z_{tj}\sim\mathcal{P}\big(-\xi_{(\infty,j)}^2,4\xi_{(\infty,j)}^2\sigma_{(\infty,j)}^2\big)$, for each t and each j, which are independent over all t and j. Independence over j's follows since by Condition B(ii), we have $(\tau_2^0-\tau_1^0)\geq T\underline{\ell}\to\infty$; consequently, for T sufficiently large, the two sums of interest are over non-overlapping indices, i.e., $\tau_1^0+\zeta_1<\tau_2^0+\zeta_2$. The second equality of (S1.31) follows analogously to (S1.19). The weak convergence follows from Condition A$'$. The remaining permutations of the signs of ζ_1,ζ_2 can be handled symmetrically to yield the same result. This completes the verification of requirement (3). An application of the

[3]Here $\mathcal{G}_H\big(\mathbf{c_u}T^{-1},0\big)=\mathcal{G}_{j_1}\big(c_uT^{-1},0\big)\times....\times\mathcal{G}_{j_{|H|}}\big(c_uT^{-1},0\big)$, for $H=\{j_1,....,j_{|H|}\}$.

Argmax theorem together with the equality (S1.28) now yields

$$(\tilde{\tau}_H - \tau_H^0) = (\breve{\tau}_H - \tau_H^0) \Rightarrow \arg\max_{\zeta\in\mathbb{Z}^{|H|}} \sum_{j\in H} \mathcal{C}_{(\infty,j)}(\zeta_j)$$

thereby establishing the first claim of the theorem. Next, note that

$$\arg\max_{\zeta\in\mathbb{Z}^{|H|}} \sum_{j\in H} \mathcal{C}_{(\infty,j)}(\zeta_j) = \Big(\arg\max_{\zeta_1\in\mathbb{Z}} \mathcal{C}_{(\infty,1)}(\zeta_1),\ \arg\max_{\zeta_2\in\mathbb{Z}} \mathcal{C}_{(\infty,2)}(\zeta_2)\Big)^T. \tag{S1.31}$$

This equality follows along the same lines as (S1.28).Also note that (S1.31) is an exact equality and not just equality in distribution. The independence of $\mathcal{C}_{(\infty,j)}(\zeta_j)$ over j as discussed earlier implies

$$\Big(\arg\max_{\zeta_1\in\mathbb{Z}} \mathcal{C}_{(\infty,1)}(\zeta_1),\ \arg\max_{\zeta_2\in\mathbb{Z}} \mathcal{C}_{(\infty,2)}(\zeta_2)\Big)^T =^d \Pi_{j\in H} \arg\max_{\zeta_j\in\mathbb{Z}} \mathcal{C}_{(\infty,j)}(\zeta_j),$$

which proves the second claim of this theorem. The final claim of asymptotic independence of $\tilde{\tau}_j$, over $j \in H$ now follows by comparing (3.11) to the marginal distributions obtained in Theorem 4. This completes the proof of the theorem. $\square$

Lemma S1.4. *Suppose Conditions A′, B and D hold, and let $H \subseteq \{1, 2, ..., N\}$ be any finite subset and assume the non-vanishing jump size regime of $\xi_j \to \xi_{(\infty,j)}$, $0 < \xi_{(\infty,j)} < \infty$, $\forall j \in H$. Let $\mathcal{C}_{(\infty,j)}(\zeta_j)$ be as defined in (3.7). Then, the $\mathbb{Z}^{|H|} \to \mathbb{R}$ map $\sum_{j\in H} \mathcal{C}_{(\infty,j)}(\zeta_j)$ is continuous with respect to the domain space. Additionally, $\arg\max_{\zeta\in\mathbb{Z}^{|H|}} \sum_{j\in H} \mathcal{C}_{(\infty,j)}(\zeta_j)$ possesses an almost sure unique maximum at ω_∞ which as a random map in $\mathbb{Z}^{|H|}$ is tight.*

Proof of Lemma S1.4. This proof has been adapted from Lemma A.3 of Kaul et al. (2021) for the process under consideration. Continuity of sample paths of the random field $\sum_{j\in H}\mathcal{C}_{(\infty,j)}(\zeta_j)$ follows trivially since the domain space $\mathbb{Z}^{|H|}$ is discrete. Next, from Condition A$'$ we have that incremental distributions are continuous, thus, if $\max_{\zeta\in\mathbb{Z}^{|H|}}\sum_{j\in H}\mathcal{C}_{(\infty,j)}(\zeta_j) < \infty$ a.s. then ω_∞ must be unique and tight. Consequently, the only thing that remains to show is that $\max_{\zeta\in\mathbb{Z}^{|H|}}\sum_{j\in H}\mathcal{C}_{(\infty,j)}(\zeta_j) < \infty$ a.s., for this purpose, w.l.o.g. let $|H| = 2$ with $H = \{1,2\}$. Now consider any fixed $\zeta_2 = \zeta_2^0$, and note that $\sum_{j\in H}\mathcal{C}_{(\infty,j)}(\zeta_j)$ is a two sided random walk over ζ_1. Moreover, under the assumed non-vanishing jump size this two sided random walk is negative drift, from Condition D the incremental variances are finite, and from the assumed underlying subexponential distribution, all moments of incremental distributions exist. Consequently, we have $\sum_{j\in H}\mathcal{C}_{(\infty,j)}(\zeta_j) \to -\infty$, as $\zeta_1 \to \infty$ or $\zeta_1 \to \infty$, a.s. (strong law of large numbers). This implies that $\max_{\zeta_1\in\mathbb{Z}}\sum_{j\in H}\mathcal{C}_{(\infty,j)}(\zeta_j) < \infty$, a.s. (follows from the Hewitt-Savage 0-1 law, see, e.g. (1.1) and (1.2) on Page 172, 173 of Durrett (2010)). Next, applying union bounds over the countable collection of $\zeta_2 \in \mathbb{Z}$, yields $\max_{\zeta_2\in\mathbb{Z}}\max_{\zeta_1\in\mathbb{Z}}\sum_{j\in H}\mathcal{C}_{(\infty,j)}(\zeta_j) < \infty$, a.s. (countable intersection of a.s. events is a.s.), thereby completing the proof of the lemma. □

Proof of Theorem 6. The main idea of this proof is first to prove the weak convergence of the two underlying stochastic processes, i.e., the two sided random walk and the Brownian motion on the lhs and rhs of (3.13), repsectively, followed by an application of continuous mapping type results to obtain the weak convergence of the desired *argmax*.

The first immediate roadblock towards this approach is the incoherence of the indexing spaces of the stochastic processes on the lhs and rhs of (3.13). To alleviate this incoherence one may consider representing the lhs of (3.13) as $\arg\max_{r\in\mathbb{R}}\mathcal{C}_\infty(\lfloor r\rfloor)$. This representation is however not well defined due to the non-uniqueness of the *argmax* functional in this case. Thus, *argmax* needs to be re-defined as the smallest maximizer: $\operatorname{sargmax} f(x) = \min\{x;\ f(x) \ge f(y)\ \forall\ y\}$. The functional sargmax has been studied in the literature, e.g., Lan et al. (2009) and Seijo and Sen (2011) whose motivations are exactly the same that arise here. Under this definition, one can re-write the lhs of (3.13) as,

$$\begin{aligned}\arg\max_{\zeta\in\mathbb{Z}}\mathcal{C}_\infty\big(\zeta,\xi_T,\sigma_T^2\big) &=^d \operatorname*{sargmax}_{\zeta'\in\mathbb{R}}\mathcal{C}_\infty\big(\lfloor\zeta'\rfloor,\xi_T,\sigma_T^2\big)\\ &=^d \lfloor\xi_T^{-2}\rfloor\operatorname*{sargmax}_{\zeta\in\mathbb{R}}\mathcal{C}_\infty\big(\lfloor\zeta\lfloor\xi_T^{-2}\rfloor\rfloor,\xi_T,\sigma_T^2\big),\end{aligned} \tag{S1.32}$$

where the second equality follows directly from a change of variables $\zeta' = \zeta\lfloor\xi_T^{-2}\rfloor$. Next, consider the random walk on the rhs of (S1.32) as per the defining relation (3.6).

$$\begin{aligned}\mathcal{C}_\infty\big(\lfloor\zeta\lfloor\xi_T^{-2}\rfloor\rfloor,\xi_T,\sigma_T^2\big) &= \begin{cases}\sum_{t=1}^{-\lfloor\zeta\lfloor\xi_T^{-2}\rfloor\rfloor}\mathcal{P}\big(-\xi_T^2,\ 4\xi_T^2\sigma_T^2\big), & \zeta\in\mathbb{R}^-\\ 0, & \zeta=0\\ \sum_{t=1}^{\lfloor\zeta\lfloor\xi_T^{-2}\rfloor\rfloor}\mathcal{P}\big(-\xi_T^2,\ 4\xi_T^2\sigma_T^2\big), & \zeta\in\mathbb{R}^+.\end{cases}\\ &=^d \begin{cases}\sum_{t=1}^{-\lfloor\zeta\lfloor\xi_T^{-2}\rfloor\rfloor}\Big[\mathcal{P}\big(0,\ 4\xi_T^2\sigma_T^2\big)-\xi_T^2\Big], & \zeta\in\mathbb{R}^-\\ 0, & \zeta=0\\ \sum_{t=1}^{\lfloor\zeta\lfloor\xi_T^{-2}\rfloor\rfloor}\Big[\mathcal{L}\big(0,\ 4\xi_T^2\sigma_T^2\big)-\xi_T^2\Big], & \zeta\in\mathbb{R}^+.\end{cases}\end{aligned}$$

where the second equality follows from the additive invariance of $\mathcal{P}$ w.r.t scalar addition.

Next, consider the positive arm ($\zeta > 0$) of this process, to obtain

$$\begin{aligned}\sum_{t=1}^{\lfloor\zeta\lfloor\xi_T^{-2}\rfloor\rfloor}\Big[\mathcal{P}\big(0,\ 4\xi_T^2\sigma_T^2\big)-\xi_T^2\Big] &=^d \ \lfloor\xi_T\rfloor\sum_{t=1}^{\lfloor\zeta\lfloor\xi_T^{-2}\rfloor\rfloor}\mathcal{P}\big(0,\ 4\lfloor\xi_T^{-2}\rfloor\xi_T^2\sigma_T^2\big)-\xi_T^2\lfloor\zeta\lfloor\xi_T^{-2}\rfloor\rfloor\\ &\Rightarrow\ 2\sigma_\infty W(\zeta)-\sigma_\infty^2\zeta \qquad\qquad \text{(S1.33)}\end{aligned}$$

The equality follows from the invariance of $\mathcal{P}$ w.r.t scalar multiplication. The weak convergence follows from the functional central limit theorem, together with the limit assumptions on the underlying sequences, specifically, $\xi_T \to 0$, $\sigma_T^2 \to \sigma_\infty^2$. Here we have also utilized the elementary result $\xi_T^2\lfloor\zeta\lfloor\xi_T^{-2}\rfloor\rfloor \to \zeta$. The relation (S1.33) together with a symmetric result on the negative arm ($\zeta < 0$) of this process yields,

$$\mathcal{C}_\infty\big(\lfloor\zeta\lfloor\xi_T^{-2}\rfloor\rfloor,\xi_T,\sigma_T^2\big)\begin{cases}2\sigma_\infty W_1(\zeta)+\sigma_\infty^2\zeta & \text{if } \zeta<0,\\ 0, & \text{if } \zeta=0,\\ 2\sigma_\infty W_2(\zeta)-\sigma_\infty^2\zeta & \text{if } \zeta>0,\end{cases} \qquad \text{(S1.34)}$$

Now applying the continuous mapping theorem for the *sargmax* functional (Lemma 3.1 of Lan et al. (2009) or Theorem 3.1 of Seijo and Sen (2011)) we obtain,

$$\begin{aligned}\underset{\zeta\in\mathbb{R}}{\text{sargmax}}\,\mathcal{C}_\infty\big(\lfloor\zeta\lfloor\xi_T^{-2}\rfloor\rfloor,\xi_T,\sigma_T^2\big) &\Rightarrow\ \arg\max_{\zeta\in\mathbb{R}} 2\sigma_\infty W(\zeta)-\sigma_\infty^2|\zeta|\\ &=^d\ \sigma_\infty^{-2}\arg\max_{\zeta\in\mathbb{R}} 2W(\zeta)-|\zeta|,\end{aligned}$$

The equality follows from a change of variables (also see, proof of Theorem 3). Also note that *sargmax* of the rhs has been replaced by *argmax*, since the rhs possesses a unique maximizer. Finally, the statement of the theorem now follows by a back substitution to the relation (S1.32) and noting that $\xi_T^2\lfloor\xi_T^{-2}\rfloor \to 1$. This completes the proof of this theorem. □

S2 Proofs of results in Section 4

The proof of Theorem 7 requires some preliminary work. For any non-negative sequence $u_T \le 1$, define the collection,

$$\overline{\mathcal{G}}(u_T) \;\; = \;\; \Big\{\tau \in \{1, ..., T-1\}^N; \; \max_{1 \le j \le N} |\tau_j - \tau_j^0| \le T u_T \Big\}$$

We begin by first examining the behavior of the estimates $\tilde{\theta}_{(j)}(\tau)$, $j = 1, ..., N+1$, uniformly over the collection $\overline{\mathcal{G}}(u_T)$. This is provided in the following theorem.

Theorem S2.1. *Suppose Conditions A and B(i, ii) hold and let $\overline{\mathcal{G}}$ be as defined in (S2.1). Let $0 \le u_T \le c_{u1}\underline{\ell}$, be any sequence with a suitably small constant $c_{u1} > 0$, and let $\psi = \max_j \|\eta_{(j)}^0\|_\infty$. Further, assume $T\underline{\ell} \ge \log(p \vee T)$ and for any constants $c_u, c_{u2}, > 0$, and $j = 1, ..., N$, let*

$$\lambda_j = \lambda = 16 \max \Big[\sigma \Big\{\frac{2c_{u2} \log(p \vee T)}{c_u T \underline{\ell}}\Big\}^{\frac{1}{2}}, \; \frac{u_T \psi}{c_u \underline{\ell}}\Big]. \tag{S2.1}$$

Then, $\hat{\theta}_{(j)}(\tau)$, $j = 1, ..., N+1$ of (4.3) satisfy the following two results

(i) For any $\tau \in \overline{\mathcal{G}}_w(u_T)$, and any $j = 1, ..., N+1$, such that $(\tau_j - \tau_{j-1}) \ge c_u T\underline{\ell}$, we have $\big\|\big(\hat{\theta}_{(j)}(\tau)\big)_{S_j^c}\big\|_1 \le 3\big\|\big(\hat{\theta}_{(j)}(\tau) - \theta_{(j)}^0\big)_{S_j}\big\|_1$, for sets S_j as defined in (3.1).

(ii) The following bound is satisfied

$$\max_{1 \le j \le N+1} \sup_{\substack{\tau \in \overline{\mathcal{G}}(u_T); \\ \min_j(\tau_j - \tau_{j-1}) \ge c_u T\underline{\ell}}} \|\hat{\theta}_{(j)}(\tau) - \theta_{(j)}^0\|_2 \le 6\sqrt{(Ns)}\lambda,$$

where both parts (i) and (ii) hold with probability at least $1-4\exp\big\{-(c_{u3}-4)\log(p\vee T)\big\}$, $c_{u3}=c_{u2}\wedge\sqrt{(c_{u2}c_u/2)}$..

Proof of Theorem S2.1. We begin with an observation that proves useful for the ensuing argument. The assumption $u_T\le c_{u1}\underline{\ell}$, yields that for any $\tau=(\tau_1,...,\tau_N)^T\in\overline{\mathcal{G}}(u_T)$, we obtain $\max_j|\tau_j-\tau_j^0|\le c_{u1}T\underline{\ell}$. recall from Condition B(ii), that all change points are separated by at least $T\underline{\ell}$, i.e., $\min_j(\tau_j-\tau_{j-1})\ge T\underline{\ell}$. Consequently, any $\tau\in\overline{\mathcal{G}}(u_T)$ must satisfy any one of the four orientations $\tau_{j-1}^0\le\tau_{j-1}<\tau_j^0\le\tau_j$, $\tau_{j-1}\le\tau_{j-1}^0<\tau_j^0\le\tau_j$, $\tau_{j-1}^0\le\tau_{j-1}<\tau_j\le\tau_j^0$ or $\tau_{j-1}\le\tau_{j-1}^0<\tau_j\le\tau_j^0$, for any $j=1,...,N$. No other orientations are feasible under these assumed conditions. In view of this observation, w.l.o.g. we assume one of the first of these four possible orientations, $\tau_{j-1}^0\le\tau_{j-1}<\tau_j^0\le\tau_j$ in the argument to follow. The remaining three permutations of the ordering of τ_{j-1},τ_j w.r.t. τ_{j-1}^0,τ_j^0, can be proved using symmetrical arguments.

Let $\tau\in\overline{\mathcal{G}}(u_T)$ additionally satisfy the relation $\min_j(\tau_j-\tau_{j-1})\ge c_uT\underline{\ell}$, then an algebraic rearrangement of the elementary inequality $\big\|\bar{x}_{(j)}(\tau)-\hat{\theta}_{(j)}(\tau)\big\|_2^2+\lambda_j\|\hat{\theta}_{(j)}(\tau)\|_1\le\big\|\bar{x}_{(j)}(\tau)-\theta_{(j)}^0\big\|_2^2+\lambda_j\|\theta_{(j)}^0\|_1$ yields,

$$
\begin{aligned}
\big\|\hat{\theta}_{(j)}(\tau)-\theta_{(j)}^0\big\|_2^2+\lambda_j\big\|\hat{\theta}_{(j)}(\tau)\big\|_1 &\le \lambda_j\big\|\theta_{(j)}^0\big\|_1+2\sum_{t=\tau_{j-1}+1}^{\tau_j}\hat{\varepsilon}_t^{*T}(\hat{\theta}_{(j)}(\tau)-\theta_{(j)}^0),\\
&= \lambda_j\big\|\theta_{(j)}^0\big\|_1+\frac{2}{(\tau_j-\tau_{j-1})}\sum_{t=\tau_{j-1}+1}^{\tau_j}\varepsilon_t^{*T}(\hat{\theta}_{(j)}(\tau)-\theta_{(j)}^0)\\
&\quad -2\frac{(\tau_j-\tau_j^0)}{(\tau_j-\tau_{j-1})}(\theta_{(j)}^0-\theta_{(j+1)}^0)^T(\hat{\theta}_{(j)}(\tau)-\theta_{(j)}^0)
\end{aligned}
$$

$$\leq \lambda_j \big\|\theta^0_{(j)}\big\|_1 + \frac{2}{(\tau_j-\tau_{j-1})}\Big\|\sum_{t=\tau_{j-1}}^{\tau_j}\varepsilon^*_t\Big\|_\infty \big\|\hat\theta_{(j)}(\tau)-\theta^0_{(j)}\big\|_1$$
$$+\frac{2u_T}{c_u\underline{\ell}}\psi\big\|\hat\theta_{(j)}(\tau)-\theta^0_{(j)}\big\|_1, \qquad \text{(S2.2)}$$

where in the first inequality we have $\hat\varepsilon^*_t = \big(x_t-\theta^0_{(j)}\big)$. The last inequality follows since $\tau\in\overline{\mathcal{G}}(u_T)$, and by definition $\|\theta^0_{(j)}-\theta^0_{(j+1)}\|_\infty\leq\psi$. Now using the bound of Lemma S3.5 we have that,

$$\frac{2}{(\tau_j-\tau_{j-1})}\Big\|\sum_{t=\tau_{j-1}}^{\tau_j}\varepsilon^*_t\Big\|_\infty \leq 4\surd(2c_{u2}/c_u)\sigma\Big\{\frac{\log(p\vee T)}{T\underline{\ell}}\Big\}^{\frac{1}{2}} \qquad \text{(S2.3)}$$

with probability at least $1-4\exp\big\{-(c_{u3}-4)\log(p\vee T)\big\}$, $c_{u3}=c_{u2}\wedge\surd(c_{u2}c_u/2)$. Consequently, upon choosing,

$$\lambda^* = 8\max\Big\{\surd(2c_{u2}/c_u)\sigma\Big\{\frac{\log(p\vee T)}{Tl_T}\Big\}^{\frac{1}{2}},\ \frac{u_T\psi}{c_u l_T}\Big\},$$

and substituting in (S2.2), we obtain

$$\big\|\hat\theta_{(j)}(\tau)-\theta^0_{(j)}\big\|_2^2+\lambda_j\big\|\hat\theta_{(j)}(\tau)\big\|_1 \leq \lambda_j\big\|\theta^0_{(j)}\big\|_1+\lambda^*\big\|\hat\theta_{(j)}(\tau)-\theta^0_{(j)}\big\|_1, \qquad \text{(S2.4)}$$

with probability at least $1-4\exp\big\{-(c_{u2}-4)\log(p\vee T)\big\}$. choosing $\lambda_j=2\lambda^*$, leads to $\|\big(\hat\theta_{(j)}(\tau)\big)_{S^c_j}\|_1\leq 3\|\big(\hat\theta_{(j)}(\tau)-\theta^0_{(j)}\big)_{S_j}\|_1$, which upon noting that the bound (S2.3) arises from Lemma S3.5 which holds uniformly over j as well as over all considered values of τ, proves part (i) of this theorem.

Next, from inequality (S2.4) we also have that,

$$\|\hat\theta_{(j)}(\tau)-\theta^0_{(j)}\|_2^2\leq\frac{3}{2}\lambda_j\|\hat\theta_{(j)}(\tau)-\theta^0_{(j)}\|_1\leq 6\lambda_j\surd(Ns)\|\hat\theta_{(j)}(\tau)-\theta^0_{(j)}\|_2 \qquad \text{(S2.5)}$$

This directly implies that $\|\hat\theta_{(j)}(\tau)-\theta^0_{(j)}\|_2\leq 6\lambda_j\surd(Ns)$, where we have used $\|\hat\theta_{(j)}(\tau)-\theta^0_{(j)}\|_1\leq 4\sqrt{(Ns)}\|\hat\theta_{(j)}(\tau)-\theta^0_{(j)}\|_2$, which follows in turn Part (i). To complete the proof of this part recall that the only stochastic bound used here is the uniform bound of Lemma S3.5; consequently, the final bound also holds uniformly over the same collection. This result can alternatively be proved using the properties of the soft-thresholding operator $k_\lambda(\cdot)$, by building uniform versions of arguments such as those in Kaul et al. (2017). $\square$

Proof of Theorem 7. To prove the first claim, note that by Condition E$'$(i) we have,

$$c_u\sigma^2\underline{\xi}^{-2}Ns\log^2(p\vee T)\le c_{u1}T\underline{\ell} \tag{S2.6}$$

This relation together with assumed properties (4.5) of the preliminary change point estimates imply that Condition C(i) is satisfied. The remaining claims of this theorem are largely an application of Theorem S2.1. Note that relations (4.5) and (S2.6) imply that $\hat\tau$ lies in the collection over which the uniform results of Theorem 7 are established, i.e., $\hat\tau\in\overline{\mathcal{G}}(u_T)$, $u_T=c_uT^{-1}\sigma^2\xi^{-2}Ns\log^2(p\vee T)\le\underline{\ell}$, and $\min_j(\hat\tau_{j-1}-\hat\tau_j)\ge c_uT\underline{\ell}$, w.p. $1-\pi_T$. Now consider λ as defined in (S2.1) with this choice of u_T,

$$\begin{aligned}
\lambda &= c_u\max\Big[\sigma\Big\{\frac{\log(p\vee T)}{T\underline{\ell}}\Big\}^{\frac12},\ \frac{\psi\sigma^2\underline{\xi}^{-2}Ns\log^2(p\vee T)}{T\underline{\ell}}\Big]\\
&= c_u\sigma\Big\{\frac{\log(p\vee T)}{T\underline{\ell}}\Big\}^{\frac12}\max\Big[1,\ \Big(\frac{\psi}{\underline{\xi}}\Big)\Big(\frac{\sigma}{\underline{\xi}}\Big)\Big\{\frac{Ns\log^{3/2}(p\vee T)}{\surd(T\underline{\ell})}\Big\}\Big]\\
&\le c_u\sigma\Big\{\frac{\log(p\vee T)}{T\underline{\ell}}\Big\}^{\frac12},
\end{aligned} \tag{S2.7}$$

wherein the inequality follows by using the assumption $\psi/\underline{\xi}=O(1)$, together with Condition E$'$(ii). The second claim and the bound (4.6) now follows from the corresponding results of Theorem S2.1. To establish the final claim, note that from Condition E$'$(ii) we also have that,

$$c_u \sigma \Big\{ \frac{Ns \log(p \vee T)}{T \underline{\ell}} \Big\}^{\frac{1}{2}} \le \frac{c_{u1} \underline{\xi}}{(Ns)^{1/2} \log(p \vee T)} \tag{S2.8}$$

Thus, (4.6) together with (S2.8) imply that $\hat{\theta}_{(j)}(\hat{\tau})$, $j = 1, ..., N+1$ satisfy all requirements of Condition C(ii). This completes the proof of the theorem. □

Proof of Corollary 1. This result is a direct consequence of Theorem 7 and the results of Section 3. In particular, under the assumed conditions, Theorem 7 yields that the preliminary estimates $\hat{\tau}$ and $\hat{\theta}$ satisfy all requirements of Condition C. All claims of this now follow from corresponding results of Section 3. □

The proof of Lemma 1 relies on the following Lemma that provides an ℓ_∞ bound for the sample covariance obtained by centering the data based on estimated mean and change point parameters.

Lemma S2.1. *Let $\mathcal{A}$ be the event in (4.9) and assume conditions of Corolllary 1. Let $\tilde{\Sigma}$ be as defined in 4.8 then on the event $\mathcal{A}$, we have,*

$$\|\tilde{\Sigma} - \Sigma\|_\infty \le C \Big\{ \frac{\log(p \vee T)}{T \underline{\ell}} \Big\}^{1/2},$$

where C is a finite constant and $\Sigma = E\varepsilon_t \varepsilon_t^T$ as defined in Condition B.

Proof of Lemma S2.1. The proof of this result relies on a recent result in the literature. Define a version of the sample covariance computed via the true mean and change point parameters as,

$$\hat{\Sigma} = \frac{1}{T} \sum_{j=1}^{N+1} \sum_{t=\tau_{j-1}^0}^{\tau_j^0} (x_t - \theta_{(j)}^0)(x_t - \theta_{(j)}^0)^T = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^* \varepsilon_t^{*T} \tag{S2.9}$$

Then, we have

$$\|\hat{\Sigma} - \Sigma\|_\infty \leq C\Big\{\frac{\log(p \vee T)}{T}\Big\}^{1/2}, \tag{S2.10}$$

with probability at least $1 - o(1)$. The result S2.10 is well known under subgaussian distributions, see e.g., Yuan (2010). In the case of subexponential distributions as assumed in our article, the same result has recently been established in Kuchibhotla and Chakrabortty (2022) (see, Theorem 4.1 and Remark 4.1 therein)[4].

The remainder of this proof establishes $\|\tilde{\Sigma} - \hat{\Sigma}\|_\infty \leq C\sqrt{\log(p \vee T)/T\underline{\ell}}$ on the event $\mathcal{A}$. This in turn shall complete the proof of this result. For this purpose we shall utilize the following bounds that are induced by the set $\mathcal{A}$ and have also been proved in the supplement. Specifically,

$$\begin{aligned}
&(i) \quad \max_j |\tilde{\tau}_j - \tau_j^0| \leq C\underline{\xi}^{-2}\log^2 T,\\
&(ii) \quad \max_j \|\hat{\theta}_{(j)} - \theta_{(j)}^0\|_2 \leq C\Big\{\frac{Ns\log(p \vee T)}{T\underline{\ell}}\Big\}^{1/2}\\
&(iii) \quad \max_j \|\hat{\theta}_{(j)} - \theta_{(j)}^0\|_1 \leq CNs\Big\{\frac{\log(p \vee T)}{T\underline{\ell}}\Big\}^{1/2}\\
&(iv) \quad \max_j \|\hat{\theta}_{(j)} - \theta_{(j+1)}^0\|_2 \leq C\underline{\xi}, \quad \max_j \|\hat{\theta}_{(j)} - \theta_{(j+1)}^0\|_1 \leq C\sqrt{(s)}\underline{\xi}
\end{aligned} \tag{S2.11}$$

The first two follow directly from the set $\mathcal{A}$. The third has been illustrated in last para in the Proof of Theorem S2.1. The final two can be obtained directly by a triangle inequality as well as by recalling Condition B(iii) on sparsity s of $\eta_{(j)}^0 = (\theta_{(j)}^0 - \theta_{(j+1)}^0)$.

[4]Note that the article Kuchibhotla and Chakrabortty (2022) establishes (S2.10) for $\frac{1}{T}\sum_{t=1}^T \varepsilon_t\varepsilon_t^T$, it can be shown by arguments similar to those in Section S3 that the impact of $\bar{\varepsilon}$ in $\varepsilon_t^* = \varepsilon_t - \bar{\varepsilon}$ is ignorable.

We shall also require the following bounds on stochastic terms,

$$(v) \quad \max_j \Big\| \sum_{t=\tau_j^0+1}^{\tilde{\tau}_j} \varepsilon_t^* \Big\|_\infty \le C \underline{\xi}^{-1} \log(p \vee T) \log T$$

$$(vi) \quad \max_j \frac{1}{T} \Big| \sum_{t=\tau_j^0+1}^{\tau_{j+1}^0} \varepsilon_t^{*T} \delta \Big| \le C \frac{\log T}{\surd(T\underline{\ell})}, \tag{S2.12}$$

where δ is a unit vector. While both are stochastic bounds, however these are both implicit in the event $\mathcal{A}$, as both are utilized in order to obtain the results assumed in event $\mathcal{A}$. The first inequality can be obtained by (i) of Lemma S3.3 by substituting $Tu_T = C\underline{\xi}^{-2} \log^2 T$ which in turn follows from the restriction in set $\mathcal{A}$. The final inequality can be obtained by arguments analogous to (ii) of Lemma S3.3.

For ease of exposition, we illustrate the following relation in the case of two change points $N = 2$. A reproduction of the same arguments yields the same bound in the more general case an arbitrary N. Let $(\tilde{\tau}_1, \tilde{\tau}_2)$ represent the change point estimates of (τ_1^0, τ_2^0), recall that all arguments are on event $\mathcal{A}$ thus we implicitly assume $\hat{N} = N$ under this event. Consider the orientation $\tau_1^0 \le \tilde{\tau}_1 < \tau_2^0 \le \tilde{\tau}_2$, then we have,

$$\begin{aligned}
\big(\tilde{\Sigma} - \hat{\Sigma}\big) &= \frac{1}{T} \sum_{j=1}^{N+1} \sum_{t=\tilde{\tau}_{j-1}}^{\tilde{\tau}_j} (x_t - \hat{\theta}_{(j)})(x_t - \hat{\theta}_{(j)})^T - \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^* \varepsilon_t^{*T} \\
&= \frac{1}{T} \Big[\sum_{t=1}^{\tau_1^0} (x_t - \hat{\theta}_{(1)})(x_t - \hat{\theta}_{(1)})^T + \sum_{t=\tau_1^0+1}^{\tilde{\tau}_1} (x_t - \hat{\theta}_{(1)})(x_t - \hat{\theta}_{(1)})^T \\
&\quad + \sum_{t=\tilde{\tau}_1+1}^{\tau_2^0} (x_t - \hat{\theta}_{(2)})(x_t - \hat{\theta}_{(2)})^T + \sum_{t=\tau_2^0+1}^{\tilde{\tau}_2} (x_t - \hat{\theta}_{(2)})(x_t - \hat{\theta}_{(2)})^T \\
&\quad + \sum_{t=\tilde{\tau}_2^0+1}^{T} (x_t - \hat{\theta}_{(3)})(x_t - \hat{\theta}_{(3)})^T \Big] - \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^* \varepsilon_t^{*T}
\end{aligned} \tag{S2.13}$$

In order to bound the lhs of the above expression, decompose each term in the rhs as follows:

$$
\begin{aligned}
\sum_{t=1}^{\tau_1^0}(x_t-\hat\theta_{(1)})(x_t-\hat\theta_{(1)})^T &= \sum_{t=1}^{\tau_1^0}(\varepsilon_t^*-(\hat\theta_{(1)}-\theta^0_{(1)})(\varepsilon_t^*-(\hat\theta_{(1)}-\theta^0_{(1)})^T\\
&= \sum_{t=1}^{\tau_1^0}\varepsilon_t^*\varepsilon_t^{*T}+\tau_1^0\|\hat\theta_{(1)}-\theta^0_{(1)}\|_2^2-2\sum_{t=1}^{\tau_1^0}\varepsilon_t^*(\hat\theta_{(1)}-\theta^0_{(1)})^T\\
\sum_{t=\tau_1^0+1}^{\tilde\tau_1}(y_t-\hat\theta_{(1)})(y_t-\hat\theta_{(1)})^T &= \sum_{t=\tau_1^0+1}^{\tilde\tau_1}(\varepsilon_t^*-(\hat\theta_{(1)}-\theta^0_{(2)}))(\varepsilon_t^*-(\hat\theta_{(1)}-\theta^0_{(2)})^T\\
&= \sum_{t=\tau_1^0+1}^{\tilde\tau_1}\varepsilon_t^*\varepsilon_t^{*T}+(\tilde\tau_1-\tau_1^0)\|\hat\theta_{(1)}-\theta^0_{(2)}\|_2^2\\
&\quad -2\sum_{t=\tau_1^0+1}^{\tilde\tau_1}\varepsilon_t^*(\hat\theta_{(1)}-\theta^0_{(2)})^T\\
\sum_{t=\tilde\tau_1+1}^{\tau_2^0}(y_t-\hat\theta_{(2)})(y_t-\hat\theta_{(2)})^T &= \sum_{t=\tilde\tau_1+1}^{\tau_2^0}\varepsilon_t^*\varepsilon_t^{*T}+(\tau_2^0-\tilde\tau_1)\|\hat\theta_{(2)}-\theta^0_{(2)}\|_2^2\\
&\quad -2\sum_{t=\tilde\tau_1+1}^{\tau_2^0}\varepsilon_t^*(\hat\theta_{(2)}-\theta^0_{(2)})^T\\
&= \sum_{t=\tilde\tau_1+1}^{\tau_2^0}\varepsilon_t^*\varepsilon_t^{*T}+(\tau_2^0-\tau_1^0)\|\hat\theta_{(2)}-\theta^0_{(2)}\|_2^2-(\tilde\tau_1-\tau_1^0)\|\hat\theta_{(2)}-\theta^0_{(2)}\|_2^2\\
&\quad -2\sum_{t=\tau_1^0+1}^{\tau_2^0}\varepsilon_t^*(\hat\theta_{(2)}-\theta^0_{(2)})^T+2\sum_{t=\tau_1^0+1}^{\tilde\tau_1}\varepsilon_t^*(\hat\theta_{(2)}-\theta^0_{(2)})^T\\
\sum_{t=\tau_2^0+1}^{\tilde\tau_2}(y_t-\hat\theta_{(2)})(y_t-\hat\theta_{(2)})^T &= \sum_{t=\tau_2^0+1}^{\tilde\tau_2}\varepsilon_t^*\varepsilon_t^{*T}+(\tilde\tau_2-\tau_2^2)\|\hat\theta_{(2)}-\theta^0_{(3)}\|_2^2\\
&\quad -2\sum_{t=\tau_2^0+1}^{\tilde\tau_2}\varepsilon_t^*(\hat\theta_{(2)}-\theta^0_{(3)})^T\\
\sum_{t=\tilde\tau_2+1}^{T}(y_t-\hat\theta_{(3)})(y_t-\hat\theta_{(3)})^T &= \sum_{t=\tilde\tau_2+1}^{T}\varepsilon_t^*\varepsilon_t^{*T}+(T-\tilde\tau_2)\|\hat\theta_{(3)}-\theta^0_{(3)}\|_2^2
\end{aligned}
$$

$$-2\sum_{t=\tilde{\tau}_2+1}^{T}\varepsilon_t^*(\hat{\theta}_{(3)}-\theta_{(3)}^0)^T$$

Aggregating the above decompositions and substituing in (S2.13) yields,

$$\begin{aligned}
\|\tilde{\Sigma}-\hat{\Sigma}\|_\infty \;\leq\; & \max_j\|\hat{\theta}_{(j)}-\theta_{(j)}^0\|_2^2+\frac{N}{T}\max_j|\tilde{\tau}_j-\tau_j^0|\max_j\|\hat{\theta}_{(j)}-\theta_j^0\|_2^2\\
&+\frac{N}{T}\max_j|\tilde{\tau}_j-\tau_j^0|\max_j\|\hat{\theta}_{(j)}-\theta_{(j+1)}^0\|_2^2\\
&+\frac{2N}{T}\max_j\|\hat{\theta}_{(j)}-\theta_{(j)}^2\|_2\max_j\Big|\sum_{t=\tau_j^0+1}^{\tau_{j+1}^0}\varepsilon_t^T\delta\Big|\\
&+\frac{2N}{T}\max_j\|\hat{\theta}_{(j)}-\theta_{(j)}^0\|_1\max_j\Big\|\sum_{t=\tau_j^0+1}^{\tilde{\tau}_j}\varepsilon_t\Big\|_\infty\\
&+\frac{2N}{T}\max_j\|\hat{\theta}_{(j)}-\theta_{(j+1)}^0\|_1\max_j\Big\|\sum_{t=\tau_j^0+1}^{\tilde{\tau}_j}\varepsilon_t\Big\|_\infty \qquad \text{(S2.14)}
\end{aligned}$$

where δ is a unit vector. It can be verified that the final inequality (S2.14) holds for any orientation of $\tilde{\tau}$ w.r.t τ^0 allowed under the event $\mathcal{A}$, moreover also holds for any number of change points N. Substituing the bounds of (S2.11) and (S2.12) in (S2.14) and utilizing the rate assumption of Condition E(iii) it can be observed that the slowest remainder term is of order $C\{\log(p\vee T)/T\underline{\ell}\}^{1/2}$, i.e., it yields,

$$\|\tilde{\Sigma}-\hat{\Sigma}\|_\infty\leq C\Big\{\frac{\log(p\vee T)}{T\underline{\ell}}\Big\}, \qquad \text{(S2.15)}$$

on the event $\mathcal{A}$. Combining this inequality with (S2.10) via a triangle inequality yields the statement of this result, i.e.,

$$\|\tilde{\Sigma}-\Sigma\|_\infty\leq\|\tilde{\Sigma}-\hat{\Sigma}\|_\infty+\|\hat{\Sigma}-\Sigma\|_\infty\leq C\Big\{\frac{\log(p\vee T)}{T\underline{\ell}}\Big\}+C\Big\{\frac{\log(p\vee T)}{T}\Big\}$$

This completes the proof. $\square$

Proof of Lemma 1. The structure of this proof is as follows. First recall the event $\mathcal{A}$ from (4.9) and note that we have already shown that $P(\mathcal{A}) \to 1$, as part of different results of this paper; specifically, this holds by an aggregation of Theorems2 and S2.1 and (S2.7) in the Proof of Theorem 7. Next, we apply the bounds of this event $\mathcal{A}$ along with Lemma S2.1 in order to establish the desired consistency.

We begin by proving part (i), and consider for any $j = 1, ..., N$

$$\begin{aligned}
\tilde{\xi}_j^2 &= \|\hat{\theta}_{(j)} - \hat{\theta}_{(j+1)}\|_2^2 = \|(\hat{\theta}_{(j)} - \theta^0_{(j)}) - (\hat{\theta}_{(j+1)} - \theta^0_{(j+1)}) + \eta^0_{(j)}\|_2^2 \\
&\le \|\hat{\theta}_{(j)} - \theta^0_{(j)}\|_2^2 + \|\hat{\theta}_{(j+1)} - \theta^0_{(j+1)}\|_2^2 + \|\eta^0_{(j)}\|_2^2 \\
&\quad +2\|\hat{\theta}_{(j)} - \theta^0_{(j)}\|_2\|\eta^0_{(j)}\|_2 + 2\|\hat{\theta}_{(j+1)} - \theta^0_{(j+1)}\|_2\|\eta^0_{(j)}\|_2 \\
&\quad +2\|\hat{\theta}_{(j)} - \theta^0_{(j)})\|_2\|\hat{\theta}_{(j+1)} - \theta^0_{(j+1)})\|_2 \\
&\le \xi_j^2 + C\Big\{\frac{Ns\log(p\vee T)}{T\underline{\ell}}\Big\} + C\xi_j\Big\{\frac{NS\log(p\vee T)}{T\underline{\ell}}\Big\}^{1/2} \qquad \text{(S2.16)}
\end{aligned}$$

The final inequality follows by noting that by definition $\xi_j^2 = \|\eta^0_{(j)}\|_2^2$, and that $\max_j \|\hat{\theta}_{(j)} - \hat{\theta}_{(j+1)}\|_2 \le C\big\{Ns\log(p\vee T)/T\underline{\ell}\big\}^{1/2}$, by restriction on the event $\mathcal{A}$. Next, dividing both sides of (S2.16) by ξ_j^2 and utilizing the rate restriction of Condition E(iii) yields $(\tilde{\xi}_j^2/\xi_j^2) \le 1 + o(1)$. Repeating a similar argument as (S2.16) from below can be utilzed to analogously obtain $(\tilde{\xi}_j^2/\xi_j^2) \ge 1 - o(1)$. Combining both bounds from above and below yields $(\tilde{\xi}_j^2/\xi_j^2) \to 1$, which also directly implies Part (i) of this lemma.

We proceed to the proof of Part (ii). For this purpose we require the following bounds

$$(i) \quad \|\tilde{\Sigma} - \Sigma\|_\infty \le C\Big\{\frac{\log(p\vee T)}{T\underline{\ell}}\Big\}^{1/2} \qquad \text{(S2.17)}$$

$$(ii) \quad \|\hat{\eta}_{(j)} - \eta^0_{(j)}\|_2 \le C\Big\{\frac{Ns\log(p\vee T)}{T\underline{\ell}}\Big\}^{1/2}$$

$$(iii) \quad \|\hat{\eta}_{(j)}\|_2 \le C\xi_j, \quad (iv)\ \|\hat{\eta}_{(j)}\|_1 \le C\surd(s)\xi_j,$$

for all $j = 1, ..., N$. The first inequality is proved in Lemma S2.1. The second follows from the triangle inequality $\|\hat{\eta}_{(j)} - \eta^0_{(j)}\|_2 \le \|\hat{\theta}_{(j)} - \theta^0_{(j)}\|_2 + \|\hat{\theta}_{(j+1)} - \theta^0_{(j+1)}\|_2$, and then bounding individual terms directly from the restriction in event $\mathcal{A}$. The third and fourth can be obtained quite analogously by straightforward triangle inequalities and utilizing the same restriction in event $\mathcal{A}$, the sparsity of $\eta^0_{(j)}$ and the rate assumption of Condition E(iii). Next, consider the decomposition,

$$\begin{aligned}(\hat{\eta}^T_{(j)}\tilde{\Sigma}\hat{\eta}_{(j)} - \eta^{0T}_{(j)}\Sigma\eta^0_{(j)}) &= \hat{\eta}^T_{(j)}(\tilde{\Sigma}-\Sigma)\hat{\eta}_{(j)} + \eta^{0T}_{(j)}\Sigma(\hat{\eta}_{(j)} - \eta^0_{(j)}) + (\hat{\eta}^T_{(j)} - \eta^0)\Sigma\hat{\eta}_{(j)} \\ &= R1 + R2 + R3, \end{aligned} \tag{S2.18}$$

and then upper bound each of $R1, R2$ and $R3$ by utilizing (S2.17) as follows,

$$\begin{aligned} R1 &= \hat{\eta}^T_{(j)}(\tilde{\Sigma}-\Sigma)\hat{\eta}_{(j)} \le \|(\tilde{\Sigma}-\Sigma)\|_\infty \|\hat{\eta}_{(j)}\|_1^2 \le C\xi_j^2 s\Big\{\frac{\log(p\vee T)}{T\underline{\ell}}\Big\}^{1/2} \\ |R2| &\le \phi^2\|\hat{\eta}_{(j)}\|_2\|\hat{\eta}_{(j)} - \eta^0_{(j)}\|_2 \le C\phi^2\xi_j\Big\{\frac{Ns\log(p\vee T)}{T\underline{\ell}}\Big\}^{1/2} \\ |R3| &\le \phi^2\|\eta^0_{(j)}\|_2\|\hat{\eta}_{(j)} - \eta^0_{(j)}\|_2 \le C\phi^2\xi_j\Big\{\frac{Ns\log(p\vee T)}{T\underline{\ell}}\Big\}^{1/2} \end{aligned} \tag{S2.19}$$

The bound for $R1$ utilizes (i) and (iv) of (S2.17). The bounds for $R2, R3$ utilize (ii) and (iii) of (S2.17) along with the bounded maximum eigenvalue from Condition B(i). From (S2.18), we have the following upper and lower bounds,

$$\Big(\frac{\hat{\eta}^T_{(j)}\tilde{\Sigma}\hat{\eta}_{(j)}}{\eta^{0T}_{(j)}\Sigma\eta^0_{(j)}} - 1\Big) = \frac{1}{(\eta^{0T}_{(j)}\Sigma\eta^0_{(j)})}(R1 + R2 + R3)$$

$$\begin{aligned} &\leq \frac{1}{\kappa^2 \xi_j^2}(R1 + |R2| + |R3|), \quad \text{and,} \\ \Big(\frac{\hat{\eta}_{(j)}^T \tilde{\Sigma} \hat{\eta}_{(j)}}{\eta_{(j)}^{0T} \Sigma \eta_{(j)}^0} - 1\Big) &\geq \frac{1}{\phi^2 \xi_j^2}(R1 - |R2| - |R3|), \end{aligned} \tag{S2.20}$$

The first inequality follows directly by the bounded minimum eigenvalue of Condition B(i) and the second from the maximum eigenvalue of the same condition. Now substitute bounds of (S2.19) in (S2.20) and apply the rate assumption of Condition E(iii). This yields that the rhs in both inequalities converges to zero. Combining this result with Part (i) yields the statement of the Lemma. □

Proof of Corollary 2. The proof of this result is consequence of the consistency of estimates established in Lemma 1. Consider the non-vanishing regime and the corresponding two sided random walk defined in (3.6) for any given $j = 1, ..., N$, under estimated parameters, then we have the following convergence,

$$\mathcal{C}_\infty(\zeta, \tilde{\xi}_j, \tilde{\sigma}_j^2) \Rightarrow \mathcal{C}_\infty(\zeta, \xi_j, \sigma_{\infty,j}^2) = \mathcal{C}_{(\infty,j)}(\zeta)$$

This convergence follows directly from Slutsky's Theorem together with Lemma 1. The convergence of *argmax*, now follows from Lemma 3.1 of Lan et al. (2009) which is a version of the continuous mapping theorem, i.e., we have, $\arg\max_{\zeta \in \mathbb{Z}} \mathcal{C}_\infty(\zeta, \tilde{\xi}_j, \tilde{\sigma}_j^2) \Rightarrow \arg\max_{\zeta \in \mathbb{Z}} \mathcal{C}_{(\infty,j)}(\zeta)$. This in turn implies that for any given $0 < \alpha < 1$, the quantiles $\tilde{q}_\alpha^{nv}$ and q_α^{nv} obtained from either of $\arg\max_{\zeta \in \mathbb{Z}} \mathcal{C}_\infty(\zeta, \tilde{\xi}_j, \tilde{\sigma}_j^2)$ or $\arg\max_{\zeta \in \mathbb{Z}} \mathcal{C}_{(\infty,j)}(\zeta)$, are asymptotically euivalent, i.e, $(\tilde{q}_\alpha^{nv}/q_\alpha^{nv}) \to 1$. The statement of the result is now a direct consequence. Analogous arguments also provide the corresponding validity in the vanshing case as well as the case for simultaneous intervals. □

S3 Deviation bounds

Lemma S3.1. *Assume Condition A holds and let* $\bar{\varepsilon} = \sum_{t=1}^{T} \varepsilon_t / T$*. Then, for any* $c_u > 0$*, we get*

$$\|\bar{\varepsilon}\|_\infty \leq \begin{cases} \sigma\surd\{2c_u \log(p \vee T)/T\}, & \text{when } T \geq 2c_u \log(p \vee T) \\ 2c_u\sigma \log(p \vee T)/\surd T, & \text{when } T \geq 1, \end{cases} \tag{S3.1}$$

with probability at least $1 - 2\exp\{-(c_u - 1)\log(p \vee T)\}$*. Further, for any non-random* $\delta \in \mathbb{R}^p$*,* $\|\delta\|_2 = 1$*, we have,* $\surd T \delta^T \bar{\varepsilon} = O_p(1)$*. More precisely, for any* $0 < a < 1$*, choosing* $c_a = \surd(1/a)$*, we have,* $pr\big(|\surd T \delta^T \bar{\varepsilon}| > 4c_a\sigma\big) \leq a$*.*

Proof of Lemma S3.1. Applying Bernstein's inequality (Lemma S4.4) for each $k = 1, ..., p,$, we obtain

$$pr\Big(\big|\sum_{t=1}^{T} \varepsilon_{tk}\big| > dT\Big) \leq 2\exp\Big\{-\frac{T}{2}\Big(\frac{d^2}{\sigma^2} \wedge \frac{d}{\sigma}\Big)\Big\}. \tag{S3.2}$$

In the case where $T \geq 2c_u \log(p \vee T)$, select $d = \sigma\{2c_u \log(p \vee T)/T\}^{1/2}$ to get $(d^2/\sigma^2) \wedge (d/\sigma) = d^2/\sigma^2$. Substituting d in (S3.2) and applying union bounds over $k = 1, ...p$ yields the desired bound for this case. In the case where $T \geq 1$, select $d = 2c_u\sigma\{\log^2(p \vee T)/T\}^{1/2}$, and note that

$$\frac{T}{2}\Big(\frac{d^2}{\sigma^2}\Big) = 2c_u \log^2(p \vee T), \quad \text{and} \quad \frac{T}{2}\Big(\frac{d}{\sigma}\Big) \geq c_u \log(p \vee T).$$

Since in this case the latter expression is smaller, substituting this choice of d in (S3.2) and applying union bounds over $k = 1, ...p$ yields the desired bound. The second claim follows from the Markov inequality upon noting that $\surd T \delta^T \bar{\varepsilon} \sim \text{subE}(\sigma^2)$ (Lemma

S4.3) together with a second moment bound for subexponential distributions (Lemma S4.2). $\square$

Lemma S3.2. *Assume that Conditions A and B(i) hold and* ε_t^*, $t = 1, ..., T$ *be as defined in (2.1). Let* $0 \le v_T \le u_T \le 1$, *be any non-negative sequences. Then, for any* $c_u \ge 1$, *we get*

$$\sup_{\substack{\tau_j \in \mathcal{G}_j(u_T, v_T);\\ \tau_j \ge \tau_j^0}} \Big\| \sum_{t=\tau_j^0+1}^{\tau_j} \varepsilon_t^* \Big\|_\infty \le 4c_u \sigma \log(p \vee T) \sqrt{\big(Tu_T\big)}, \text{ for given } j = 1, ..., N.$$

with probability at least $1 - 4\exp\{-(c_u - 2)\log(p \vee T)\}$.

Proof of Lemma S3.2. Without loss of generality assume $v_T \ge (1/T)$ (else, the sum of interest is over an empty set of indices and trivially zero). Consider any $k \in \{1, 2, ..., p\}$ and any $\tau_j > \tau_j^0$, and apply Bernstein's inequality (Theorem S4.4) for any $d > 0$ to obtain

$$pr\Big(\big|\sum_{t=\tau_j^0+1}^{\tau_j} \varepsilon_{tk}\big| > d(\tau_j - \tau_j^0)\Big) \le 2\exp\Big\{ - \frac{(\tau_j - \tau_j^0)}{2}\Big(\frac{d^2}{\sigma^2} \wedge \frac{d}{\sigma}\Big)\Big\}. \tag{S3.3}$$

Select $d = 2c_u\sigma\{\log^2(p \vee T)/(\tau_j - \tau_j^0)\}^{1/2}$, and note that

$$\begin{aligned} (\tau_j - \tau_j^0)\frac{d^2}{2\sigma^2} &= 2c_u^2 \log^2(p \vee T), \quad \text{and,} \\ (\tau_j - \tau_j^0)\frac{d}{2\sigma} &\ge c_u \log(p \vee T), \end{aligned}$$

where we have used $(\tau_j - \tau_j^0) \ge Tv_T \ge 1$ to obtain the inequality. Substituting this choice of d in (S3.3), we obtain

$$\big|\sum_{t=\tau_j^0+1}^{\tau_j} \varepsilon_{tk}\big| \le 2c_u\sigma(\tau_j - \tau_j^0)^{1/2}\{\log^2(p \vee T)\}^{1/2} \le 2c_u\sigma\{Tu_T \log^2(p \vee T)\}^{1/2},$$

w.p. at least $1-2\exp\{-c_u\log(p\vee T)\}$. Applying union bounds over $k=1,...,p$ and T possible distinct values of τ_j, yields,

$$\sup_{\substack{\tau_j\in\mathcal{G}_j(u_T,v_T);\\ \tau_j\ge\tau_j^0}}\Big\|\sum_{t=\tau_j^0+1}^{\tau_j}\varepsilon_t\Big\|_\infty\le 2c_u\sigma\{Tu_T\log^2(p\vee T)\}^{1/2},$$

w.p. at least $1-2\exp\{-(c_u-2)\log(p\vee T)\}$. Finally, recall from (2.1) that $\varepsilon_t^*=\varepsilon_t-\bar{\varepsilon}$, consequently,

$$\begin{aligned}\sup_{\substack{\tau_j\in\mathcal{G}_j(u_T,v_T);\\ \tau_j\ge\tau_j^0}}\Big\|\sum_{t=\tau_j^0+1}^{\tau_j}\varepsilon_t^*\Big\|_\infty &\le \sup_{\substack{\tau_j\in\mathcal{G}_j(u_T,v_T);\\ \tau_j\ge\tau_j^0}}\Big\|\sum_{t=\tau_j^0+1}^{\tau_j}\varepsilon_t\Big\|_\infty+Tu_T\|\bar{\varepsilon}\|_\infty\\ &\le 2c_u\sigma\{Tu_T\log^2(p\vee T)\}^{\frac{1}{2}}+2c_u\sigma u_T\{T\log^2(p\vee T)\}^{\frac{1}{2}}\\ &\le 2c_u\sigma\{Tu_T\log^2(p\vee T)\}^{\frac{1}{2}}\big[1+\sqrt{u_T}\big]\\ &\le 4c_u\sigma\{Tu_T\log^2(p\vee T)\}^{\frac{1}{2}},\end{aligned}\tag{S3.4}$$

w.p. at least $1-2\exp\{-(c_u-2)\log(p\vee T)\}-2\exp\{-(c_u-1)\log(p\vee T)\}\ge 1-4\exp\{-(c_u-2)\log(p\vee T)\}$. The second inequality follows from Lemma S3.1 and the final inequality follows from $u_T\le 1$. This completes the proof of the lemma. □

Lemma S3.3. *Assume that Conditions A and B(i) hold and ε_t^*, $t=1,...,T$ be as defined in (2.1). Let $0\le v_T\le u_T\le 1$, be any non-negative sequences. Then, for any $c_u\ge 1$, we have*

$$(i)\ \max_{1\le j\le N}\sup_{\substack{\tau_j\in\mathcal{G}_j(u_T,v_T);\\ \tau_j\ge\tau_j^0}}\Big\|\sum_{t=\tau_j^0+1}^{\tau_j}\varepsilon_t^*\Big\|_\infty\le 4c_u\sigma\log(p\vee T)\sqrt{\big(Tu_T\big)},\tag{S3.5}$$

with probability at least $1-4\exp\{-(c_u-3)\log(p\vee T)\}$. Additionally, let $\bar{\xi}$ be as in

(1.2), then,

$$(ii)\ \max_{1\le j\le N}\sup_{\substack{\tau_j\in\mathcal{G}_j(u_T,v_T);\\ \tau_j\ge\tau_j^0}}\Big|\sum_{t=\tau_j^0+1}^{\tau_j}\varepsilon_t^{*T}\eta_{(j)}^0\Big|\le 2c_u\bar{\xi}\sigma\{Tu_T\log^2 T\}^{1/2}, \tag{S3.6}$$

with probability at least $1-4\exp\{-(c_u-2)\log T\}$.

Proof of Lemma S3.3. The first part of this lemma is a direct application of Lemma S3.2 and is obtained by supplying an additional union bound over $j=1,..,N$, and noting that $N\le T$. To establish Part (ii), we have $\varepsilon_t^T\eta_{(j)}^0\sim\text{subE}(\xi_j^2\sigma^2)$, for each $t=1,...,T$. Now proceed analogously to Lemma S3.2 by applying Bernstein's inequality to obtain for any given $j=1,...,N$ and $d>0$,

$$pr\Big(\big|\sum_{t=\tau_j^0+1}^{\tau_j}\varepsilon_t^T\eta_{(j)}^0\big|>d(\tau_j-\tau_j^0)\Big)\le 2\exp\Big\{-\frac{(\tau_j-\tau_j^0)}{2}\Big(\frac{d^2}{\xi_j^2\sigma^2}\wedge\frac{d}{\xi_j\sigma}\Big)\Big\}. \tag{S3.7}$$

Selecting $d=2c_u\xi_j\sigma\{\log^2 T/(\tau_j-\tau_j^0)\}^{1/2}$ and substituting in (S3.7), we obtain

$$\big|\sum_{t=\tau_j^0+1}^{\tau_j}\varepsilon_t^T\eta_{(j)}^0\big|\le 2c_u\xi_j\sigma\{Tu_T\log^2 T\}^{1/2},$$

w.p. at least $1-2\exp\{-c_u\log T\}$. Supplying union bounds over T possible distinct values of τ_j and over $j=1,...,N$, and that by definition $\xi_j\le\bar{\xi}$, we obtain

$$\max_{1\le j\le N}\sup_{\substack{\tau_j\in\mathcal{G}_j(u_T,v_T);\\ \tau_j\ge\tau_j^0}}\big|\sum_{t=\tau_j^0+1}^{\tau_j}\varepsilon_t^T\eta_{(j)}^0\big|\le 2c_u\bar{\xi}\sigma\{Tu_T\log^2 T\}^{1/2},$$

w.p. at least $1-2\exp\{-(c_u-2)\log T\}$. In order to obtain the analogous bound w.r.t. ε_t^*, note that $\sqrt{T}\bar{\varepsilon}^T\eta_{(j)}^0\sim\text{subE}(\xi_j^2\sigma^2)$. Again employing Bernstein's inequality together with union bounds over $j=1,...,N$, we get

$$\max_{1\le j\le N}\big|\bar{\varepsilon}^T\eta_{(j)}^0\big|\le 2c_u\bar{\xi}\sigma\{\log^2 T/T\}^{1/2}, \tag{S3.8}$$

w.p. at least $1-2\exp\{-(c_u-1)\log T\}$. Next, proceeding as in (S3.4), we obtain

$$
\begin{aligned}
\max_{1\le j\le N}\sup_{\substack{\tau_j\in\mathcal{G}_j(u_T,v_T);\\ \tau_j\ge\tau_j^0}}\Big|\sum_{t=\tau_j^0+1}^{\tau_j}\varepsilon_t^{*T}\eta_{(j)}^0\Big| &\le \max_{1\le j\le N}\sup_{\substack{\tau_j\in\mathcal{G}_j(u_T,v_T);\\ \tau_j\ge\tau_j^0}}\Big|\sum_{t=\tau_j^0+1}^{\tau_j}\varepsilon_t^{T}\eta_{(j)}^0\Big| + Tu_T\max_{1\le j\le N}\big|\bar{\varepsilon}^T\eta_{(j)}^0\big| \\
&\le 2c_u\bar{\xi}\sigma\{Tu_T\log^2 T\}^{1/2}+2c_u\bar{\xi}\sigma\{T\log^2 T\}^{1/2} \\
&\le 4c_u\bar{\xi}\sigma\{Tu_T\log^2 T\}^{1/2},
\end{aligned}
$$

w.p. at least $1-4\exp\{-(c_u-2)\log T\}$, which completes the proof of the lemma. □

Lemma S3.4. *Assume that Conditions A and B(i) hold and let* u_T, v_T *be any non-negative sequences satisfying* $0\le v_T\le u_T\le 1$. *Then, for any* $0<a<1$, *with* $c_a\ge\surd(1/a)$ *and for any given* $j=1,...,N$, *we get*

$$
\sup_{\substack{\tau_j\in\mathcal{G}_j(u_T,v_T);\\ \tau_j\ge\tau_j^0}}\Big|\sum_{t=\tau_j^0+1}^{\tau_j}\varepsilon_t^{*T}\eta_{(j)}^0\Big|\le 8c_a\sigma\xi_j\surd(Tu_T),
$$

with probability at least $1-2a$.

Proof of Lemma S3.4. This result is largely an application of Kolmogorov's inequality (Theorem S4.1). For any given $j=1,...,N$, we have

$$
\mathrm{var}(\varepsilon_t^T\eta_{(j)}^0)\le 16\xi_j^2\sigma^2
$$

where the inequality follows from Lemma S4.2. Next, note that there are at most Tu_T distinct values of τ_j in the set $\mathcal{G}_j(u_T,v_T)$. Now apply Kolmogorov's inequality (Theorem S4.1) for any $d>0$ to obtain

$$
pr\Big(\sup_{\substack{\tau_j\in\mathcal{G}_j(u_T,v_T);\\ \tau_j\ge\tau_j^0}}\Big|\sum_{t=\tau_j^0+1}^{\tau_j}\varepsilon_t^{T}\eta_{(j)}^0\Big|>d\Big)\le\frac{Tu_T}{d^2}16\sigma^2\xi_j^2.
$$

Selecting $d = 4c_a\sigma\xi_j\surd(Tu_T)$ with $c_a \geq \surd(1/a)$ yields

$$\sup_{\substack{\tau_j\in\mathcal{G}_j(u_T,v_T);\\ \tau_j\geq\tau_j^0}}\Big|\sum_{t=\tau_j^0+1}^{\tau_j}\varepsilon_t^T\eta_{(j)}^0\Big| \leq 4c_a\sigma\xi_j\surd(Tu_T),$$

w.p. at least $1-a$. The analogous bound w.r.t ε_t^* can be obtained as

$$\begin{aligned}\sup_{\substack{\tau_j\in\mathcal{G}_j(u_T,v_T);\\ \tau_j\geq\tau_j^0}}\Big|\sum_{t=\tau_j^0+1}^{\tau_j}\varepsilon_t^{*T}\eta_{(j)}^0\Big| &\leq \sup_{\substack{\tau_j\in\mathcal{G}_j(u_T,v_T);\\ \tau_j\geq\tau_j^0}}\Big|\sum_{t=\tau_j^0+1}^{\tau_j}\varepsilon_t^T\eta_{(j)}^0\Big| + Tu_T|\bar\varepsilon^T\eta_{(j)}^0|\\ &\leq 4c_a\sigma\xi_j\surd(Tu_T) + 4c_a\sigma\xi_j u_T\surd T\\ &\leq 8c_a\sigma\xi_j\surd(Tu_T),\end{aligned}$$

w.p. at least $1-2a$. Here the second inequality follows from (S3.9) together with the second claim of Lemma S3.1. This completes the proof. $\square$

Lemma S3.5. *Assume Conditions A and B(i) hold and that $T\underline{\ell} \geq \log(p\vee T)$. Then, for any $c_u, c_{u1} > 0$, we have*

$$\max_{1\leq j\leq N+1}\sup_{\substack{\tau_{j-1},\tau_j\in\{1,....,T-1\};\\ (\tau_j-\tau_{j-1})\geq c_uT\underline{\ell}}}\frac{1}{(\tau_{(j)}-\tau_{(j-1)})}\Big\|\sum_{t=\tau_{j-1}+1}^{\tau_j}\varepsilon_t^*\Big\|_\infty \leq 2\sigma\Big\{\frac{2c_{u1}\log(p\vee T)}{c_uT\underline{\ell}}\Big\}^{\frac{1}{2}}$$

with probability at least $1-4\exp\big\{-(c_{u2}-4)\log(p\vee T)\big\}$, *where* $c_{u2} = c_{u1}\wedge\surd(c_uc_{u1}/2)$.

Proof of Lemma S3.5. For any given $j = 1,...,N$ consider any $\tau_{j-1} < \tau_j \in \{1,...,T\}$ satisfying $(\tau_j - \tau_{j-1}) \geq c_uT\underline{\ell}$, and any $k \in \{1,...,p\}$. Then applying the Bernstein's inequality (Lemma S4.4) for any $d > 0$, we obtain,

$$pr\Big(\Big|\sum_{t=\tau_{j-1}+1}^{\tau_j}\varepsilon_{tk}\Big| > d(\tau_j-\tau_{j-1})\Big) \leq$$

$$2\exp\Big\{-\frac{(\tau_j-\tau_{(j-1)})}{2}\Big(\frac{d^2}{\sigma^2}\wedge\frac{d}{\sigma}\Big)\Big\}. \tag{S3.9}$$

Choose $d=\sigma\Big\{2c_{u1}\log(p\vee T)\big/\big(\tau_{(j)}-\tau_{(j-1)}\big)\Big\}^{1/2}$, then, we have,

$$\begin{aligned}\big(\tau_j-\tau_{j-1}\big)\frac{d^2}{2\sigma^2} &= c_{u1}\log(p\vee T),\quad \text{and,}\\ \big(\tau_j-\tau_{j-1}\big)\frac{d}{2\sigma} &\geq \surd(c_{u1}/2)(c_uT\underline{\ell})^{1/2}\{\log(p\vee T)\}^{1/2}\\ &\geq \surd(c_uc_{u1}/2)\log(p\vee T).\end{aligned}$$

The first inequality follows since by choice $\big(\tau_{j+1}-\tau_{j-1}\big)\geq c_uT\underline{\ell}$, and the second inequality follows by assumption $T\underline{\ell}\geq\log(p\vee T)$. Substituting this choice of d in (S3.9), we obtain,

$$\begin{aligned}\frac{1}{\big(\tau_j-\tau_{j-1}\big)}\Big|\sum_{t=\tau_{j-1}+1}^{\tau_j}\varepsilon_{tk}\Big| &\leq \sigma\Big\{2c_{u1}\log(p\vee T)\big/\big(\tau_j-\tau_{j-1}\big)\Big\}^{1/2}\\ &\leq \sigma\Big\{\frac{2c_{u1}\log(p\vee T)}{c_uT\underline{\ell}}\Big\}^{1/2}\end{aligned}$$

with probability at least $1-2\exp\{-c_{u2}\log(p\vee T)\}$, where $c_{u2}=c_{u1}\wedge\surd(c_uc_{u1}/2)$. Applying union bounds over $k=1,...,p$, the upper bound T^2 of at most distinct combinations of τ_{j-1} and τ_{j+1}, and then over $j=1,...,N+1$, $(N\leq T)$ yields,

$$\max_{1\leq j\leq N+1}\sup_{\substack{\tau_{j-1},\tau_j\in\{1,....,T-1\};\\ (\tau_j-\tau_{j-1})\geq c_uT\underline{\ell}}}\frac{1}{\big(\tau_j-\tau_{j-1}\big)}\Big|\sum_{t=\tau_{j-1}+1}^{\tau_{j+1}}\varepsilon_{tk}\Big|\leq\sigma\Big\{\frac{2c_{u1}\log(p\vee T)}{c_uT\underline{\ell}}\Big\}^{1/2},$$

w.p. at least $1-2\exp\big\{-(c_{u2}-4)\log(p\vee T)\big\}$. Finally utilizing the form $\varepsilon_t^*=\varepsilon_t-\bar{\varepsilon}$, $t=1,...,T$, together with the first bound for $\|\bar{\varepsilon}\|_\infty$ of Lemma S3.1 by an argument analogous to that in (S3.4) yields the statement of the lemma. $\square$

S4 Definitions and auxiliary results

The following definitions and results provide basic properties of subexponential distributions. These are largely reproduced from Vershynin (2019) and Rigollet (2015). Theorem S4.1 and S4.2 below reproduce Kolmogorov's inequality and the Argmax Theorem.

Definition S4.1. [Subexponential r.v.] A random variable $X \in \mathbb{R}$ is said to be subexponential with parameter $\sigma^2 > 0$ $\big($denoted by $X \sim \mathrm{subE}(\sigma^2)\big)$ if $E(X) = 0$ and its moment generating function

$$E(\mathrm{e}^{tX}) \leq \mathrm{e}^{t^2\sigma^2/2}, \qquad \forall\ |t| \leq \frac{1}{\sigma}$$

Definition S4.2. A random vector $X \in \mathbb{R}^p$ is subexponential with parameter σ^2, if the inner product $\langle X, v\rangle \sim \mathrm{subE}(\sigma^2)$, respectively, for any $v \in \mathbb{R}^p$ with $\|v\|_2 = 1$.

Following is the elementary definition of uniform tightness of a sequence of random variables reproduced from Page 166, Chapter 2 of Durrett (2010).

Definition S4.3. A sequence of random variables X_n is said to be uniformly tight if for every $\epsilon > 0$, there is a compact set K such that $pr(X_n \in K) > 1 - \epsilon$.

Lemma S4.1. *[Tail bounds] If $X \sim \mathrm{subE}(\sigma^2)$, then*

$$pr(|X| \geq \lambda) \leq 2\exp\Big\{-\frac{1}{2}\Big(\frac{\lambda^2}{\sigma^2} \wedge \frac{\lambda}{\sigma}\Big)\Big\}.$$

Lemma S4.2 (Moment bounds)**.** *If $X \sim \mathrm{subE}(\sigma^2)$, then*

$$E|X|^k \leq 4\sigma^k k^k, \qquad k > 0.$$

Lemma S4.3. *Assume that* $X \sim \mathrm{subE}(\sigma^2)$, *and that* $\alpha \in \mathbb{R}$, *then* $\alpha X \sim \mathrm{subE}(\alpha^2\sigma^2)$. *Moreover, assume that* $X_1 \sim \mathrm{subE}(\sigma_1^2)$ *and* $X_2 \sim \mathrm{subE}(\sigma_2^2)$, *then* $X_1 + X_2 \sim \mathrm{subE}((\sigma_1 + \sigma_2)^2)$, *additionally, if* X_1 *and* X_2 *are independent, then* $X_1 + X_2 \sim \mathrm{subE}(\sigma_1^2 + \sigma_2^2)$.

Lemma S4.4 (Bernstein's inequality). *Let* $X_1, X_2, ..., X_T$ *be independent random variables such that* $X_t \sim \mathrm{subE}(\lambda^2)$. *Then for any* $d > 0$ *we have,*

$$pr(|\bar{X}| > d) \leq 2 \exp\Big\{ -\frac{T}{2}\Big(\frac{d^2}{\lambda^2} \wedge \frac{d}{\lambda}\Big)\Big\}$$

The next result is Kolmogorov's inequality reproduced from Hájek and Rényi (1955)

Theorem S4.1 (Kolmogorov's inequality). *If* $\xi_1, \xi_2, ...$ *is a sequence of mutually independent random variables with mean values* $E(\xi_k) = 0$ *and finite variance* $\mathrm{var}(\xi_k) = D_k^2$ $(k = 1, 2, ...)$, *we have, for any* $\varepsilon > 0$,

$$pr\Big(\max_{1\leq k\leq m} \big|\xi_1 + \xi_2 + ... + \xi_k\big| > \varepsilon\Big) \leq \frac{1}{\varepsilon^2}\sum_{k=1}^{m} D_k^2$$

Next, we provide the Argmax Theorem reproduced from Theorem 3.2.2 of Vaart and Wellner (1996).

Theorem S4.2 (Argmax Theorem). *Let* $\mathcal{M}_n, \mathcal{M}$ *be stochastic processes indexed by a metric space* H *such that* $\mathcal{M}_n \Rightarrow \mathcal{M}$ *in* $\ell^\infty(K)$ *for every compact set* $K \subseteq H$. *Suppose that almost all sample paths* $h \to \mathcal{M}(h)$ *are upper semicontinuous and posses a unique maximum at a (random) point* $\hat{h}$, *which as a random map in* H *is tight. If the sequence* $\hat{h}_n$ *is uniformly tight and satisfies* $\mathcal{M}_n(\hat{h}_n) \geq \sup_h \mathcal{M}_n(h) - o_p(1)$, *then* $\hat{h}_n \Rightarrow \hat{h}$ *in* H.

Theorem S4.3. *Let* $X_1, ..., X_n$ *be independent marginally sub-Weibull random vectors* *in* $\mathbb{R}^p$ *satisfying*

$$\max_{1\le i\le n}\|X_i\|_{M,\psi_n}\le K_{n,p}<\infty,\quad \text{forsome}\quad 0<\alpha\le 2.$$

Fix $n,p\ge 1$. *Then for any* $t\ge 0$, *with probability at least* $1-3\exp(-t)$,

$$\|\hat{\Sigma}_n-\Sigma_n\|_\infty\le 7A_{n,p}\sqrt{\frac{t+2\log p}{n}}+\frac{C_\alpha K_{n,p}^2\big(\log(2n)\big)^{2/\alpha}(t+2\log p)^{2/\alpha}}{n},$$

where $C_\alpha>0$ *is a constant depending only on* α, *and* $A_{n,p}^2$ *is given by,*

$$A_{n,p}^2=\max_{1\le j\le k\le p}\frac{1}{n}\sum_{i=1}^n\mathrm{Var}\big(X_i(j)X_i(k)\big)\tag{S4.1}$$

S5 Additional details and numerical results

S5.1 Estimation of drifts, asymptotic variances and quantiles

Next, we provide a discussion on the estimation of ξ_j, and $\sigma^2_{(\infty,j)}$, $j=1,...,N$, employed to obtain confidence intervals for $\tau^0=(\tau^0_1,...,\tau^0_N)^T$, using the results of Theorems 3, 4 and 5.

First, to alleviate finite sample regularization biases we employ refitted mean estimates computed as $\tilde{\theta}_{(j)}=\big[\bar{x}_{(j)}(\tilde{\tau})\big]_{\hat{S}_j}$ $j=1,...,N$ wherein $\tilde{\tau}$ is the change point estimate of Algorithm 1. Here $\hat{S}_j=\{k\ \hat{\theta}_{(j)k}\ne 0\}$, $j=1,...,N$ correspond to the estimated sparsity sets, where $\hat{\theta}_{(j)}$, $j=1,...,N$ are the Step 2 mean estimates of Algorithm 1. All remaining indices of these mean estimates are set to zero. It is known that refitted mean estimates preserve the rate of convergence of the regularized version while reducing finite sample biases, e.g. Belloni et al. (2011). The jump vectors and jump sizes $\tilde{\eta}_{(j)}$ and $\tilde{\xi}_j$, $j = 1, ..., N$, are then evaluated as plug-in estimates per the defining relations (1.2).

Next, consider the asymptotic variances $\sigma^2_{(\infty,j)}$, $j = 1, ..., N$ of Condition D. Note the finite sample representation of this parameter, $\xi_j^{-2}\eta^{0T}_{(j)}\Sigma\eta^0_{(j)}$. Plug-in versions $\tilde{\sigma}^2_{(\infty,j)}$, $j = 1, ..., N$, are computed by employing the above described estimated parameters. The covariance matrix Σ is estimated as the sample covariance $\tilde{\Sigma}$ computed by utilizing the entire data set centered with the estimated mean parameters $\tilde{\theta}_{(j)}$, $j = 1, ..., N$ over estimated partitions induced by $\tilde{\tau}$ of Algorithm 1. Note that since we are not interested in the estimation of Σ itself, but instead the quadratic form described above, employing the sample covariance is effectively identical to employing the refitted covariance on the adjacency matrix estimated by the jump vectors $\tilde{\eta}'_{(j)}s$, in turn making this shortcut valid despite potential high dimensionality.

Finally, for quantiles of the limiting distributions characterized in Theorems 3 and 4 in the vanishing and non-vanishing regimes, respectively, we note the following: in the former case, we employ the cdf of this distribution which was first presented in Yao (1987). In the latter case, we assume in all calculations that the underlying distribution is Gaussian and consequently the distribution of the increments $\mathcal{P}$ of Condition A$'$ is also Gaussian. The above estimated parameters are then used to produce realizations of the increments' distribution, and thus realizations of the two-sided random walk and in turn those of its *argmax*. The quantiles are then estimated by a Monte Carlo approximation.

S5.2 Additional numerical results of Section 5

Results of Scenarios A an B (Gaussian errors): Tables 1 and 2 below provide results for these scenarios for $N = 4$ change points, respectively.

$N=4$, $s=4$		haus.d (sd)	Comp. coverage (av. ME) $(1-\alpha)=0.95$		Simul. Coverage $(1-\alpha)^N = 0.814$
T	p		Vanishing	Non-Vanishing	
450	50	1.31 (1.28)	0.964 (2.08)	0.978 (2.02)	0.794
450	200	1.31 (1.15)	0.936 (2.07)	0.958 (2)	0.79
450	350	1.33 (1.42)	0.946 (2.07)	0.978 (2.01)	0.798
450	500	1.35 (1.37)	0.954 (2.12)	0.976 (2.06)	0.79
600	50	1.41 (1.32)	0.96 (2.12)	0.976 (2.03)	0.79
600	200	1.31 (1.3)	0.934 (2.08)	0.96 (2.02)	0.808
600	350	1.38 (1.17)	0.952 (2.05)	0.978 (1.99)	0.798
600	500	1.36 (1.19)	0.94 (2.05)	0.97 (1.98)	0.772
750	50	1.39 (1.22)	0.958 (2.12)	0.972 (2.02)	0.786
750	200	1.33 (1.28)	0.946 (2.09)	0.962 (2.02)	0.788
750	350	1.41 (1.41)	0.95 (2.09)	0.968 (2)	0.768
750	500	1.42 (1.38)	0.954 (2.08)	0.972 (2.02)	0.774

Table 1: Results of Scenario A with $N = 4$ based on 500 replicates. Coverage metrics rounded to three decimals, all other metrics rounded to two decimals.

Method	$N=4$, $s=4$ T	p	haus.d (sd)	N-match	Comp. cov. (av. ME) $\vert\hat{N}=N$ $(1-\alpha)=0.95$ Vanishing	Non-Vanishing	Simul. cov. $\vert\hat{N}=N$ $(1-\alpha)^N=0.814$
KFJS+ BS+ LR	450	50	2.81 (6.46)	0.95	0.951 (2.08)	0.975 (1.99)	0.753
	450	200	5.09 (10.6)	0.88	0.939 (2.08)	0.964 (2.02)	0.722
	450	350	4.15 (9.57)	0.90	0.924 (2.09)	0.955 (2.02)	0.753
	450	500	4.27 (9.33)	0.89	0.921 (2.11)	0.946 (2.05)	0.742
	600	50	3.79 (9.86)	0.94	0.932 (2.11)	0.951 (2.02)	0.756
	600	200	4.86 (11.71)	0.92	0.954 (2.07)	0.963 (2.00)	0.771
	600	350	5.19 (12.6)	0.92	0.948 (2.10)	0.969 (2.03)	0.734
	600	500	4.57 (11.4)	0.92	0.939 (2.07)	0.961 (2.02)	0.747
	750	50	6.18 (16.19)	0.91	0.956 (2.14)	0.978 (2.04)	0.776
	750	200	7.10 (17.78)	0.90	0.942 (2.10)	0.969 (2.02)	0.752
	750	350	6.78 (16.79)	0.90	0.951 (2.11)	0.976 (2.03)	0.734
	750	500	6.16 (16.14)	0.91	0.941 (2.10)	0.956 (2.03)	0.742
WS +LR	450	50	12.54 (17.97)	0.64	0.934 (2.06)	0.959 (1.99)	0.723
	450	200	14.16 (22.25)	0.64	0.938 (2.07)	0.959 (1.99)	0.716
	450	350	24.83 (53.64)	0.58	0.911 (2.08)	0.955 (2.01)	0.712
	450	500	84.14 (116.91)	0.49	0.918 (2.07)	0.951 (2.00)	0.694
	600	50	15.77 (23.71)	0.62	0.926 (2.10)	0.945 (2.00)	0.746
	600	200	18.55 (25.95)	0.62	0.942 (2.04)	0.971 (1.97)	0.759
	600	350	18.09 (29.49)	0.64	0.928 (2.09)	0.950 (2.00)	0.704
	600	500	30.79 (68.63)	0.60	0.957 (2.06)	0.973 (2.01)	0.739
	750	50	20.57 (30.58)	0.60	0.94 (2.15)	0.967 (2.05)	0.779
	750	200	20.14 (29.03)	0.62	0.935 (2.09)	0.955 (2.00)	0.718
	750	350	19.38 (29.51)	0.66	0.961 (2.10)	0.976 (2.02)	0.721
	750	500	23.25 (38.99)	0.65	0.923 (2.09)	0.944 (2.02)	0.728

Table 2: Results of Scenario B with $N=4$ based on 500 replicates. Coverage metrics rounded to three decimals, all other metrics rounded to two decimals.

Results of Scenarios A$'$ and B$'$ (subexponential errors): Tables 3 and 5 below provide results for these scenarios with $N = 2$ change points, and Tables 4 and 6 with $N = 4$.. Under Scenario B$'$ with subexponential errors, the method WS for preliminary estimation was found to have a very low proportion of replicates wherein $\hat{N} = N$. Hence, it was rendered unsuitable for calculating the coverage metrics. Consequently in Scenario B$'$ we only report results obtained by the KFJS+BS+LR method.

$N = 2$, $s = 4$		haus.d (sd)	Comp. coverage (av. ME) $(1-\alpha) = 0.95$		Simul. Coverage
T	p		Vanishing	Non-Vanishing	$(1-\alpha)^N = 0.902$
450	50	0.79 (1.00)	0.962 (2.16)	0.982 (2.05)	0.888
450	200	0.76 (1.10)	0.956 (2.15)	0.968 (2.05)	0.894
450	350	0.75 (1.02)	0.950 (2.12)	0.964 (2.05)	0.882
450	500	0.79 (1.06)	0.954 (2.11)	0.968 (2.02)	0.878
600	50	0.71 (1.11)	0.958 (2.18)	0.964 (2.05)	0.908
600	200	0.81 (1.13)	0.940 (2.15)	0.952 (2.03)	0.886
600	350	0.84 (1.03)	0.946 (2.16)	0.964 (2.04)	0.862
600	500	0.70 (0.96)	0.948 (2.14)	0.972 (2.03)	0.898
750	50	0.74 (0.99)	0.958 (2.17)	0.966 (2.03)	0.888
750	200	0.81 (1.07)	0.954 (2.17)	0.962 (2.03)	0.864
750	350	0.70 (0.97)	0.956 (2.16)	0.968 (2.02)	0.888
750	500	0.72 (0.92)	0.962 (2.17)	0.974 (2.03)	0.898

Table 3: Results of Scenario A$'$ with $N = 2$ based on 500 replicates. Coverage metrics rounded to three decimals, all other metrics rounded to two decimals.

$N = 4$, $s = 4$		haus.d (sd)	Comp. coverage (av. ME) $(1-\alpha) = 0.95$		Simul. Coverage $(1-\alpha)^N = 0.814$
T	p		Vanishing	Non-Vanishing	
450	50	1.28 (1.33)	0.962 (2.08)	0.972 (2.00)	0.808
450	200	1.34 (1.23)	0.958 (2.06)	0.976 (1.98)	0.788
450	350	1.31 (1.31)	0.958 (2.08)	0.968 (2.00)	0.796
450	500	1.37 (1.15)	0.946 (2.09)	0.964 (2.02)	0.770
600	50	1.32 (1.24)	0.958 (2.11)	0.980 (2.04)	0.798
600	200	1.26 (1.20)	0.940 (2.06)	0.958 (1.98)	0.798
600	350	1.40 (1.30)	0.956 (2.07)	0.982 (2.00)	0.768
600	500	1.44 (1.47)	0.956 (2.06)	0.974 (1.99)	0.770
750	50	1.29 (1.25)	0.940 (2.14)	0.952 (2.04)	0.814
750	200	1.36 (1.31)	0.946 (2.09)	0.962 (2.02)	0.772
750	350	1.29 (1.25)	0.956 (2.09)	0.970 (2.00)	0.806
750	500	1.46 (1.35)	0.950 (2.10)	0.972 (2.03)	0.784

Table 4: Results of Scenario A$'$ with $N = 4$ based on 500 replicates. Coverage metrics rounded to three decimals, all other metrics rounded to two decimals.

Method	$N=2$, $s=4$		haus.d (sd)	N-match	Comp. coverage (av. ME) $(1-\alpha)=0.95$		Simul. Coverage $(1-\alpha)^N=0.902$
	T	p			Vanishing	Non-Vanishing	
KFJS+ BS+ LR	450	50	18.14 (25.7)	0.65	0.957 (2.16)	0.957 (2.12)	0.877
	450	200	18.96 (26.59)	0.66	0.955 (2.16)	0.970 (2.12)	0.882
	450	350	17.16 (26.31)	0.70	0.940 (2.18)	0.974 (2.11)	0.883
	450	500	16.66 (26.04)	0.71	0.924 (2.21)	0.941 (2.18)	0.876
	600	50	25.75 (34.03)	0.60	0.940 (2.18)	0.957 (2.10)	0.881
	600	200	26.55 (35.15)	0.62	0.913 (2.14)	0.926 (2.06)	0.859
	600	350	25.90 (35.85)	0.65	0.938 (2.18)	0.969 (2.12)	0.864
	600	500	23.78 (34.47)	0.66	0.952 (2.16)	0.967 (2.08)	0.891
	750	50	24.18 (39.17)	0.70	0.936 (2.21)	0.954 (2.09)	0.858
	750	200	31.50 (42.29)	0.62	0.964 (2.18)	0.964 (2.09)	0.877
	750	350	33.38 (44.07)	0.61	0.944 (2.17)	0.958 (2.06)	0.869
	750	500	33.99 (45.38)	0.62	0.929 (2.18)	0.936 (2.09)	0.878

Table 5: Results of Scenario B′ with $N = 2$ based on 500 replicates. Coverage metrics rounded to three decimals, all other metrics rounded to two decimals.

Method	$N=4$, $s=4$ T	p	haus.d (sd)	N-match	Comp. coverage (av. ME) $(1-\alpha)=0.95$ Vanishing	Non-Vanishing	Simul. Coverage $(1-\alpha)^N=0.814$
KFJS+ BS+ LR	450	50	3.57 (8.71)	0.93	0.961 (2.09)	0.970 (2.08)	0.773
	450	200	4.32 (9.75)	0.91	0.928 (2.08)	0.952 (2.08)	0.740
	450	350	3.86 (8.63)	0.90	0.934 (2.11)	0.962 (2.09)	0.723
	450	500	4.01 (9.21)	0.91	0.940 (2.14)	0.962 (2.14)	0.753
	600	50	4.42 (11.27)	0.93	0.950 (2.11)	0.974 (2.07)	0.767
	600	200	4.92 (12.32)	0.92	0.939 (2.08)	0.961 (2.05)	0.762
	600	350	5.51 (13.30)	0.91	0.945 (2.08)	0.967 (2.06)	0.750
	600	500	4.50 (11.71)	0.93	0.940 (2.09)	0.970 (2.08)	0.755
	750	50	5.64 (14.81)	0.92	0.948 (2.14)	0.952 (2.08)	0.754
	750	200	5.55 (14.99)	0.93	0.942 (2.09)	0.952 (2.05)	0.769
	750	350	5.47 (14.83)	0.93	0.959 (2.09)	0.974 (2.05)	0.754
	750	500	6.26 (16.56)	0.91	0.949 (2.08)	0.967 (2.04)	0.786

Table 6: Results of Scenario B′ with $N = 4$ based on 500 replicates. Coverage metrics rounded to three decimals, all other metrics rounded to two decimals.

Setup and results of Scenario C: The design of this simulation is largely identical to that of Scenario B (Gaussian errors) described in Section 5, with the only distinction being that we consider larger values of the sampling period $T \in \{1000, 1500\}$. Results are provided in Table 7 below.

Method	$N = 4$, $s = 4$ T	p	haus.d (sd)	N-match	Comp. coverage (av. ME) $(1-\alpha) = 0.95$ Vanishing	Non-Vanishing	Simul. Coverage $(1-\alpha)^N = 0.814$
KFJS+ BS+ LR	1000	200	8.01 (21.17)	0.90	0.967 (2.12)	0.978 (2.01)	0.792
	1000	350	9.85 (23.37)	0.87	0.945 (2.12)	0.956 (2.02)	0.793
	1000	500	8.91 (22.70)	0.89	0.935 (2.11)	0.957 (2.02)	0.763
	1500	200	12.85 (34.33)	0.89	0.957 (2.18)	0.969 (2.02)	0.804
	1500	350	13.78 (35.74)	0.88	0.957 (2.17)	0.973 (2.02)	0.758
	1500	500	11.77 (32.69)	0.89	0.955 (2.16)	0.971 (2.01)	0.786
WS+ LR	1000	200	24.70 (41.09)	0.66	0.964 (2.12)	0.976 (2.00)	0.781
	1000	350	24.99 (38.80)	0.63	0.937 (2.13)	0.947 (2.03)	0.767
	1000	500	21.60 (37.41)	0.71	0.949 (2.11)	0.966 (2.02)	0.764
	1500	200	34.72 (61.29)	0.67	0.941 (2.17)	0.962 (2.01)	0.776
	1500	350	35.65 (61.65)	0.68	0.959 (2.17)	0.974 (2.01)	0.752
	1500	500	37.09 (63.98)	0.68	0.950 (2.16)	0.962 (2.01)	0.776

Table 7: Results of Scenario C with $N = 4$ based on 500 replicates. Coverage metrics rounded to three decimals, all other metrics rounded to two decimals.

S5.3 Description of second method (KFJS+BS) employed for preliminary estimation in Section 5

Kaul et al. (2021) considers a mean shift model with a single change point under potential high dimensionality, i.e., model (1.1) with $N = 1$. They propose a two step algorithmic procedure which yields an estimate that is optimal is its rate of convergence (this estimate is the same to $\tilde{\tau}$ in Algorithm 1 in Section 3 for $N = 1$). While not of direct interest, the paper also establishes that the first update in their algorithm is near optimal, i.e., obeys the bound (4.5) with $N = 1$, under identical assumptions as those assumed here, including the relaxation to subexponential distributions. Remark 4.2 of the paper provides an ℓ_0 regularization that also enables boundary selection of the change point estimate, i.e., identifying that a change point *is not present*. This estimator is compiled as Algorithm 2 below.

Algorithm 2 (KFJS): Near optimal estimation of τ^0 with boundary selection (under $N = 1$)

(Initialize): Select a preliminary evenly spaced coarse grid $\mathcal{D} \subset \{1, ..., T\}$ of cardinality $\log T$. Select an initializer $\check{\tau} \in \mathcal{D}$ as the best fitting value to the data $\{x_t\}_{t=1}^{T}$.

Step 1: Obtain mean estimates $\check{\theta}_{(j)} = \hat{\theta}_{(j)}(\check{\tau})$, $j = 1, 2$, and update change point estimates as

$$\hat{\tau} = \underset{\tau \in \{1,...,(T-1)\}}{\arg\min} Q(\tau, \check{\theta}),$$

and perform an ℓ_0 regularization as

$$\hat{\tau}^* = \begin{cases} T(\text{no change}) & \text{if } \{Q(T, \check{\theta}) - Q(\hat{\tau}, \check{\theta})\} < \gamma \\ \hat{\tau} & \text{else.} \end{cases}$$

(Output): $\hat{\tau}^*$

The mean estimates $\hat{\theta}(\tau)$ of Algorithm 2 are the soft-thresholded sample means as defined in (4.3), and $Q(\tau, \theta)$ represents the squared loss under a single change point assumption defined as

$$Q(\tau,\theta)=\sum_{t=1}^{\tau}\|x_t-\theta_{(1)}\|_2^2+\sum_{t=\tau+1}^{T}\|x_t-\theta_{(2)}\|_2^2.$$

It can be observed that the regularization carried out in Step 1 of Algorithm 2 is equivalent to

$$\hat\tau^*=\operatorname*{arg\,min}_{\tau\in\{1,\ldots,(T-1)\}}\big\{Q(\tau,\check\theta)+\gamma\mathbf{1}[\tau\neq T]\big\}.$$

Further, the BIC criterion to tune this regularization reduces to $\gamma=(|\hat S_2|+1)\log T$. Note that at the boundary value $\tau=T$, the model has $|S_2|$ fewer mean parameters and one less change point parameter. It can be shown that in addition to near optimal estimation of τ^0, Algorithm 2 also provides selection consistency, i.e., $pr(\hat\tau^*=T)\to 1$ when $\tau^0=T$. A natural extension of Algorithm 2 is employed in Section 5 by leveraging binary segmentation, i.e., recursive application of Algorithm 2 on estimated partitions, performed until no further change points are detected. This extension is summarized in Algorithm 3.

Bibliography

Belloni, A., V. Chernozhukov, and L. Wang (2011). Square-root lasso: pivotal recovery of sparse signals via conic programming. *Biometrika 98*(4), 791–806.

Durrett, R. (2010). *Probability: theory and examples.* Cambridge university press.

Algorithm 3 (KJFS+BS): Extension of KJFS to multiple changes via binary segmentation

(Initialize): $\hat{\tau}_{\text{st}} = \phi$ collecting all change points to be estimated.

Implement $\hat{\tau}=$ Alg. 2 $\big(\{1, ..., T\}\big)$.

If $\hat{\tau} = T$ (no change) then **Stop**

Else $\hat{\tau}_{\text{up}} = (\tau_{\text{st}}, \hat{\tau})$ (updated vector of estimated change points)

While length$(\hat{\tau}_{up}) >$ length$(\hat{\tau}_{st})$ **do**

$\hat{\tau}_{st} = \hat{\tau}_{up}$

for $m \in 1 : (\text{length}(\tau_{st}) + 1)$ **do**

$\text{partition}_m = \{\tau_{st(m-1)}, ..., \tau_{st(m)}\}$

$\hat{\tau} = \text{Alg. } 2(\text{partition}_{\text{m}})$

If $\hat{\tau}$ is away from boundary of sampling period of partition **then**

$\hat{\tau}_{up} = (\hat{\tau}_{st}, \hat{\tau})$

(Output): all estimated change points of vector $\hat{\tau}_{up}$ sorted in ascending order.

Hájek, J. and A. Rényi (1955). Generalization of an inequality of kolmogorov. *Acta Mathematica Hungarica 6*(3-4), 281–283.

Kaul, A., O. Davidov, and S. D. Peddada (2017). Structural zeros in high-dimensional data with applications to microbiome studies. *Biostatistics 18*(3), 422–433.

Kaul, A., S. B. Fotopoulos, V. K. Jandhyala, A. Safikhani, et al. (2021). Inference on the change point under a high dimensional sparse mean shift. *Electronic Journal of Statistics 15*(1), 71–134.

Kuchibhotla, A. K. and A. Chakrabortty (2022, jun). Moving beyond sub-gaussianity in high-dimensional statistics: applications in covariance estimation and linear regression. *Information and Inference: A Journal of the IMA 11*(4), 1389–1456.

Lan, Y., M. Banerjee, and G. Michailidis (2009). Change-point estimation under adaptive sampling. *The Annals of Statistics 37*(4), 1752–1791.

Rigollet, P. (2015). 18. s997: High dimensional statistics. *Lecture Notes), Cambridge, MA, USA: MIT Open-CourseWare.*

Seijo, E. and B. Sen (2011). A continuous mapping theorem for the smallest argmax functional. *Electronic Journal of Statistics 5*, 421–439.

Vaart, A. W. and J. A. Wellner (1996). *Weak convergence and empirical processes: with applications to statistics*. Springer.

Vershynin, R. (2019). *High-Dimensional Probability.* Cambridge, UK: Cambridge University Press.

Yao, Y.-C. (1987). Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. *The Annals of Statistics 15*(3), 1321–1328.

Yuan, M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *Journal of Machine Learning Research 11*(Aug), 2261–2286.