LOCALIZING MULTIVARIATE CAVIAR

Supplementary Material

Without loss of generality in Sections S1–S3 we assume that the interval of interest is the whole observed data set, i.e. $\mathcal{I} = [1, T]$. For this reason we neglect the index "$\mathcal{I}$" where applies, for instance, $L(\widetilde{\boldsymbol{\theta}})$ instead of $L_{\mathcal{I}}(\widetilde{\boldsymbol{\theta}}_{\mathcal{I}})$.

S1 Proof of Lemma 2.1

Denote,

$$\tilde{\mathbf{g}}_t(\boldsymbol{\theta}) = \mathbf{g}_t(\boldsymbol{\theta}) - \sum_i \nabla q_{it}(\boldsymbol{\theta}^*) \mathbf{1}^c[Y_{it} \le q_{it}(\boldsymbol{\theta})],$$

where for $\mathcal{F}_{t-1}$–measurable Z we set $\mathbf{1}^c[Y_{it} \le Z] = \mathbf{1}[Y_{it} \le Z] - \mathsf{P}(Y_{it} \le Z \mid \mathcal{F}_{t-1})$. Since $q_{it}(\boldsymbol{\theta})$ are $\mathcal{F}_{t-1}$–measurable, we obviously have $\mathsf{E}\tilde{\mathbf{g}}_t(\boldsymbol{\theta}) = \boldsymbol{\lambda}_t(\boldsymbol{\theta})$. For any two $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta$ consider the decomposition,

$$\begin{aligned} \mathbf{g}_t(\boldsymbol{\theta}) - \mathbf{g}_t(\boldsymbol{\theta}') = & \sum_i \{\nabla q_{it}(\boldsymbol{\theta}) - \nabla q_{it}(\boldsymbol{\theta}')\} \psi_{\tau_i}(Y_{it} - q_{it}(\boldsymbol{\theta})) \\ & + \sum_i \nabla q_{it}(\boldsymbol{\theta}^*) \left\{\mathsf{P}[Y_{it} \le q_{it}(\boldsymbol{\theta}) \mid \mathcal{F}_{it}] - \mathsf{P}[Y_{it} \le q_{it}(\boldsymbol{\theta}') \mid \mathcal{F}_{it}]\right\} \\ & + \sum_i \nabla q_{it}(\boldsymbol{\theta}^*) \left\{\mathbf{1}^c[Y_{it} \le q_{it}(\boldsymbol{\theta})] - \mathbf{1}^c[Y_{it} \le q_{it}(\boldsymbol{\theta}')]\right\}, \end{aligned}$$

and, similarly, the difference $\tilde{\mathbf{g}}_t(\boldsymbol{\theta}) - \tilde{\mathbf{g}}_t(\boldsymbol{\theta}^*)$ has only two first terms in this decomposition. In the proof of Theorem 2 of WKM it is shown that with Assumption 2.3

$$\|\tilde{\mathbf{g}}_t(\boldsymbol{\theta}) - \tilde{\mathbf{g}}_t(\boldsymbol{\theta}')\| \le D_2(np + f_0 D_1)\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|.$$

Let us fix some unit $\boldsymbol{\gamma} \in \mathbb{R}^p$ and apply Theorem 1 of Merlevède et al. (2009) to the sum $\sum_t \boldsymbol{\gamma}^\top\{\tilde{\mathbf{g}}_t(\boldsymbol{\theta}) - \tilde{\mathbf{g}}_t(\boldsymbol{\theta}')\}$. Since by Assumption 2.4 it holds $\alpha(k) \le \exp(-ck)$, we have a Hoeffding-type inequality for each $\mathtt{x} \ge 0$,

$$\boldsymbol{\gamma}^\top \Big\{\sum_t \tilde{\mathbf{g}}_t(\boldsymbol{\theta}) - \boldsymbol{\lambda}_t(\boldsymbol{\theta}) - \tilde{\mathbf{g}}_t(\boldsymbol{\theta}') + \boldsymbol{\lambda}_t(\boldsymbol{\theta}')\Big\} > C_1\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|(\sqrt{\mathtt{x}T} + \mathtt{x}\log^2 T) \qquad \text{(S1.1)}$$

with probability $\ge 1 - C_2 e^{-\mathtt{x}}$, where C_1 and C_2 only depend on γ. Further we apply

Theorem 2.2.27 of Talagrand (2014) to get for any $\mathtt{x} \geq 0$

$$\mathsf{P}\left(\sup_{\boldsymbol{\theta}\in\Theta\,:\,\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|\leq r} \left\| \sum_t \tilde{\mathbf{g}}_t(\boldsymbol{\theta}) - \boldsymbol{\lambda}_t(\boldsymbol{\theta}) - \tilde{\mathbf{g}}_t(\boldsymbol{\theta}') + \boldsymbol{\lambda}_t(\boldsymbol{\theta}') \right\| > LA(r,\mathtt{x}) \right) \leq LC_2 e^{-\mathtt{x}},$$

where $A(r,\mathtt{x}) = \sqrt{T}\gamma_2(rB_1, \|\cdot\|)\sqrt{\mathtt{x}} + (\log^2 T)\gamma_1(rB_1, \|\cdot\|)\mathtt{x}$, with L being a generic constant, B_1 is a unit ball in $\mathbb{R}^p$, and $\gamma_{1,2}(T, \|\cdot\|)$ are Talagrand gamma-functional, precisely, see Definition 2.2.18 in Talagrand (2014). In the case of finite dimensional space, we have $\gamma_{1,2}(rB_1(0), \|\cdot\|) \leq rC$, where $C = C(p)$ only depends on the dimension. We therefore can rewrite the above inequality,

$$\mathsf{P}\left(\sup_{\boldsymbol{\theta}\in\Theta\,:\,\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|\leq r} \left\| \sum_t \tilde{\mathbf{g}}_t(\boldsymbol{\theta}) - \boldsymbol{\lambda}_t(\boldsymbol{\theta}) - \tilde{\mathbf{g}}_t(\boldsymbol{\theta}') + \boldsymbol{\lambda}_t(\boldsymbol{\theta}') \right\| > Cr(\sqrt{\mathtt{x}T} + \mathtt{x}\log^2 T) \right) \leq e^{-\mathtt{x}},$$

where C is a positive constant and only depends on n and γ, and $\mathtt{x} \geq 1$.

Consider a δ-net $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N\}$ of the set $\Theta_0(r)$, so that for each $\boldsymbol{\theta} \in \Theta_0(r)$ there is $j = 1..N$ with $\|\boldsymbol{\theta} - \boldsymbol{\theta}_j\| \leq \delta$. It is known that there is such a set with $\log N \leq Cp \log \frac{r}{\delta}$ elements. By Bernstein-type inequality, Theorem 2 in Merlevède et al. (2009), it holds

$$\begin{aligned} \left\| \sum_t \sum_i \nabla q_{it}(\boldsymbol{\theta}^*)(\mathbf{1}^c[Y_{it} \leq q_{it}(\boldsymbol{\theta}_k)] - \mathbf{1}^c[Y_{it} \leq q_{it}(\boldsymbol{\theta}^*)]) \right\| \leq\ & C\{\sqrt{rT}\sqrt{\mathtt{x} + \log N} \\ & + (\log T)^2(\mathtt{x} + \log N)\}, \end{aligned}$$

uniformly for all $k = 1, \ldots, N$ with probability at least $1 - e^{-\mathtt{x}}$, and the constant only depend on n, γ. Here we use the fact that the terms $\mathbf{1}^c[Y_{it} \leq q_{it}(\boldsymbol{\theta})]$ are centred conditioned on $\mathcal{F}_{t-1}$, while $\nabla q_{it}(\boldsymbol{\theta})$ are $\mathcal{F}_t$ measurable.

Furthermore, taking into account part (iii) of Assumption 2.4 we can use Theorem 5.2 from Boucheron et al. (2005) to get that for any $i = 1, \ldots, n$

$$|\{t : \varepsilon_{it} \in [a,b]\}| \leq Tf_0(b-a) + C\sqrt{Tf_0(b-a)\mathtt{x}} + C\mathtt{x}$$

with probability at least $1 - 4e^{-\mathtt{x}}$ uniformly over all intervals, with some universal constant C. By definition, for any $\boldsymbol{\theta} \in \Theta_0(r)$ there is some k such that $|g_{it}(\boldsymbol{\theta}) - g_{it}(\boldsymbol{\theta}_k)| \leq D_1\delta$ for each i, t. Therefore, the amount of indices i, t, for which the values of $\mathbf{1}[Y_{it} - q_{it}(\boldsymbol{\theta})]$ and $\mathbf{1}[Y_{it} - q_{it}(\boldsymbol{\theta}_k)]$ differ is bounded by $C(T\delta + \sqrt{T\delta\mathtt{x}} + \mathtt{x})$, constant C does not depend on $T, \mathtt{x}, r$ and δ. We come to the conclusion that choosing $\delta = rT^{-1/2}$, on the intersection of the events listed above it holds,

$$\left\|\sum_t \sum_i \nabla q_{it}(\boldsymbol{\theta}^*)\{\mathbf{1}[Y_{it} \le q_{it}(\boldsymbol{\theta})] - \mathbf{1}[Y_{it} \le q_{it}(\boldsymbol{\theta}_k)]\}\right\| \lesssim T^{1/2} r + \sqrt{T^{1/2} r \mathtt{x}} + \mathtt{x}.$$

Putting the inequalities together we get the result.

S1.1 Proof of Proposition 2.1

The claim follows directly from a slightly flexible version, which we are using for the consistency of bootstrap estimator as well.

Lemma S1.1. *Let Assumptions 2.1-2.5 hold on the interval $\mathcal{I}$. Then there are positive constants T_0, a_0 such that whenever $|\mathcal{I}| \ge T_0$, $a \le a_0$ and $\mathtt{x} \le |\mathcal{I}|$ the following implication takes place with probability $\ge 1 - 6e^{-\mathtt{x}}$. Each $\boldsymbol{\theta} \in \Theta$ that satisfies,*

$$L_{\mathcal{I}}(\boldsymbol{\theta}) - L_{\mathcal{I}}(\boldsymbol{\theta}^*) \ge -|\mathcal{I}| a$$

satisfies as well

$$\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| \le \sqrt{a/b} + C_0 \sqrt{\frac{\mathtt{x} + \log |\mathcal{I}|}{|\mathcal{I}|}},$$

where the positive constants b, C_0 do not depend on $|\mathcal{I}|$ and $\mathtt{x}$.

First, we present a uniform bound for the score. Similar to (S1.1) it holds $\|\nabla \zeta(\boldsymbol{\theta}^*)\| \le C(\sqrt{\mathtt{x}T} + \mathtt{x} \log^2 T)$ with probability $\ge 1 - e^{-\mathtt{x}}$, while by Lemma 2.1 we have, with probability $\ge 1 - e^{-\mathtt{x}}$, that

$$\sup_{\boldsymbol{\theta} \in \Theta_0} \|\nabla \zeta(\boldsymbol{\theta}) - \nabla \zeta(\boldsymbol{\theta}^*)\| \le C(\sqrt{T}\sqrt{\mathtt{x} + \log T} + \mathtt{x} \log^2 T),$$

using the fact that the set Θ_0 is bounded. Using a simple triangle inequality we have,

$$\|\nabla \zeta_{\mathcal{I}}(\boldsymbol{\theta})\| \le C(\sqrt{T}\sqrt{\mathtt{x} + \log T} + \mathtt{x} \log^2 T) \tag{S1.2}$$

with probability $\ge 1 - 2e^{-\mathtt{x}}$ uniformly for each $\boldsymbol{\theta} \in \Theta_0$, with C not depending on $T, \mathtt{x}$.

Next we present a technical lemma, that shows quadratic deviation of the expectation of log-likelihood in the neighbourhood of true parameter. The resulting inequality is akin to condition $(\mathcal{L}_r)$ of Spokoiny (2017).

Lemma S1.2. *Suppose Assumptions 2.1-2.3 and 2.5 hold. Then, there are positive constants r_0, b that do not depend on $|\mathcal{I}|$, such that for each $\boldsymbol{\theta} \in \Theta$ satisfying $\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| \geq r$ it holds $\mathsf{E} L_{\mathcal{I}}(\boldsymbol{\theta}) - \mathsf{E} L_{\mathcal{I}}(\boldsymbol{\theta}^*) \leq -b|\mathcal{I}|(r^2 \wedge r_0^2)$.*

The proof of this lemma is postponed to Section S5.

Proof of Lemma S1.1. By (S1.2) we have for $\mathtt{x} \leq |\mathcal{I}|$,

$$\begin{aligned}\frac{1}{|\mathcal{I}|}\mathsf{E} L_{\mathcal{I}}(\boldsymbol{\theta}) - \frac{1}{|\mathcal{I}|}\mathsf{E} L_{\mathcal{I}}(\boldsymbol{\theta}^*) &\geq L_{\mathcal{I}}(\boldsymbol{\theta}) - L_{\mathcal{I}}(\boldsymbol{\theta}^*) - \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| \sup_{\boldsymbol{\theta}\in\Theta} \|\nabla \zeta_{\mathcal{I}}(\boldsymbol{\theta})\| \\ &\geq -a - C_2\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\||\mathcal{I}|^{-1/2}\sqrt{\mathtt{x} + \log|\mathcal{I}|} \\ &\geq -a_0 - C_2 R|\mathcal{I}|^{-1/2}\sqrt{\mathtt{x} + \log|\mathcal{I}|}\end{aligned}$$

with probability at least $1 - 2e^{-\mathtt{x}}$. By Lemma S1.2 this implies,

$$b\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|^2 \leq a + C_2\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\||\mathcal{I}|^{-1/2}\sqrt{\mathtt{x} + \log|\mathcal{I}|},$$

and it is left to notice that $x^2 \leq \alpha + \beta x$ implies $x \leq \sqrt{\alpha} + \beta$. Additionally, $L(\widetilde{\boldsymbol{\theta}}) \geq L(\boldsymbol{\theta}^*)$ pointwise, thus the deviation bound for the estimator takes place. □

S2 Proof of Proposition 2.2

First of all, by Proposition 2.1 , it holds with probability $\geq 1 - 7e^{-\mathtt{x}}$ that $\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| \leq r_0 = C_0\sqrt{T^{-1}(\mathtt{x} + \log T)}$. Applying Lemma 2.1 with this radius, we get that with probability $\geq 1 - 13e^{-\mathtt{x}}$ additionally this holds for each $\boldsymbol{\theta} \in \Theta_0(r_0)$:

$$\frac{1}{\sqrt{T}}\left\|\sum_t \mathbf{g}_t(\boldsymbol{\theta}) - \boldsymbol{\lambda}_t(\boldsymbol{\theta}) - \mathbf{g}_t(\boldsymbol{\theta}^*) + \boldsymbol{\lambda}_t(\boldsymbol{\theta}^*)\right\| \lesssim \delta_{T,\mathtt{x}} = \frac{(\mathtt{x} + \log T)^{3/4}}{T^{1/4}}. \tag{S2.3}$$

With $\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}$ and using $\sum_t \mathbf{g}_t(\widetilde{\boldsymbol{\theta}}) = 0$, $\sum_t \boldsymbol{\lambda}_t(\boldsymbol{\theta}^*) = 0$ we get,

$$\left\|\sqrt{T}Q(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) - \frac{1}{\sqrt{T}}\sum_t \mathbf{g}_t(\boldsymbol{\theta}^*)\right\| \lesssim \delta_{T,\mathtt{x}}.$$

It is shown in WKM (see formula (24)) that for each $\boldsymbol{\theta} \in \Theta$,

$$\left\|\sum_{t\in\mathcal{I}} \boldsymbol{\lambda}_t(\boldsymbol{\theta}) - \sum_{t\in\mathcal{I}} \boldsymbol{\lambda}_t(\boldsymbol{\theta}^*) + |\mathcal{I}|Q(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\right\| \leq C_2|\mathcal{I}|\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|^2, \tag{S2.4}$$

with C_2 is a positive constant and does not depend on the interval length.

Similar to the proof of Theorem 2.3 in Spokoiny (2017), introducing the error of quadratic approximation of log-likelihood near the true parameter and provided (S2.4) and (S2.3), one can show that the square root of log-likelihood ratio is approximated with the same rate, i.e. $\left|\sqrt{2L(\boldsymbol{\theta}) - 2L(\boldsymbol{\theta}^*)} - \|\boldsymbol{\xi}\|\right| \le \delta_{T,\mathtt{x}}$. Scaling $\mathtt{x} \leftarrow \mathtt{x} + \log 13$ provides the result.

S3 Proof of Proposition 3.1

Similar to the original likelihood,

$$\zeta^\circ(\boldsymbol{\theta}) = L^\circ(\boldsymbol{\theta}) - \mathsf{E}^\circ L^\circ(\boldsymbol{\theta}) = \sum_t (w_t - 1)\ell_t(\boldsymbol{\theta})$$

denotes the stochastic part of the likelihood in the bootstrap world.

Lemma S3.1. *Suppose Assumptions 2.2, 2.3 and 3.1 hold, for each* $\mathtt{x} \ge 1$ *with probability* $\ge 1 - 4e^{-\mathtt{x}}$ *w.r.t. to the data, the probability of*

$$\sup_{\boldsymbol{\theta}\in\Theta(r)} \frac{1}{T^{1/2}} \left\| \sum_t (w_t - 1)\{\mathbf{g}_t(\boldsymbol{\theta}) - \mathbf{g}_t(\boldsymbol{\theta}^*)\} \right\| \le \diamondsuit^\flat(T, r, \mathtt{x})$$

conditioned on the data is at least $1 - 3e^{-\mathtt{x}}$*, where*

$$\diamondsuit^\flat(T, r, \mathtt{x}) = C_3 \left(r \vee \sqrt{r} + T^{-1/4}\{(r\mathtt{x})^{1/2} \vee (r\mathtt{x})^{1/4}\} + T^{-1/2}\mathtt{x}\right) \sqrt{\mathtt{x} + \log T},$$

with a positive constant C_3 *not depending on* $T, r, \mathtt{x}$.

Proof. The proof is similar to that of Lemma 2.1. □

Corollary S3.1. *For* $\mathtt{x} \le \sqrt{T}$ *it holds with probability at least* $1 - 6e^{-\mathtt{x}}$*,*

$$\mathsf{P}^\circ \left(\sup_{\boldsymbol{\theta}\in\Theta} \|\nabla\zeta^\circ(\boldsymbol{\theta})\| \le C_5 T^{1/2} \sqrt{\mathtt{x} + \log T} \right) \le 1 - 5e^{-\mathtt{x}},$$

where C_5 *is a positive constant and does not depend on* $T, \mathtt{x}$.

Now we are ready to state the global concentration result for the bootstrap estimator.

Proposition S3.1. *Suppose Assumptions 2.2-2.5 and 3.1 hold. Then, on a set of probability at least* $1 - 12e^{-\mathtt{x}}$ *it holds with probability at least* $1 - 5e^{-\mathtt{x}}$ *conditioned on the data,*

$$\|\widetilde{\boldsymbol{\theta}}^{\circ} - \boldsymbol{\theta}^*\| \le C\sqrt{\frac{\mathtt{x} + \log T}{T}}.$$

Proof. Denote $r = \|\widetilde{\boldsymbol{\theta}}^{\circ} - \boldsymbol{\theta}\|$. Using Corollary S3.1 and the fact that $L^{\circ}(\widetilde{\boldsymbol{\theta}}^{\circ}) \ge L^{\circ}(\boldsymbol{\theta}^*)$, we have that on the event of probability at least $1 - 6e^{-\mathtt{x}}$ w.r.t. data, with probability at least $1 - 5e^{-\mathtt{x}}$ conditioned on the data,

$$\begin{aligned} L(\widetilde{\boldsymbol{\theta}}) - L(\boldsymbol{\theta}^*) &\ge L^{\circ}(\widetilde{\boldsymbol{\theta}}^{\circ}) - L^{\circ}(\boldsymbol{\theta}^*) - \|\widetilde{\boldsymbol{\theta}}^{\circ} - \boldsymbol{\theta}^*\| \times \sup \|\nabla \zeta^{\circ}(\boldsymbol{\theta})\| \\ &\ge -C_5 T^{1/2} r \sqrt{\mathtt{x} + \log T}. \end{aligned}$$

Using Proposition 2.1, we have additionally that on the other event of probability $1 - 6e^{-\mathtt{x}}$ it holds $r \lesssim \sqrt{r\sqrt{\frac{\mathtt{x}+\log T}{T}}} + \sqrt{\frac{\mathtt{x}+\log T}{T}}$, which yields the result. □

The rest can be accomplished using linear approximation of the score. Similar to the original likelihood, with $r_0 = \|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| \vee \|\widetilde{\boldsymbol{\theta}}^{\circ} - \boldsymbol{\theta}^*\|$ it follows from (S2.4),

$$\left\| \sum_t \boldsymbol{\lambda}_t(\widetilde{\boldsymbol{\theta}}^{\circ}) - \sum_t \boldsymbol{\lambda}_t(\widetilde{\boldsymbol{\theta}}) + TQ^2(\widetilde{\boldsymbol{\theta}}^{\circ} - \widetilde{\boldsymbol{\theta}}) \right\| \le 2C_2 T r_0^2.$$

Here, $\sum_t \boldsymbol{\lambda}_t(\boldsymbol{\theta})$ stands for the expectation of gradient of the likelihood. With help of Proposition 2.1 we first replace it with just the gradient, then, using Lemma S3.1 we replace it with the gradient of bootstrap likelihood. This finally leads to the proof of the proposition.

S4 Proof of Theorem 1

W.l.o.g. we have an interval $\mathcal{I} = \{1, \ldots, T\}$ and a set of break points $\mathcal{S}(\mathcal{I}) \subset \mathcal{I}$ to be considered. Let us denote $\underline{T} = \alpha_0 T$ with $\alpha_0 > 0$ from the conditions of the theorem. We have by Proposition 2.2 that, with probability at least $1 - e^{-\mathtt{x}}$, it holds for each $s \in \mathcal{S}(\mathcal{I})$,

$$\left| L_{A_{\mathcal{I},s}}(\widetilde{\boldsymbol{\theta}}_{A_{\mathcal{I},s}}) - L_{A_{\mathcal{I},s}}(\boldsymbol{\theta}^*) - \|\boldsymbol{\xi}_{A_{\mathcal{I},s}}\|^2/2 \right| \le \diamondsuit, \qquad \left| L_{B_{\mathcal{I},s}}(\widetilde{\boldsymbol{\theta}}_{B_{\mathcal{I},s}}) - L_{B_{\mathcal{I},s}}(\boldsymbol{\theta}^*) - \|\boldsymbol{\xi}_{B_{\mathcal{I},s}}\|^2/2 \right| \le \diamondsuit,$$

$$\left| L_{\mathcal{I}}(\widetilde{\boldsymbol{\theta}}_{\mathcal{I}}) - L_{\mathcal{I}}(\boldsymbol{\theta}^*) - \|\boldsymbol{\xi}_{A_{\mathcal{I}}}\|^2/2 \right| \le \diamondsuit,$$

where $\diamondsuit = CT^{-1/4}(\mathtt{x} + \log T + \log(1 + 2|\mathcal{S}(\mathcal{I})|))^{3/4}$, implying

$$\left| L_{A_{\mathcal{I},s}}(\widetilde{\boldsymbol{\theta}}_{A_{\mathcal{I},s}}) + L_{B_{\mathcal{I},s}}(\widetilde{\boldsymbol{\theta}}_{B_{\mathcal{I},s}}) - L_{\mathcal{I}}(\widetilde{\boldsymbol{\theta}}_{\mathcal{I}}) - (\|\boldsymbol{\xi}_{A_{\mathcal{I},s}}\|^2 + \|\boldsymbol{\xi}_{B_{\mathcal{I},s}}\|^2 - \|\boldsymbol{\xi}_{\mathcal{I}}\|^2)/2 \right| \le 3\diamondsuit.$$

By definition, $|\mathcal{I}|^{1/2}\boldsymbol{\xi}_{\mathcal{I}} = |A_{\mathcal{I},s}|^{1/2}\boldsymbol{\xi}_{A_{\mathcal{I},s}} + |B_{\mathcal{I},s}|^{1/2}\boldsymbol{\xi}_{B_{\mathcal{I},s}}$, therefore for $\alpha = |A_{\mathcal{I},s}|/|\mathcal{I}|$ and $\beta = |B_{\mathcal{I},s}|/|\mathcal{I}| = 1 - \alpha$ we have,

$$\begin{aligned}
\|\boldsymbol{\xi}_{A_{\mathcal{I},s}}\|^2 + \|\boldsymbol{\xi}_{B_{\mathcal{I},s}}\|^2 - \|\boldsymbol{\xi}_{\mathcal{I}}\|^2 &= \|\boldsymbol{\xi}_{A_{\mathcal{I},s}}\|^2 + \|\boldsymbol{\xi}_{B_{\mathcal{I},s}}\|^2 - \|\alpha^{1/2}\boldsymbol{\xi}_{A_{\mathcal{I},s}} + \beta^{1/2}\boldsymbol{\xi}_{B_{\mathcal{I},s}}\|^2 \\
&= \beta\|\boldsymbol{\xi}_{A_{\mathcal{I},s}}\|^2 + \alpha\|\boldsymbol{\xi}_{B_{\mathcal{I},s}}\|^2 - 2\alpha^{1/2}\beta^{1/2}\boldsymbol{\xi}_{A_{\mathcal{I},s}}^\top\boldsymbol{\xi}_{B_{\mathcal{I},s}} \\
&= \|\beta^{1/2}\boldsymbol{\xi}_{A_{\mathcal{I},s}} - \alpha^{1/2}\boldsymbol{\xi}_{B_{\mathcal{I},s}}\|^2
\end{aligned}$$

Obviously, similar expansion holds for the bootstrap counterpart, so that denoting

$$\begin{aligned}
S_{\mathcal{I},s} &= \frac{1}{\sqrt{|\mathcal{I}|}}\left[\sqrt{\frac{|B_{\mathcal{I},s}|}{|A_{\mathcal{I},s}|}}\sum_{t\in A_{\mathcal{I},s}} Q^{-1}\mathbf{g}_t(\boldsymbol{\theta}^*) - \sqrt{\frac{|A_{\mathcal{I},s}|}{|B_{\mathcal{I},s}|}}\sum_{t\in B_{\mathcal{I},s}} Q^{-1}\mathbf{g}_t(\boldsymbol{\theta}^*)\right], \\
S^\circ_{\mathcal{I},s} &= \frac{1}{\sqrt{|\mathcal{I}|}}\left[\sqrt{\frac{|B_{\mathcal{I},s}|}{|A_{\mathcal{I},s}|}}\sum_{t\in A_{\mathcal{I},s}} Q^{-1}w_t\mathbf{g}_t(\boldsymbol{\theta}^*) - \sqrt{\frac{|A_{\mathcal{I},s}|}{|B_{\mathcal{I},s}|}}\sum_{t\in B_{\mathcal{I},s}} Q^{-1}w_t\mathbf{g}_t(\boldsymbol{\theta}^*)\right],
\end{aligned}$$

we have

$$\left|\max_s T_{\mathcal{I},s} - \max_s \|S_{\mathcal{I},s}\|^2\right| \le 3\diamondsuit, \qquad \left|\max_s T^\circ_{\mathcal{I},s} - \max_s \|S^\circ_{\mathcal{I},s}\|^2\right| \le 3\diamondsuit. \tag{S4.5}$$

For a single break point $s \in \mathcal{S}(\mathcal{I})$ by Azuma-Hoeffding inequality for all $\mathtt{x} > 0$ it holds,

$$\mathsf{P}\left(\|S_{\mathcal{I},s}\| \lesssim 1 + \sqrt{\mathtt{x}}\right) \ge 1 - e^{-\mathtt{x}},$$

so that it holds with probability $\ge 1 - e^{-\mathtt{x}}$,

$$\max_s \|S_{\mathcal{I},s}\| \lesssim \sqrt{\log T} + \sqrt{\mathtt{x}}, \qquad \max_s \|S^\circ_{\mathcal{I},s}\| \lesssim \sqrt{\log T} + \sqrt{\mathtt{x}}.$$

Additionally, for each $A \subset \mathcal{I}$ the covariance

$$\mathsf{Var}^\circ(\boldsymbol{\xi}^\circ_A) = \frac{1}{|A|}\sum_{t\in A} Q^{-1}\mathbf{g}_t(\boldsymbol{\theta}^*)\mathbf{g}_t(\boldsymbol{\theta}^*)^\top Q^{-1}.$$

is concentrated near $\Sigma = \mathsf{Var}(Q^{-1}\mathbf{g}_1(\boldsymbol{\theta}^*)) = Q^{-1}V^2Q^{-1}$, e.g. by Azuma-Hoeffding

$$\mathsf{P}\left(\|\mathsf{Var}^\circ(\boldsymbol{\xi}^\circ_A) - \Sigma\| \lesssim \sqrt{\frac{1+\mathtt{x}}{|A|}}\right) \ge 1 - e^{-\mathtt{x}},$$

so that taking into account (13) , it holds with probability $\geq 1 - e^{-\mathtt{x}}$ that for each $A = A_{\mathcal{I},s}$ or $A = B_{\mathcal{I},s}$ with $s \in \mathcal{S}(\mathcal{I})$,

$$\|\mathsf{Var}^{\circ}(\boldsymbol{\xi}_A^{\circ}) - \Sigma\| \lesssim \sqrt{\frac{\log T + \mathtt{x}}{T}}. \tag{S4.6}$$

Now we want to use Lemma S7.2 with $n = T$. Since $\delta > 1$ by Assumption 2.4 , we can choose $c_2, c' > 0$ such that $(1+\delta)/2 - (1+2\delta)c_2 > 1 + c'$. Then, we can have $a, \epsilon > 0$ such that $a + \epsilon < \frac{1}{2} - 2c_2$ and $c_2 + (1+\delta)a > 1 + c'$. Setting $b = a + \gamma + \epsilon$, we have that

$$1 - b - \gamma a < -c', \qquad b < \frac{1}{2} - c_2, \qquad b - a > c_2.$$

This means that taking $q = \lceil T^a \rceil$ and $r = \lceil T^b \rceil$ and $D_n \lesssim \sqrt{\log n}$ by Assumption 3.1 , the conditions of Lemma S7.2 are satisfied. Moreover, by (S4.6) we have $\Delta \lesssim \sqrt{\log T / T}$ with probability $\geq 1 - 1/(2T)$, so that for each $t, y \in \mathbb{R}$

$$\left| \mathsf{P}(\max_s \|S_{\mathcal{I},s}\| > t) - \mathsf{P}(\max_s \|S_{\mathcal{I},s}^{\circ}\| > t + y) \right| \lesssim T^{-c \wedge c'} + |y| \log^{1/2} T. \tag{S4.7}$$

Thus, for $|y| \leq 6\diamondsuit$ taken for $\mathtt{x} = C \log T$, we have for each $t, y \in \mathbb{R}$

$$\sup_t \left| \mathsf{P}(\max_s T_{\mathcal{I},s} > t + y) - \mathsf{P}(\max_s T_{\mathcal{I},s}^{\circ} > t) \right| \lesssim T^{-c \wedge c'} + |y| \log^{1/2} T$$

with probability $\geq 1 - 1/T$.

S5 Proof of Lemma S1.2

Note that integrating the inequality (S2.4) with $Q = \sum_{i=1}^{n} \mathsf{E} f_{it}(0) \nabla q_{it}(\boldsymbol{\theta}^*)[\nabla q_{it}(\boldsymbol{\theta}^*)]^{\top}$, we get second-order approximation in the neighbourhood of $\boldsymbol{\theta}^*$,

$$\left| \frac{1}{T} \mathsf{E} L(\boldsymbol{\theta}) - \frac{1}{T} \mathsf{E} L(\boldsymbol{\theta}^*) + \|Q(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\|^2/2 \right| \leq C \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|^3,$$

therefore we get that for $\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| > r$ and $r \leq r_0 = \lambda_{\min}(Q^2)/(4C)$ we have

$$\frac{1}{T} \mathsf{E} L(\boldsymbol{\theta}) - \frac{1}{T} \mathsf{E} L(\boldsymbol{\theta}^*) < -b_{loc} r^2, \qquad b_{loc} = \lambda_{\min}(Q^2)/4.$$

Next, notice that if a r.v. Z has τ quantile 0, then for $\delta > 0$

$$\begin{aligned}
\mathsf{E}\rho_\tau(Z+\delta) - \mathsf{E}\rho_\tau(Z) &= \mathsf{E}(Z+\delta)(\tau - \mathbf{1}[Z+\delta \le 0]) - \mathsf{E}Z(\tau - \mathbf{1}[Z \le 0]) \\
&= \delta \mathsf{E}(\tau - \mathbf{1}(Z \le \delta) + \mathbf{1}[Z \in (-\delta, 0)]) + \mathsf{E}Z\mathbf{1}(Z \in (-\delta, 0)) \\
&= \mathsf{E}(Z+\delta)\mathbf{1}(Z \in (-\delta; 0)) \\
&\ge \frac{\delta}{2}\mathsf{E}\mathbf{1}(Z \in (-\delta/2; 0)) \\
&\ge \frac{\underline{f}\delta}{2}\left(\frac{\delta}{2} \wedge \delta_0\right),
\end{aligned}$$

and by analogy same bound takes place for $\mathsf{E}\rho_\tau(Z-\delta) - \mathsf{E}\rho_\tau(Z)$. Therefore,

$$-\mathsf{E}\ell_t(\boldsymbol{\theta}) + \mathsf{E}\ell_t(\boldsymbol{\theta}^*) \le \mathsf{E}\sum_{i=1}^n \frac{\underline{f}|q_{it} - q_{it}^*|}{2}\left(\frac{|q_{it} - q_{it}^*|}{2} \wedge \delta_0\right),$$

where due to (7), the right-hand side is bounded by $\underline{f}\delta(\delta \wedge \delta_0)/4$ with $\delta = \delta(r_0)$. Setting $b_{glob} = \underline{f}\delta(\delta \wedge \delta_0)/(4r_0^2)$, we get that the required inequality is satisfied with $b = b_{loc} \wedge b_{glob}$.

S6 Proof of Corollary 3.1

Let $z(\alpha)$ denotes $(1-\alpha)$-quantile of the test T, and $z^\circ(\alpha)$ is that of T° with respect to the bootstrap probability (here for convenience we write the confidence level in the brackets). Since $\mathsf{P}(X + Y > a + b) \le \mathsf{P}(X > a) + \mathsf{P}(Y \ge b)$ for arbitrary random variables X, Y and real numbers a, b, we have for each $\delta \in (0; \alpha)$

$$\begin{aligned}
\mathsf{P}(T > z^\circ(\alpha)) \le& \mathsf{P}(T > z(\alpha+\delta)) + \mathsf{P}(z^\circ(\alpha) \le z(\alpha+\delta)) \\
=& \alpha + \delta + \mathsf{P}(z^\circ(\alpha) \le z(\alpha+\delta)), \\
\mathsf{P}(T > z^\circ(\alpha)) \ge& \mathsf{P}(T > z(\alpha-\delta)) - \mathsf{P}(z^\circ(\alpha) \ge z(\alpha-\delta)) \\
=& \alpha - \delta - \mathsf{P}(z^\circ(\alpha) \ge z(\alpha-\delta)).
\end{aligned} \tag{S6.8}$$

Furthermore,

$$\begin{aligned}
\mathsf{P}(z^\circ(\alpha) \ge z(\alpha-\delta)) &= \mathsf{P}\left\{\mathsf{P}^\circ(T^\circ > z(\alpha-\delta)) \ge \alpha\right\}, \\
\mathsf{P}(z^\circ(\alpha) \le z(\alpha+\delta)) &= \mathsf{P}\left\{\mathsf{P}^\circ(T^\circ > z(\alpha+\delta)) \le \alpha\right\}.
\end{aligned}$$

By Theorem 1 , we have that on a set of probability $\geq 1 - 1/T$,

$$\sup_t |P(T > t) - P^\circ(T^\circ > t)| \leq CT^{-c}.$$

Taking $\delta = 2CT^{-c}$ and $t = z(\alpha - \delta)$ we have,

$$P^\circ(T^\circ > z(\alpha - \delta)) \leq \alpha - \delta + CT^{-c} < \alpha$$

and in a similar way,

$$P^\circ(T^\circ > z(\alpha + \delta)) \geq \alpha + \delta - CT^{-c} > \alpha.$$

Thus, with this choice of δ it holds,

$$\mathsf{P}(z^\circ(\alpha) \leq z(\alpha + \delta)) \leq 1/T, \qquad \mathsf{P}(z^\circ(\alpha) \geq z(\alpha - \delta)) \leq 1/T,$$

which via (S6.8) concludes the proof.

S7 Technical tools

One contribution of our paper is that we use the multiplier bootstrap technique to construct the critical values, which is model-free and data-driven, see Spokoiny and Zhilova (2015). Theory 1 in main text ensures that the distribution of the bootstrap statistics $T_{\mathcal{I}}^\circ$ mimics the unknown distribution of the original test statistics $T_{\mathcal{I}}$, hence we can construct critical values for $T_{\mathcal{I}}$ by using the multiplier bootstrap statistics, see (14). Here we present some technical tools in order to prove Theorem 1 in part S4. Basically Lemma S7.2 is used to prove Theorem 1 in part S4, while Lemma S7.1 is used to derive Lemma S7.2.

Let $X_1, \ldots, X_n \in \mathbb{R}^d$ be a martingale difference sequence (MDS) with coefficients b_k, and set

$$\overline{\sigma}^2(q) = \max_{j=1,\ldots,d} \max_I \mathsf{Var}\left(q^{-1/2} \sum_{i \in I} X_{ij}\right),$$

$$\underline{\sigma}^2(q) = \min_{j=1,\ldots,d} \min_I \mathsf{Var}\left(q^{-1/2} \sum_{i \in I} X_{ij}\right),$$

where $\max_I, \min_I$ are taken with respect to the subsets $I \subset [1, n]$ of form $I = [i+1, \dots, i+q]$. Let additionally, with probability one

$$|X_{ij}| \le D_n, \qquad 1 \le i \le n; 1 \le j \le p.$$

Denote the statistics,

$$\check{T} = \max_{j=1,\dots,d} n^{-1/2} \sum_{i=1}^{n} X_{ij}, \tag{S7.9}$$

and let $\check{Y} = (\check{Y}_1, \dots, \check{Y}_d)^\top$ be normal with zero mean and covariance $\mathsf{E}\check{Y}\check{Y}^\top = \Sigma := \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}X_i X_i^\top$.

Theorem S7.1 (Chernozhukov et al. (2013), Theorem B.1)**.** *Suppose there are positive constants r, q such that $r + q \le n/2$ and for some positive constants c_1, C_1, $0 < c_2 < 1/4$, $c_1 \le \underline{\sigma}(q) \le \overline{\sigma}(q) \vee \overline{\sigma}(r) \le C_1$ for each $i = 1, \dots, n$, $j = 1, \dots, d$, it holds $(r/q)\log^2 d \le C_1 n^{-c_2}$, and*

$$\max\left\{qD_n \log^{1/2} d, rD_n \log^{3/2} d, \sqrt{q} D_n \log^{7/2} d\right\} \le C_1 n^{1/2-c_2}.$$

Then, there are positive constants c, C that only depend on c_1, c_2, C_1, such that

$$\sup_t \left| \mathsf{P}(\check{T} < t) - \mathsf{P}(\max_{j\le d} \check{Y}_j < t) \right| \le Cn^{-c} + 2(n/q - 1)b_r.$$

Suppose we have another MDS $X_1', \dots, X_n'$, from which we construct a similar to (S7.9) statistic $\check{T}'$. Suppose, the sequence has β-mixing coefficients bounded by the same values b_k and the values of the vectors bounded a.s. by the same D_n. Finally, let us set $\Sigma' = \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}X_i X_i^\top$. Combining the result above with Gaussian comparison and anti-concentration we get the following corollary.

Lemma S7.1. *Suppose there are positive constants q, r such that $q+r < n/2$, and positive constants c_1, C_1, $0 < c_2 < 1/4$ such that $c_1 \le \underline{\sigma}(q) \le \overline{\sigma}(q) \vee \overline{\sigma}(r) \le C_1$ holds for both (X_i) and (X_i'). Let $|\Sigma_{jk} - \Sigma'_{jk}| \le \Delta$ for each $j, k = 1, \dots, d$. Then under conditions of Theorem S7.1 it holds for each $t, \delta \in \mathbb{R}$,*

$$\left| \mathsf{P}(\check{T} > t + \delta) - \mathsf{P}(\check{T}' > t) \right| \le C\Delta^{1/3} \log^{2/3} p + C|\delta| \log^{1/2} p + Cn^{-c} + 2(n/q-1)b_r,$$

where positive constants $c, C > 0$ only depend on c_1, c_2, C_1.

Proof. Simply apply Theorem S7.1, together with Theorem 2 of Chernozhukov et al. (2015) and Theorem 1 of Chernozhukov et al. (2017). □

Let now $X_1, \dots, X_n \in \mathbb{R}^p$ be a martingale difference sequence, with β-mixing coefficients b_k and $\mathsf{Var}(X_i) = V$. We need to bring the statistics

$$\hat{T} = \max_{s \in \mathcal{S}} \frac{1}{\sqrt{n}} \left\| \sqrt{\frac{n-s}{s}} \sum_{i=1}^{s} X_i - \sqrt{\frac{s}{n-s}} \sum_{i=s+1}^{n} X_i \right\|$$

into the above form. Following Zhilova (2015) we consider the following approximation. Let G_ϵ be an ϵ-net of the unit sphere in $\mathbb{R}^p$, such that for each $\mathbf{a} \in \mathbb{R}^p$ it holds,

$$(1-\epsilon)\|\mathbf{a}\| \le \max_{\boldsymbol{\gamma} \in G_\epsilon} \boldsymbol{\gamma}^\top \mathbf{a} \le (1+\epsilon)\|\mathbf{a}\|.$$

Let $G_\epsilon = \{\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_{|G_\epsilon|}\}$ be fixed and set,

$$[X]_{G_\epsilon} = (\boldsymbol{\gamma}_1^\top X, \dots, \boldsymbol{\gamma}_{|G_\epsilon|}^\top X) \in \mathbb{R}^{|G_\epsilon|},$$

and having $\mathcal{S} = \{s_1 < s_2 < \dots < s_{|\mathcal{S}|}\}$ set for each $i = 1, \dots, n$ a stacked vector,

$$\widetilde{X}_i = \left(\alpha_{n,s_1}(i)[X_i]_{G_\epsilon}^\top, \dots, \alpha_{n,s_{|S|}}(i)[X_i]_{G_\epsilon}^\top\right)^\top \in \mathbb{R}^{|S| \times |G_\epsilon|},$$
$$\alpha_{n,s}(i) = sign(s-i+1/2)\left(\frac{n-s}{s}\right)^{sign(s-i+1/2)/2},$$

which implies that

$$(1-\epsilon)\hat{T} \le \max_j \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{X}_{ij} \le (1+\epsilon)\hat{T}.$$

For sake of simplicity assume, $a^{-1} \le s/(n-s) \le a$ for each $s \in \mathcal{S}$. Note that for each j and $|I| = q$ it holds for some $\boldsymbol{\gamma}$ that,

$$\mathsf{Var}\left(q^{-1/2} \sum_{i \in I} \widetilde{X}_{ij}\right) = \mathsf{Var}\left(q^{-1/2} \sum_{i \in I} \boldsymbol{\gamma}^\top X_i\right) \in [\sigma_{\min}(V), \sigma_{\max}(V)].$$

Suppose, there is another MDS $X_1', \dots, X_n'$ with same mixing properties and set for each interval I of observations,

$$V_I' = \frac{1}{q} \sum_{i \in I} \mathsf{E} X_i' [X_i']^\top, \qquad |I| = q,$$

and assume that for each such I it holds,

$$\|V_I' - V\| \le \Delta_I, \qquad \Delta_q = \max_{|I|=q} \Delta_I.$$

Denote by analogy the test statistics $\hat{T}'$ and the vectors $\widetilde{X}_i'$. In what follows we assume that the dimension p is constant and the size of $\mathcal{S}$ is growing with n. Moreover, assume that $|X_{ij}|, |X_{ij}'| \le D_n$ for each i, j and that $\hat{T}, \hat{T}' \le A_n$, all with probability $\ge 1 - 1/n$.

Lemma S7.2. *Suppose there are positive constants r, q such that $r + q \le n/2$ and for some positive constants $c_1, C_1 > 0$, $0 < c_2 < 1/4$ such that $c_1 \le \sigma_{\min}(V) \le \sigma_{\max}(V) \le C_1$ for each $i = 1, \dots, n$, $j = 1, \dots, d$, it holds $(r/q)\log^2 n \le C_1 n^{-c_2}$, and*

$$\max\left\{qD_n \log^{1/2} n, rD_n \log^{3/2} n, \sqrt{q}D_n \log^{7/2} n\right\} \le C_1 n^{1/2-c_2}.$$

Moreover, assume $\Delta_r, \Delta_q \le c_1/2$. Then, for any $C_2 > 0$ there are positive constants $c, C > 0$ that only depend on c_1, c_2, C_1, C_2, such that for each $t, \delta \in \mathbb{R}$ it holds,

$$\left|\mathsf{P}(\hat{T} > t + \delta) - \mathsf{P}(\hat{T}' > t)\right| \le C\Delta^{1/3}\log^{2/3} n + C(A_n n^{-C_2} + |\delta|)\log^{1/2} n$$
$$+ Cn^{-c} + 2(n/q - 1)b_r,$$

where $\Delta = \max_{s\in\mathcal{S}}\{\Delta_{[1,s]}, \Delta_{(s,n]}, \Delta_n\}$.

Proof. Take $\epsilon = n^{-C_2}$, then we can have $\log|G_\epsilon| \lesssim \log n$, so that if d is dimension of $\widetilde{X}$, then $\log p \lesssim \log n$. In order to apply Lemma S7.1 with $\delta = \epsilon A_n + \delta$, it is left to bound the covariance difference Δ. We have that (assuming $s_1 \le s_2$)

$$\begin{aligned}\frac{1}{n}\sum_{i=1} n\mathsf{E}\widetilde{X}_{ij}\widetilde{X}_{ik} &= \frac{1}{n}\sum_{i=1}^{n} a_{s_1,n}(i)a_{s_2,n}(i)\boldsymbol{\gamma}_1^\top \mathsf{E}X_iX_i^\top\boldsymbol{\gamma}_2\\ &= \gamma_1^\top\left[\frac{s_1\frac{n-s_1}{s_1}\frac{n-s_2}{s_2} - (s_2-s_1)\frac{s_1}{n-s_1}\frac{n-s_2}{s_2} + (n-s_2)\frac{s_1}{n-s_1}\frac{s_2}{n-s_2}}{n}V\right]\boldsymbol{\gamma}_2,\end{aligned}$$

while

$$\begin{aligned}\frac{1}{n}\sum_{i=1} n\mathsf{E}\widetilde{X}'_{ij}\widetilde{X}'_{ik} &= \frac{1}{n}\sum_{i=1}^{n} sign(s_1 - i + 1/2)sign(s_2 - i + 1/2)\boldsymbol{\gamma}_1^\top \mathsf{E}X'_i[X'_i]^\top\boldsymbol{\gamma}_2\\ &= \boldsymbol{\gamma}_1^\top\left[\frac{s_1\frac{n-s_1}{s_1}\frac{n-s_2}{s_2}V_{[1,s_1]} - (s_2-s_1)\frac{s_1}{n-s_1}\frac{n-s_2}{s_2}V_{(s_1,s_2]}}{n}\right.\\ &\qquad\left.+\frac{(n-s_2)\frac{s_1}{n-s_1}\frac{s_2}{n-s_2}V_{(s_2,n]}}{n}\right]\boldsymbol{\gamma}_2.\end{aligned}$$

Observe that $(s_2 - s_1)V_{(s_1,s_2]} = nV_{[1,n]} - s_1V_{[1,s_1]} - (n-s_2)V_{(s_2,n]}$. Therefore, the difference between two is bounded by,

$$\begin{aligned}|\Sigma_{jk} - \Sigma'_{jk}| &\le \frac{a^2 s_1}{n}\|V_{[1,s_1]} - V\| + \frac{a^2(n-s_2)}{n}\|V_{(s_2,n]} - V\| + a^2\|V_{[1,n]} - V\|\\ &\le 2a^2\max_{s\in\mathcal{S}}\{\Delta_{[1,s]}, \Delta_{(s,n]}, \Delta_n\},\end{aligned}$$

thus the statement follows. □

S8 Additional application results

This part we present additional application results. We consider two stock markets, namely, the S&P 500 and DAX series. Daily index returns are obtained from Datastream and our data cover the period from 3 January 2005 to 29 December 2017, in total 3390 trading days, see Figure S1. Table T1 collects the summary statistics.

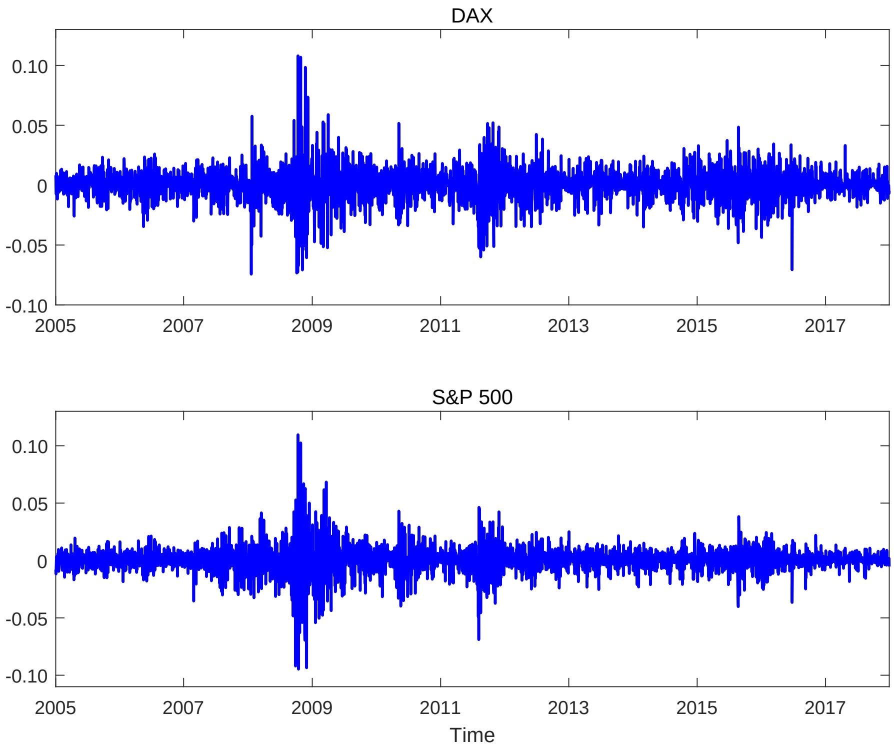

Figure S1: Selected index return time series from 3 January 2005 to 29 December 2017 (3390 trading days).

Index	Mean	Median	Min	Max	Std	Skew.	Kurt.
S&P 500	0.0002	0.0003	-0.0947	0.1096	0.0121	-0.3403	14.6949
DAX	0.0003	0.0007	-0.0743	0.1080	0.0137	-0.0406	9.2297

Table T1: Descriptive statistics for the selected index return time series from 3 January 2005 to 29 December 2017 (3390 trading days): mean, median, minimum (Min), maximum (Max), standard deviation (Std), skewness (Skew.) and kurtosis (Kurt.).

Figures S2 and S3 show the dynamics of estimated parameters with MV-CAViaR model in rolling window estimation. Parameter estimates are indeed more volatile when fitting the MV-CAViaR over shorter intervals (60 days), see e.g. More precisely, we display the estimated MV-CAViaR parameters $\hat{\beta}_{11}, \hat{\beta}_{12}, \hat{\beta}_{21}, \hat{\beta}_{22}$ in model (19) in rolling window exercises from 1 January 2007 to 29 December 2017. The upper (lower) panel at each figure shows the estimated parameter values if 60 (500) observations are included in the respective window.

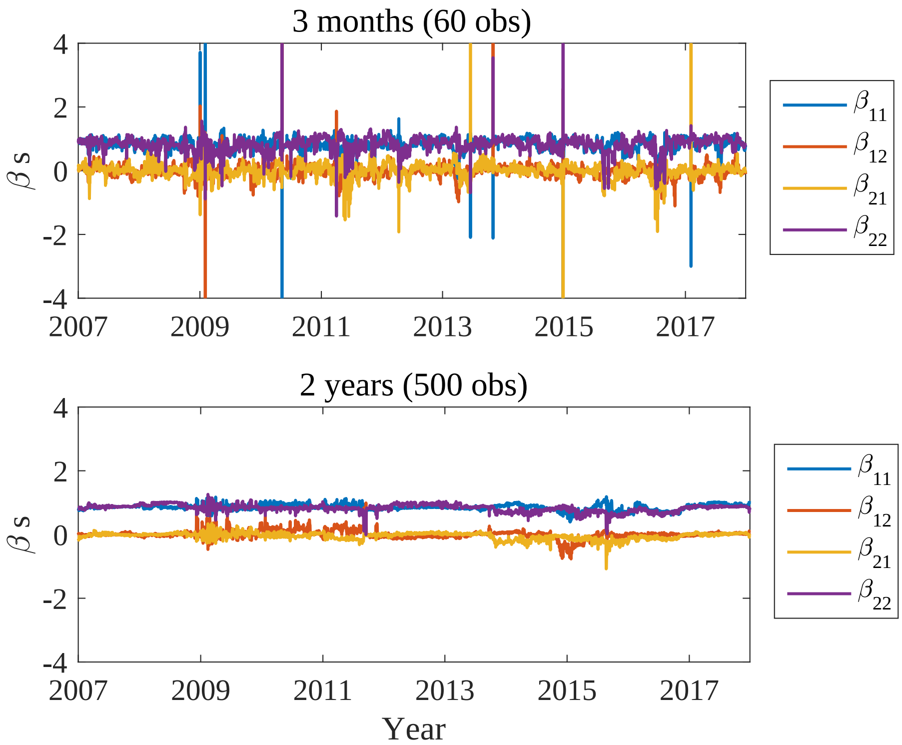

Figure S2: Estimated parameters $\hat{\beta}_{11}, \hat{\beta}_{12}, \hat{\beta}_{21}, \hat{\beta}_{22}$ at quantile level $\tau = 0.05$ for the selected two stock markets from 1 January 2007 to 29 December 2017, with 60 (upper panel) and 500 (lower panel) observations used in the rolling window exercises.

Figure S4 presents the time-varying coefficients γ_{21} and γ_{12} at quantile level $\tau = 0.01$ and $\tau = 0.05$ between DAX and S&P 500. γ_{21} denotes the tail effects of DAX from the absolute return of S&P 500 while γ_{12} denotes the tail effects of S&P 500 from the absolute return of DAX in equation (19). The blue lines show results of the conservative risk case $\alpha = 0.8$ and the red lines depict results of the modest risk case $\alpha = 0.9$.

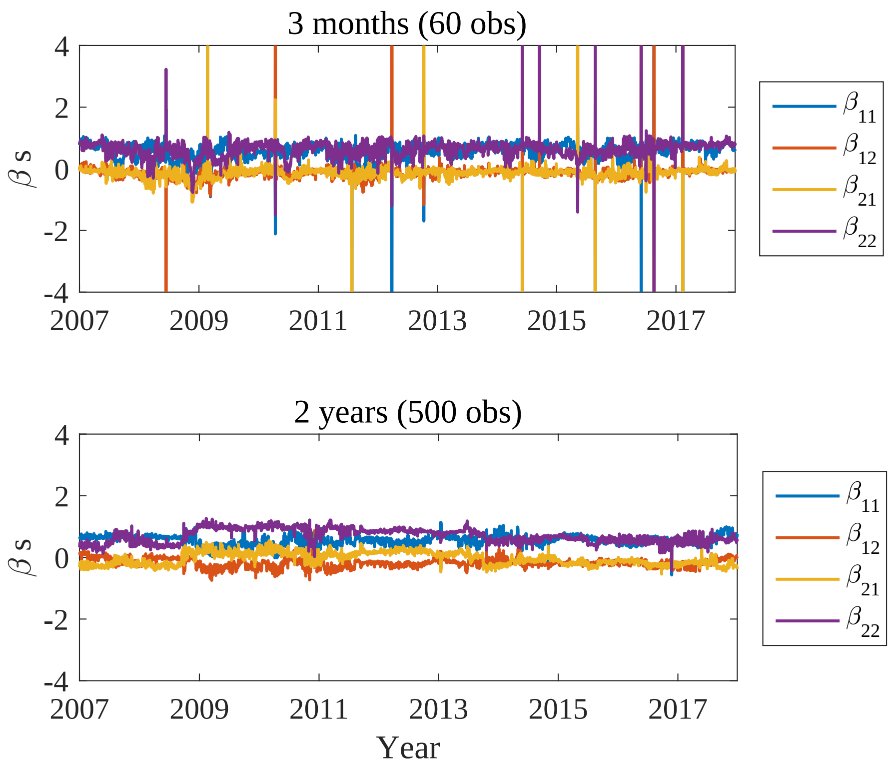

Figure S3: Estimated parameters $\hat{\beta}_{11}, \hat{\beta}_{12}, \hat{\beta}_{21}, \hat{\beta}_{22}$ at quantile level $\tau = 0.01$ for the selected two stock markets from 1 January 2007 to 29 December 2017, with 60 (upper panel) and 500 (lower panel) observations used in the rolling window exercises.

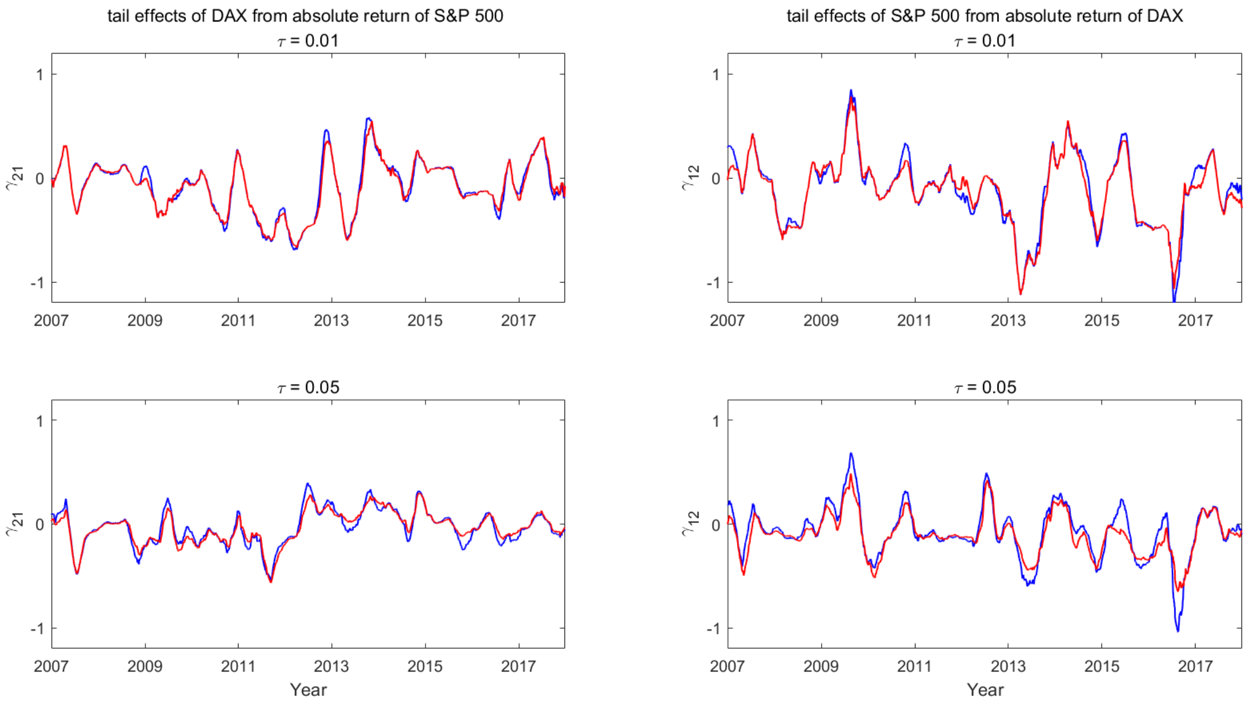

Figure S4: Time-varying coefficients γ_{21} and γ_{12} at quantile level $\tau = 0.01$ (upper panel) and $\tau = 0.05$ (lower panel) between DAX and S&P 500. The blue lines show results of the conservative risk case $\alpha = 0.8$ and the red lines depict results of the modest risk case $\alpha = 0.9$.

References

Boucheron, S., Bousquet, O. and Lugosi, G. (2005). Theory of classification: A survey of some recent advances, *ESAIM: probability and statistics* **9**: 323–375.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2013). Testing many moment inequalities, *arXiv preprint arXiv:1312.7614* .

Chernozhukov, V., Chetverikov, D. and Kato, K. (2015). Comparison and anti-concentration bounds for maxima of gaussian random vectors, *Probability Theory and Related Fields* **162**(1-2): 47–70.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2017). Detailed proof of Nazarov's inequality, *arXiv preprint arXiv:1711.10696* .

Merlevède, F., Peligrad, M., Rio, E. et al. (2009). Bernstein inequality and moderate deviations under strong mixing conditions, *High dimensional probability V: the Luminy volume*, Institute of Mathematical Statistics, pp. 273–292.

Spokoiny, V. (2017). Penalized maximum likelihood estimation and effective dimension, *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, Vol. 53, Institut Henri Poincaré, pp. 389–429.

Spokoiny, V. and Zhilova, M. (2015). Bootstrap confidence sets under model misspecification, *The Annals of Statistics* **43**(6): 2653–2675.

Talagrand, M. (2014). *Upper and lower bounds for stochastic processes: modern methods and classical problems*, Vol. 60, Springer Science & Business Media.

Zhilova, M. (2015). Simultaneous likelihood-based bootstrap confidence sets for a large number of models, *arXiv preprint arXiv:1506.05779* .