Supplement to "Multiscale Inference for Nonparametric Time Trends"

Marina Khismatullina
University of Bonn

Michael Vogt
University of Bonn

In this supplement, we provide the proofs that are omitted in the paper. Specifically, we prove Proposition A.3 and derive the technical results from Sections 4 and 5. We employ the same notation as summarized at the beginning of the Appendix in the paper.

Proof of Proposition 4.2

We only need to prove part (b). The arguments closely follow those for the proof of Proposition 3.3. To start with, we introduce the notation $\widehat{\psi}'_T(u,h) = \widehat{\psi}'^A_T(u,h) + \widehat{\psi}'^B_T(u,h)$, where $\widehat{\psi}'^A_T(u,h) = \sum_{t=1}^T w'_{t,T}(u,h)\varepsilon_t$ and $\widehat{\psi}'^B_T(u,h) = \sum_{t=1}^T w'_{t,T}(u,h) m_T(\frac{t}{T})$. We further write $m_T(\frac{t}{T}) = m_T(u) + m'_T(\xi_{u,t,T})(\frac{t}{T} - u)$, where $\xi_{u,t,T}$ is an intermediate point between u and t/T. The local linear weights $w'_{t,T}(u,h)$ are constructed such that $\sum_{t=1}^T w'_{t,T}(u,h) = 0$, which implies that

$$\widehat{\psi}'^B_T(u,h) = \sum_{t=1}^T w'_{t,T}(u,h)\Big(\frac{\frac{t}{T} - u}{h}\Big) h m'_T(\xi_{u,t,T}). \tag{S.1}$$

By assumption, there exists $(u_0,h_0) \in \mathcal{G}_T$ with $[u_0 - h_0, u_0 + h_0] \subseteq [0,1]$ such that $m'_T(w) \ge c_T\sqrt{\log T/(Th_0^3)}$ for all $w \in [u_0 - h_0, u_0 + h_0]$. (The case that $-m'_T(w) \ge c_T\sqrt{\log T/(Th_0^3)}$ for all w can be treated analogously.) Since the kernel K is symmetric and $u_0 = t/T$ for some t, it holds that $S_{T,1}(u_0,h_0) = 0$, which in turn implies that

$$\begin{aligned} & w'_{t,T}(u_0,h_0)\Big(\frac{\frac{t}{T} - u_0}{h_0}\Big) \\ & \quad = K\Big(\frac{\frac{t}{T} - u_0}{h_0}\Big)\Big(\frac{\frac{t}{T} - u_0}{h_0}\Big)^2 \Big/ \Big\{\sum_{t=1}^T K^2\Big(\frac{\frac{t}{T} - u_0}{h_0}\Big)\Big(\frac{\frac{t}{T} - u_0}{h_0}\Big)^2\Big\}^{1/2} \ge 0. \end{aligned}$$

From this and the assumption that $m'_T(w) \ge c_T\sqrt{\log T/(Th_0^3)}$ for all $w \in [u_0 - h_0, u_0 + h_0]$, we get that

$$\widehat{\psi}'^B_T(u_0,h_0) \ge c_T\sqrt{\frac{\log T}{Th_0}}\sum_{t=1}^T w'_{t,T}(u_0,h_0)\Big(\frac{\frac{t}{T} - u_0}{h_0}\Big). \tag{S.2}$$

Analogous to (A.11), we can show that for any $(u,h) \in \mathcal{G}_T$ with $[u-h,u+h] \subseteq [0,1]$,

$$\Big| \sum_{t=1}^{T} w'_{t,T}(u,h) \Big(\frac{\frac{t}{T} - u}{h} \Big) - \kappa' \sqrt{Th} \Big| \le \frac{C}{\sqrt{Th}}, \tag{S.3}$$

where $\kappa' = (\int K(\varphi)\varphi^2 d\varphi)/(\int K^2(\varphi)\varphi^2 d\varphi)^{1/2}$ and the constant C does not depend on u, h and T. (S.3) implies that $\sum_{t=1}^{T} w'_{t,T}(u,h)(\frac{t}{T} - u)/h \ge \kappa'\sqrt{Th}/2$ for sufficiently large T and any $(u,h) \in \mathcal{G}_T$ with $[u-h,u+h] \subseteq [0,1]$. From this and (S.2), we can infer that

$$\widehat{\psi}_T'^{B}(u_0,h_0) \ge \frac{\kappa' c_T \sqrt{\log T}}{2} \tag{S.4}$$

for sufficiently large T. Furthermore, by arguments very similar to those for the proof of Proposition A.1, it follows that

$$\max_{(u,h)\in\mathcal{G}_T} |\widehat{\psi}_T'^{A}(u,h)| = O_p(\sqrt{\log T}). \tag{S.5}$$

With the help of (S.4), (S.5) and the fact that $\lambda(h) \le \lambda(h_{\min}) \le C\sqrt{\log T}$, we can finally conclude that

$$\begin{aligned}
\widehat{\Psi}'_T &\ge \max_{(u,h)\in\mathcal{G}_T} \frac{|\widehat{\psi}_T'^{B}(u,h)|}{\widehat{\sigma}} - \max_{(u,h)\in\mathcal{G}_T} \Big\{ \frac{|\widehat{\psi}_T'^{A}(u,h)|}{\widehat{\sigma}} + \lambda(h) \Big\} \\
&= \max_{(u,h)\in\mathcal{G}_T} \frac{|\widehat{\psi}_T'^{B}(u,h)|}{\widehat{\sigma}} + O_p(\sqrt{\log T}) \\
&\ge \frac{\kappa' c_T \sqrt{\log T}}{2\widehat{\sigma}} + O_p(\sqrt{\log T}).
\end{aligned} \tag{S.6}$$

Since $q'_T(\alpha) = O(\sqrt{\log T})$ for any fixed $\alpha \in (0,1)$, (S.6) immediately yields that $\mathbb{P}(\widehat{\Psi}'_T \le q'_T(\alpha)) = o(1)$.

Proof of Proposition 4.3

In what follows, we show that

$$\mathbb{P}(E_T^{+}) \ge (1-\alpha) + o(1). \tag{S.7}$$

The other statement of Proposition 4.3 can be verified by analogous arguments. (S.7) is a consequence of the following two observations:

(i) For all $(u,h) \in \mathcal{G}_T$ with

$$\Big| \frac{\widehat{\psi}'_T(u,h) - \mathbb{E}\widehat{\psi}'_T(u,h)}{\widehat{\sigma}} \Big| - \lambda(h) \le q'_T(\alpha) \quad \text{and} \quad \frac{\widehat{\psi}'_T(u,h)}{\widehat{\sigma}} - \lambda(h) > q'_T(\alpha),$$

it holds that $\mathbb{E}[\widehat{\psi}'_T(u,h)] > 0$.

(ii) For all $(u,h) \in \mathcal{G}_T$ with $[u-h,u+h] \subseteq [0,1]$, $\mathbb{E}[\widehat{\psi}'_T(u,h)] > 0$ implies that $m'(v) > 0$ for some $v \in [u-h, u+h]$.

Observation (i) is trivial, (ii) can be seen as follows: Let (u,h) be any point with $(u,h) \in \mathcal{G}_T$ and $[u-h,u+h] \subseteq [0,1]$. It holds that $\mathbb{E}[\widehat{\psi}'_T(u,h)] = \widehat{\psi}'^B_T(u,h)$, where $\widehat{\psi}'^B_T(u,h)$ has been defined in the proof of Proposition 4.2. There, we have already seen that

$$\widehat{\psi}'^B_T(u,h) = \sum_{t=1}^T w'_{t,T}(u,h)\Big(\frac{\frac{t}{T}-u}{h}\Big)\, hm'(\xi_{u,t,T}),$$

where $\xi_{u,t,T}$ is some intermediate point between u and t/T. Moreover, $S_{T,1}(u,h) = 0$, which implies that $w'_{t,T}(u,h)(\frac{t}{T}-u)/h \ge 0$ for any t. Hence, $\mathbb{E}[\widehat{\psi}'_T(u,h)] = \widehat{\psi}'^B_T(u,h)$ can only take a positive value if $m'(v) > 0$ for some $v \in [u-h,u+h]$.
We now proceed as in the proof of Proposition 3.4. From observations (i) and (ii), we can infer the following: On the event

$$\big\{\widehat{\Phi}'_T \le q'_T(\alpha)\big\} = \Big\{ \max_{(u,h)\in\mathcal{G}_T} \Big(\Big|\frac{\widehat{\psi}'_T(u,h) - \mathbb{E}\widehat{\psi}'_T(u,h)}{\widehat{\sigma}}\Big| - \lambda(h)\Big) \le q'_T(\alpha)\Big\},$$

it holds that for all $(u,h) \in \mathcal{A}^+_T$, $m'(v) > 0$ for some $v \in I_{u,h} = [u-h,u+h]$. We thus obtain that $\{\widehat{\Phi}'_T \le q'_T(\alpha)\} \subseteq E^+_T$. This in turn implies that

$$\mathbb{P}(E^+_T) \ge \mathbb{P}\big(\widehat{\Phi}'_T \le q'_T(\alpha)\big) = (1-\alpha) + o(1),$$

where the last equality holds by Theorem 4.1.

Proof of Theorem 5.1

By Theorem 2.1 and Corollary 2.1 in Berkes et al. (2014), there exist a standard Brownian motion $\mathbb{B}_i$ and a sequence $\{\widetilde{\varepsilon}_{it} : t \in \mathbb{N}\}$ for each i such that the following holds: (i) $\mathbb{B}_i$ and $\{\widetilde{\varepsilon}_{it} : t \in \mathbb{N}\}$ are independent across i, (ii) $[\widetilde{\varepsilon}_{i1}, \dots, \widetilde{\varepsilon}_{iT}] \stackrel{\mathcal{D}}{=} [\varepsilon_{i1}, \dots, \varepsilon_{iT}]$ for each i and T, and (iii)

$$\max_{1\le t\le T} \Big| \sum_{s=1}^t \widetilde{\varepsilon}_{is} - \sigma_i \mathbb{B}_i(t)\Big| = o\big(T^{1/q}\big) \quad \text{a.s.}$$

for each i, where $\sigma_i^2 = \sum_{k\in\mathbb{Z}} \mathrm{Cov}(\varepsilon_{i0}, \varepsilon_{ik})$ denotes the long-run error variance. We define

$$\widetilde{\Phi}_{n,T} = \max_{1\le i<j\le N} \widetilde{\Phi}_{ij,T} \quad \text{with} \quad \widetilde{\Phi}_{ij,T} = \max_{(u,h)\in\mathcal{G}_T} \Big\{\Big|\frac{\widetilde{\phi}_{ij,T}(u,h)}{(\widetilde{\sigma}_i^2 + \widetilde{\sigma}_j^2)^{1/2}}\Big| - \lambda(h)\Big\},$$

where $\widetilde{\phi}_{ij,T}(u,h) = \sum_{t=1}^T w_{t,T}(u,h)\{(\widetilde{\varepsilon}_{it} - \bar{\widetilde{\varepsilon}}_i) - (\widetilde{\varepsilon}_{jt} - \bar{\widetilde{\varepsilon}}_j)\}$ with $\bar{\widetilde{\varepsilon}}_i = \bar{\widetilde{\varepsilon}}_{i,T} = T^{-1}\sum_{t=1}^T \widetilde{\varepsilon}_{it}$. Moreover, $\widetilde{\sigma}_i^2$ is the same estimator as $\widehat{\sigma}_i^2$ with $\widehat{Y}_{it} = (m_i(\frac{t}{T}) - \bar{m}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i)$ replaced by $\widetilde{Y}_{it} = (m_i(\frac{t}{T}) - \bar{m}_i) + (\widetilde{\varepsilon}_{it} - \bar{\widetilde{\varepsilon}}_i)$, where we set $\bar{m}_i = \bar{m}_{i,T} = T^{-1}\sum_{t=1}^T m_i(\frac{t}{T})$. By construction, $\widetilde{\Phi}_{n,T}$ has the same distribution as $\widehat{\Phi}_{n,T}$ for each $T \ge 1$. In addition, we

let
$$\Phi^*_{n,T} = \max_{1\le i<j\le N} \Phi^*_{ij,T} \quad \text{with} \quad \Phi^*_{ij,T} = \max_{(u,h)\in\mathcal{G}_T} \Big\{ \Big| \frac{\phi^*_{ij,T}(u,h)}{(\sigma_i^2+\sigma_j^2)^{1/2}} \Big| - \lambda(h) \Big\},$$
where $\phi^*_{ij,T}(u,h) = \sum_{t=1}^T w_{t,T}(u,h)\{\sigma_i(Z_{it}-\bar{Z}_i) - \sigma_j(Z_{jt}-\bar{Z}_j)\}$ and $Z_{it} = \mathbb{B}_i(t) - \mathbb{B}_i(t-1)$. Without loss of generality, we also set $Z_{it} = \mathbb{B}_i(t) - \mathbb{B}_i(t-1)$ in the Gaussian statistic $\Phi_{n,T}$ which is defined in Section 5.2.

We now follow the proof strategy for Theorem 3.1. Slightly modifying the arguments for Proposition A.1, we can show that
$$\big|\widetilde{\Phi}_{n,T} - \Phi^*_{n,T}\big| = o_p\Big(\frac{T^{1/q}}{\sqrt{Th_{\min}}} + \rho_T\sqrt{\log T}\Big). \tag{S.8}$$
Moreover, it holds that
$$\big|\Phi_{n,T} - \Phi^*_{n,T}\big| = o_p\big(\rho_T\sqrt{\log T}\big), \tag{S.9}$$
which is a consequence of the following facts: (i) the variables Z_{it} are i.i.d. standard normal, (ii) $|\mathcal{G}_T| = O(T^\theta)$ for some large but fixed constant θ, (iii) $\widehat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$, and (iv) $\max_{(u,h)\in\mathcal{G}_T} |\sum_{t=1}^T w_{t,T}(u,h)| \le CTh_{\max}$, where the constant C is independent of T. Finally, by arguments very similar to those for Proposition A.2, we obtain that
$$\sup_{x\in\mathbb{R}} \mathbb{P}\Big(\big|\Phi^*_{n,T} - x\big| \le \delta_T\Big) = o(1) \tag{S.10}$$
with $\delta_T = T^{1/q}/\sqrt{Th_{\min}} + \rho_T\sqrt{\log T}$. Combining (S.8)–(S.10) with Lemma A.4, we can infer that
$$\sup_{x\in\mathbb{R}} \big|\mathbb{P}\big(\widetilde{\Phi}_{n,T} \le x\big) - \mathbb{P}\big(\Phi^*_{n,T} \le x\big)\big| = o(1) \tag{S.11}$$
$$\sup_{x\in\mathbb{R}} \big|\mathbb{P}\big(\Phi_{n,T} \le x\big) - \mathbb{P}\big(\Phi^*_{n,T} \le x\big)\big| = o(1). \tag{S.12}$$
From (S.11) and (S.12), it immediately follows that
$$\sup_{x\in\mathbb{R}} \big|\mathbb{P}\big(\widetilde{\Phi}_{n,T} \le x\big) - \mathbb{P}\big(\Phi_{n,T} \le x\big)\big| = o(1),$$
which in turn implies that $\mathbb{P}(\widetilde{\Phi}_{n,T} \le q_{n,T}(\alpha)) = (1-\alpha) + o(1)$. Since $\widetilde{\Phi}_{n,T}$ has the same distribution as $\widehat{\Phi}_{n,T}$, this completes the proof of Theorem 5.1.

Proof of Proposition 5.3

Consider the event
$$B_{n,T} = \Big\{ \max_{1\le\ell\le N} \max_{i,j\in G_\ell} \widehat{\Psi}_{ij,T} \le q_{n,T}(\alpha) \ \text{ and } \min_{1\le\ell<\ell'\le N} \min_{\substack{i\in G_\ell,\\ j\in G_{\ell'}}} \widehat{\Psi}_{ij,T} > q_{n,T}(\alpha) \Big\}.$$

The term $\max_{1\le\ell\le N}\max_{i,j\in G_\ell}\widehat{\Psi}_{ij,T}$ is the largest multiscale distance between two time series i and j from the same group, whereas $\min_{1\le\ell<\ell'\le N}\min_{i\in G_\ell, j\in G_{\ell'}}\widehat{\Psi}_{ij,T}$ is the smallest multiscale distance between two time series from two different groups. On the event $B_{n,T}$, it obviously holds that

$$\max_{1\le\ell\le N}\max_{i,j\in G_\ell}\widehat{\Psi}_{ij,T} < \min_{1\le\ell<\ell'\le N}\min_{\substack{i\in G_\ell,\\ j\in G_{\ell'}}}\widehat{\Psi}_{ij,T}. \tag{S.13}$$

Hence, any two time series from the same class have a smaller distance than any two time series from two different classes. With the help of Theorem 5.1, it is easy to see that

$$\mathbb{P}\Big(\max_{1\le\ell\le N}\max_{i,j\in G_\ell}\widehat{\Psi}_{ij,T} \le q_{n,T}(\alpha)\Big) \ge (1-\alpha)+o(1).$$

Moreover, the same arguments as for part (b) of Proposition 5.2 show that

$$\mathbb{P}\Big(\min_{1\le\ell<\ell'\le N}\min_{\substack{i\in G_\ell,\\ j\in G_{\ell'}}}\widehat{\Psi}_{ij,T} \le q_{n,T}(\alpha)\Big) = o(1).$$

Taken together, these two statements imply that

$$\mathbb{P}\big(B_{n,T}\big) \ge (1-\alpha)+o(1). \tag{S.14}$$

In what follows, we show that on the event $B_{n,T}$, (i) $\big\{\widehat{G}_1^{[n-N]},\dots,\widehat{G}_N^{[n-N]}\big\} = \big\{G_1,\dots$ $\dots,G_N\big\}$ and (ii) $\widehat{N}=N$. From (i), (ii) and (S.14), the statements of Proposition 5.3 easily follow.

Proof of (i). Suppose we are on the event $B_{n,T}$. The proof proceeds by induction on the iteration steps r of the HAC algorithm.

Base case ($r=0$): In the first iteration step, the HAC algorithm merges two singleton clusters $\widehat{G}_i^{[0]}=\{i\}$ and $\widehat{G}_j^{[0]}=\{j\}$ with i and j belonging to the same group G_k. This is a direct consequence of (S.13). The algorithm thus produces a partition $\{\widehat{G}_1^{[1]},\dots,\widehat{G}_{n-1}^{[1]}\}$ whose elements $\widehat{G}_\ell^{[1]}$ all have the following property: $\widehat{G}_\ell^{[1]}\subseteq G_k$ for some k, that is, each cluster $\widehat{G}_\ell^{[1]}$ contains elements from only one group.

Induction step ($r \curvearrowright r+1$): Now suppose we are in the r-th iteration step for some $r<n-N$. Assume that the partition $\{\widehat{G}_1^{[r]},\dots,\widehat{G}_{n-r}^{[r]}\}$ is such that for any ℓ, $\widehat{G}_\ell^{[r]}\subseteq G_k$ for some k. Because of (S.13), the dissimilarity $\widehat{\Delta}(\widehat{G}_\ell^{[r]},\widehat{G}_{\ell'}^{[r]})$ gets minimal for two clusters $\widehat{G}_\ell^{[r]}$ and $\widehat{G}_{\ell'}^{[r]}$ with the property that $\widehat{G}_\ell^{[r]}\cup\widehat{G}_{\ell'}^{[r]}\subseteq G_k$ for some k. Hence, the HAC algorithm produces a partition $\{\widehat{G}_1^{[r+1]},\dots,\widehat{G}_{n-(r+1)}^{[r+1]}\}$ whose elements $\widehat{G}_\ell^{[r+1]}$ are all such that $\widehat{G}_\ell^{[r+1]}\subseteq G_k$ for some k.

The above induction argument shows the following: For any $r\le n-N$, the partition $\{\widehat{G}_1^{[r]},\dots,\widehat{G}_{n-r}^{[r]}\}$ consists of clusters $\widehat{G}_\ell^{[r]}$ which all have the property that $\widehat{G}_\ell^{[r]}\subseteq G_k$ for some k. This in particular holds for the partition $\{\widehat{G}_1^{[n-N]}, \ldots, \widehat{G}_N^{[n-N]}\}$, which implies that $\{\widehat{G}_1^{[n-N]}, \ldots, \widehat{G}_N^{[n-N]}\} = \{G_1, \ldots, G_N\}$. $\square$

Proof of (ii). To start with, consider any partition $\{\widehat{G}_1^{[n-r]}, \ldots, \widehat{G}_r^{[n-r]}\}$ with $r < N$ elements. Such a partition must contain at least one element $\widehat{G}_\ell^{[n-r]}$ with the following property: $\widehat{G}_\ell^{[n-r]} \cap G_k \neq \emptyset$ and $\widehat{G}_\ell^{[n-r]} \cap G_{k'} \neq \emptyset$ for some $k \neq k'$. On the event $B_{n,T}$, it obviously holds that $\widehat{\Delta}(S) > q_{n,T}(\alpha)$ for any S with the property that $S \cap G_k \neq \emptyset$ and $S \cap G_{k'} \neq \emptyset$ for some $k \neq k'$. Hence, we can infer that on the event $B_{n,T}$, $\max_{1 \le \ell \le r} \widehat{\Delta}(\widehat{G}_\ell^{[n-r]}) > q_{n,T}(\alpha)$ for any $r < N$.
Next consider the partition $\{\widehat{G}_1^{[n-r]}, \ldots, \widehat{G}_r^{[n-r]}\}$ with $r = N$ and suppose we are on the event $B_{n,T}$. From (i), we already know that $\{\widehat{G}_1^{[n-N]}, \ldots, \widehat{G}_N^{[n-N]}\} = \{G_1, \ldots, G_N\}$. Moreover, it is easy to see that $\widehat{\Delta}(G_\ell) \le q_{n,T}(\alpha)$ for any ℓ. Hence, we obtain that $\max_{1 \le \ell \le N} \widehat{\Delta}(\widehat{G}_\ell^{[n-N]}) = \max_{1 \le \ell \le N} \widehat{\Delta}(G_\ell) \le q_{n,T}(\alpha)$.
Putting everything together, we can conclude that on the event $B_{n,T}$,

$$\min \Big\{ r = 1, 2, \ldots \,\Big|\, \max_{1 \le \ell \le r} \widehat{\Delta}\big(\widehat{G}_\ell^{[n-r]}\big) \le q_{n,T}(\alpha) \Big\} = N,$$

that is, $\widehat{N} = N$. $\square$

Proof of Proposition 5.4

We consider the event

$$D_{n,T} = \Big\{ \widehat{\Phi}_{n,T} \le q_{n,T}(\alpha) \ \text{ and } \min_{1 \le \ell < \ell' \le N} \min_{\substack{i \in G_\ell, \\ j \in G_{\ell'}}} \widehat{\Psi}_{ij,T} > q_{n,T}(\alpha) \Big\},$$

where we write the statistic $\widehat{\Phi}_{n,T}$ as

$$\widehat{\Phi}_{n,T} = \max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T} \Big\{ \Big| \frac{\widehat{\psi}_{ij,T}(u,h) - \mathbb{E}\widehat{\psi}_{ij,T}(u,h)}{(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2)^{1/2}} \Big| - \lambda(h) \Big\}.$$

The event $D_{n,T}$ can be analysed by the same arguments as those applied to the event $B_{n,T}$ in the proof of Proposition 5.3. In particular, analogous to (S.14) and statements (i) and (ii) in this proof, we can show that

$$\mathbb{P}\big(D_{n,T}\big) \ge (1 - \alpha) + o(1) \tag{S.15}$$

and

$$D_{n,T} \subseteq \big\{ \widehat{N} = N \text{ and } \widehat{G}_\ell = G_\ell \text{ for all } \ell \big\}. \tag{S.16}$$

Moreover, we have that

$$D_{n,T} \subseteq \bigcap_{1 \le \ell < \ell' \le \widehat{N}} E_{n,T}(\ell, \ell'), \tag{S.17}$$

which is a consequence of the following observation: For all i, j and $(u,h) \in \mathcal{G}_T$ with

$$\Big|\frac{\widehat{\psi}_{ij,T}(u,h) - \mathbb{E}\widehat{\psi}_{ij,T}(u,h)}{(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2)^{1/2}}\Big| - \lambda(h) \le q_{n,T}(\alpha) \quad \text{and} \quad \Big|\frac{\widehat{\psi}_{ij,T}(u,h)}{(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2)^{1/2}}\Big| - \lambda(h) > q_{n,T}(\alpha),$$

it holds that $\mathbb{E}[\widehat{\psi}_{ij,T}(u,h)] \neq 0$, which in turn implies that $m_i(v) - m_j(v) \neq 0$ for some $v \in I_{u,h}$. From (S.16) and (S.17), we obtain that

$$D_{n,T} \subseteq \Big\{ \bigcap_{1 \le \ell < \ell' \le \widehat{N}} E_{n,T}(\ell,\ell') \Big\} \cap \big\{ \widehat{N} = N \text{ and } \widehat{G}_\ell = G_\ell \text{ for all } \ell \big\} = E_{n,T}.$$

This together with (S.15) implies that $\mathbb{P}(E_{n,T}) \ge (1-\alpha) + o(1)$, thus completing the proof.

Proof of Proposition A.3

The proof makes use of the following three lemmas, which correspond to Lemmas 5–7 in Chernozhukov et al. (2015).

Lemma S.1. *Let $(W_1,\dots,W_p)^\top$ be a (not necessarily centred) Gaussian random vector in $\mathbb{R}^p$ with $\mathrm{Var}(W_j) = 1$ for all $1 \le j \le p$. Suppose that $\mathrm{Corr}(W_j,W_k) < 1$ whenever $j \neq k$. Then the distribution of $\max_{1 \le j \le p} W_j$ is absolutely continuous with respect to Lebesgue measure and a version of the density is given by*

$$f(x) = f_0(x) \sum_{j=1}^p e^{\mathbb{E}[W_j]x - \mathbb{E}[W_j]^2/2}\, \mathbb{P}\big(W_k \le x \text{ for all } k \neq j \,\big|\, W_j = x\big).$$

Lemma S.2. *Let $(W_0,W_1,\dots,W_p)^\top$ be a (not necessarily centred) Gaussian random vector with $\mathrm{Var}(W_j) = 1$ for all $0 \le j \le p$. Suppose that $\mathbb{E}[W_0] \ge 0$. Then the map*

$$x \mapsto e^{\mathbb{E}[W_0]x - \mathbb{E}[W_0]^2/2}\, \mathbb{P}\big(W_j \le x \text{ for } 1 \le j \le p \,\big|\, W_0 = x\big)$$

is non-decreasing on $\mathbb{R}$.

Lemma S.3. *Let $(X_1,\dots,X_p)^\top$ be a centred Gaussian random vector in $\mathbb{R}^p$ with $\max_{1 \le j \le p} \mathbb{E}[X_j^2] \le \sigma^2$ for some $\sigma^2 > 0$. Then for any $r > 0$,*

$$\mathbb{P}\Big(\max_{1 \le j \le p} X_j \ge \mathbb{E}\Big[\max_{1 \le j \le p} X_j\Big] + r \Big) \le e^{-r^2/(2\sigma^2)}.$$

The proof of Lemmas S.1 and S.2 can be found in Chernozhukov et al. (2015). Lemma S.3 is a standard result on Gaussian concentration whose proof is given e.g. in Ledoux (2001); see Theorem 7.1 therein. We now closely follow the arguments for the proof of Theorem 3 in Chernozhukov et al. (2015). The proof splits up into three steps.

Step 1. Pick any $x \geq 0$ and set

$$W_j = \frac{X_j - x}{\sigma_j} + \frac{\overline{\mu} + x}{\underline{\sigma}}.$$

By construction, $\mathbb{E}[W_j] \geq 0$ and $\mathrm{Var}(W_j) = 1$. Defining $Z = \max_{1 \leq j \leq p} W_j$, it holds that

$$\begin{aligned}
\mathbb{P}\Big(\Big| \max_{1\leq j\leq p} X_j - x\Big| \leq \delta\Big) &\leq \mathbb{P}\Big(\Big| \max_{1\leq j\leq p} \frac{X_j - x}{\sigma_j}\Big| \leq \frac{\delta}{\underline{\sigma}}\Big) \\
&\leq \sup_{y\in\mathbb{R}} \mathbb{P}\Big(\Big| \max_{1\leq j\leq p} \frac{X_j - x}{\sigma_j} + \frac{\overline{\mu} + x}{\underline{\sigma}} - y\Big| \leq \frac{\delta}{\underline{\sigma}}\Big) \\
&= \sup_{y\in\mathbb{R}} \mathbb{P}\Big(|Z - y| \leq \frac{\delta}{\underline{\sigma}}\Big).
\end{aligned}$$

Step 2. We now bound the density of Z. Without loss of generality, we assume that $\mathrm{Corr}(W_j, W_k) < 1$ for $k \neq j$. The marginal distribution of W_j is $N(\nu_j, 1)$ with $\nu_j = \mathbb{E}[W_j] = (\mu_j/\sigma_j + \overline{\mu}/\underline{\sigma}) + (x/\underline{\sigma} - x/\sigma_j) \geq 0$. Hence, by Lemmas S.1 and S.2, the random variable Z has a density of the form

$$f_p(z) = f_0(z) G_p(z), \tag{S.18}$$

where the map $z \mapsto G_p(z)$ is non-decreasing. Define $\overline{Z} = \max_{1\leq j\leq p}(W_j - \mathbb{E}[W_j])$ and set $\overline{z} = 2\overline{\mu}/\underline{\sigma} + x(1/\underline{\sigma} - 1/\overline{\sigma})$ such that $\mathbb{E}[W_j] \leq \overline{z}$ for any $1 \leq j \leq p$. With these definitions at hand, we obtain that

$$\begin{aligned}
\int_z^\infty f_0(u)du\, G_p(z) \leq \int_z^\infty f_0(u) G_p(u) du = \mathbb{P}(Z > z) \\
\leq P(\overline{Z} > z - \overline{z}) \leq \exp\Big(-\frac{(z - \overline{z} - \mathbb{E}[\overline{Z}])_+^2}{2}\Big),
\end{aligned}$$

where the last inequality follows from Lemma S.3. Since $W_j - \mathbb{E}[W_j] = (X_j - \mu_j)/\sigma_j$, it holds that

$$\mathbb{E}[\overline{Z}] = \mathbb{E}\Big[\max_{1\leq j\leq p}\Big\{\frac{X_j - \mu_j}{\sigma_j}\Big\}\Big] =: a_p.$$

Hence, for every $z \in \mathbb{R}$,

$$G_p(z) \leq \frac{1}{1 - F_0(z)} \exp\Big(-\frac{(z - \overline{z} - a_p)_+^2}{2}\Big). \tag{S.19}$$

Mill's inequality states that for $z > 0$,

$$z \leq \frac{f_0(z)}{1 - F_0(z)} \leq z\frac{1 + z^2}{z^2}.$$

Since $(1 + z^2)/z^2 \leq 2$ for $z \geq 1$ and $f_0(z)/\{1 - F_0(z)\} \leq 1.53 \leq 2$ for $z \in (-\infty, 1)$, we can infer that

$$\frac{f_0(z)}{1-F_0(z)} \le 2(z \vee 1) \quad \text{for any } z \in \mathbb{R}.$$

This together with (S.18) and (S.19) yields that

$$f_p(z) \le 2(z \vee 1) \exp\Big(- \frac{(z - \overline{z} - a_p)_+^2}{2} \Big) \quad \text{for any } z \in \mathbb{R}.$$

Step 3. By Step 2, we get that for any $y \in \mathbb{R}$ and $u > 0$,

$$\mathbb{P}(|Z - y| \le u) = \int_{y-u}^{y+u} f_p(z) dz \le 2u \max_{z \in [y-u, y+u]} f_p(z) \le 4u(\overline{z} + a_p + 1),$$

where the last inequality follows from the fact that the map $z \mapsto z e^{-(z-a)^2/2}$ (with $a > 0$) is non-increasing on $[a+1, \infty)$. Combining this bound with Step 1, we further obtain that for any $x \ge 0$ and $\delta > 0$,

$$\mathbb{P}\Big(\Big| \max_{1 \le j \le p} X_j - x \Big| \le \delta \Big) \le 4\delta \Big\{ \frac{2\overline{\mu}}{\underline{\sigma}} + |x| \Big(\frac{1}{\underline{\sigma}} - \frac{1}{\overline{\sigma}} \Big) + a_p + 1 \Big\} / \underline{\sigma}. \tag{S.20}$$

This inequality also holds for $x < 0$ by an analogous argument, and hence for all $x \in \mathbb{R}$. Now let $0 < \delta \le \underline{\sigma}$ and define $b_p = \mathbb{E} \max_{1 \le j \le p} \{X_j - \mu_j\}$. For any $|x| \le \delta + \overline{\mu} + b_p + \overline{\sigma}\sqrt{2 \log(\underline{\sigma}/\delta)}$, (S.20) yields that

$$\begin{aligned} \mathbb{P}\Big(\Big| \max_{1 \le j \le p} X_j - x \Big| \le \delta \Big) &\le \frac{4\delta}{\underline{\sigma}} \Big\{ \overline{\mu} \Big(\frac{3}{\underline{\sigma}} - \frac{1}{\overline{\sigma}} \Big) + a_p + \Big(\frac{1}{\underline{\sigma}} - \frac{1}{\overline{\sigma}} \Big) b_p \\ &\qquad + \Big(\frac{\overline{\sigma}}{\underline{\sigma}} - 1 \Big) \sqrt{2 \log \Big(\frac{\underline{\sigma}}{\delta} \Big)} + 2 - \frac{\underline{\sigma}}{\overline{\sigma}} \Big\} \\ &\le C\delta \big\{ \overline{\mu} + a_p + b_p + \sqrt{1 \vee \log(\underline{\sigma}/\delta)} \big\} \end{aligned} \tag{S.21}$$

with a sufficiently large constant $C > 0$ that depends only on $\underline{\sigma}$ and $\overline{\sigma}$. For $|x| \ge \delta + \overline{\mu} + b_p + \overline{\sigma}\sqrt{2 \log(\underline{\sigma}/\delta)}$, we obtain that

$$\mathbb{P}\Big(\Big| \max_{1 \le j \le p} X_j - x \Big| \le \delta \Big) \le \frac{\delta}{\underline{\sigma}}, \tag{S.22}$$

which can be seen as follows: If $x > \delta + \overline{\mu}$, then $|\max_j X_j - x| \le \delta$ implies that $|x| - \delta \le \max_j X_j \le \max_j \{X_j - \mu_j\} + \overline{\mu}$ and thus $\max_j \{X_j - \mu_j\} \ge |x| - \delta - \overline{\mu}$. Hence, it holds that

$$\mathbb{P}\Big(\Big| \max_{1 \le j \le p} X_j - x \Big| \le \delta \Big) \le \mathbb{P}\Big(\max_{1 \le j \le p} \{X_j - \mu_j\} \ge |x| - \delta - \overline{\mu} \Big). \tag{S.23}$$

If $x < -(\delta + \overline{\mu})$, then $|\max_j X_j - x| \le \delta$ implies that $\max_j \{X_j - \mu_j\} \le -|x| + \delta + \overline{\mu}$.

Hence, in this case,

$$\mathbb{P}\Big(\Big|\max_{1\le j\le p} X_j - x\Big| \le \delta\Big) \le \mathbb{P}\Big(\max_{1\le j\le p}\big\{X_j-\mu_j\big\} \le -|x| + \delta + \overline{\mu}\Big)$$
$$\le \mathbb{P}\Big(\max_{1\le j\le p}\big\{X_j-\mu_j\big\} \ge |x| - \delta - \overline{\mu}\Big), \tag{S.24}$$

where the last inequality follows from the fact that for centred Gaussian random variables V_j and $v>0$, $\mathbb{P}(\max_j V_j \le -v) \le \mathbb{P}(V_1 \le -v) = P(V_1 \ge v) \le \mathbb{P}(\max_j V_j \ge v)$. With (S.23) and (S.24), we obtain that for any $|x| \ge \delta + \overline{\mu} + b_p + \overline{\sigma}\sqrt{2\log(\underline{\sigma}/\delta)}$,

$$\mathbb{P}\Big(\Big|\max_{1\le j\le p} X_j - x\Big| \le \delta\Big) \le \mathbb{P}\Big(\max_{1\le j\le p}\big\{X_j-\mu_j\big\} \ge |x| - \delta - \overline{\mu}\Big)$$
$$\le \mathbb{P}\Big(\max_{1\le j\le p}\big\{X_j-\mu_j\big\} \ge \mathbb{E}\Big[\max_{1\le j\le p}\big\{X_j-\mu_j\big\}\Big] + \overline{\sigma}\sqrt{2\log(\underline{\sigma}/\delta)}\Big) \le \frac{\delta}{\underline{\sigma}},$$

the last inequality following from Lemma S.3. To sum up, we have established that for any $0 < \delta \le \underline{\sigma}$ and any $x \in \mathbb{R}$,

$$\mathbb{P}\Big(\Big|\max_{1\le j\le p} X_j - x\Big| \le \delta\Big) \le C\delta\big\{\overline{\mu} + a_p + b_p + \sqrt{1 \vee \log(\underline{\sigma}/\delta)}\big\} \tag{S.25}$$

with some constant $C>0$ that does only depend on $\underline{\sigma}$ and $\overline{\sigma}$. For $\delta > \underline{\sigma}$, (S.25) trivially follows upon setting $C \ge 1/\underline{\sigma}$. This completes the proof.

References

Berkes, I., Liu, W. and Wu, W. B. (2014). Komlós-Major-Tusnády approximation under dependence. *Annals of Probability*, **42** 794–817.

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** 47–70.

Ledoux, M. (2001). *Concentration of Measure Phenomenon.* American Mathematical Society.