SUPPLEMENT TO "DETECTING GRADUAL CHANGES IN LOCALLY STATIONARY PROCESSES"

By Michael Vogt and Holger Dette

Universität Konstanz and Ruhr-Universität Bochum

In this supplement, we examine the finite sample performance of our method by further simulations. In addition, we provide the proofs that are omitted in the paper.

1. Simulations. In what follows, we continue the simulation study from Section 7.1 of the paper. As announced there, we examine a volatility model together with a multivariate extension of it. The univariate model is

$$X_{t,T} = \sigma\Big(\frac{t}{T}\Big)\varepsilon_t, \tag{S.1}$$

where σ is a time-varying volatility function and ε_t are i.i.d. residuals that are normally distributed with zero mean and unit variance. This is the same model as discussed in the application on the S&P 500 returns in Section 7.3 of the paper. Our aim is to estimate the time point where the volatility function σ starts to vary over time. We consider two different specifications of σ,

$$\sigma_1(u) = 1(u < 0.5) + 2 \cdot 1(u \geq 0.5)$$
$$\sigma_2(u) = 1(u < 0.5) + \{1 + 10(u - 0.5)\} \cdot 1(0.5 < u < 0.6) + 2 \cdot 1(u \geq 0.6),$$

both of which are equal to 1 on the interval $[0, 0.5]$ and then start to vary over time. Thus, $u_0 = 0.5$ in both cases. Analogously to the time-varying mean setting, σ_1 has a jump at $u_0 = 0.5$, whereas σ_2 smoothly deviates from its baseline value 1.

The multivariate extension of model (S.1) is given by the equation

$$X_{t,T} = \Sigma\Big(\frac{t}{T}\Big)\varepsilon_t, \tag{S.2}$$

where $X_{t,T} = (X_{t,T,1}, X_{t,T,2})^\intercal$ are bivariate random variables, $\Sigma(u)$ is a 2×2-matrix for each time point u and $\varepsilon_t = (\varepsilon_{t,1}, \varepsilon_{t,2})^\intercal$ are bivariate standard normal i.i.d. residuals. Since $\Sigma^2(\frac{t}{T}) := \Sigma(\frac{t}{T})\Sigma^\intercal(\frac{t}{T}) = \mathbb{E}[X_{t,T}X_{t,T}^\intercal]$, the time-varying matrix $\Sigma^2(\frac{t}{T})$ is the covariance matrix of $X_{t,T}$. Our aim is to estimate the time point where this matrix starts to vary over time. Put differently, we want to localize the time point where the covariance structure of $X_{t,T}$ starts to change. The stochastic feature of interest is thus the vector of covariances $\lambda_{t,T} = (\nu_{t,T}^{(1,1)}, \nu_{t,T}^{(1,2)}, \nu_{t,T}^{(2,2)})^\intercal$, where $\nu_{t,T}^{(i,j)} = \mathbb{E}[X_{t,T,i} X_{t,T,j}]$. We consider two different specifications of the volatility matrix Σ,

$$\Sigma_1(u) = \sigma_1(u) \cdot A$$
$$\Sigma_2(u) = \sigma_2(u) \cdot A,$$

where

$$AA^\intercal = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}, \quad \text{or put differently,} \quad A \approx \begin{pmatrix} 0.87 & -0.5 \\ 0.87 & 0.5 \end{pmatrix}$$

and $\sigma_1(u)$ along with $\sigma_2(u)$ are defined above. Both matrices $\Sigma_1(u)$ and $\Sigma_2(u)$ are constant on the interval $[0, 0.5]$ and then start to vary over time. Hence, as in the univariate case, $u_0 = 0.5$.

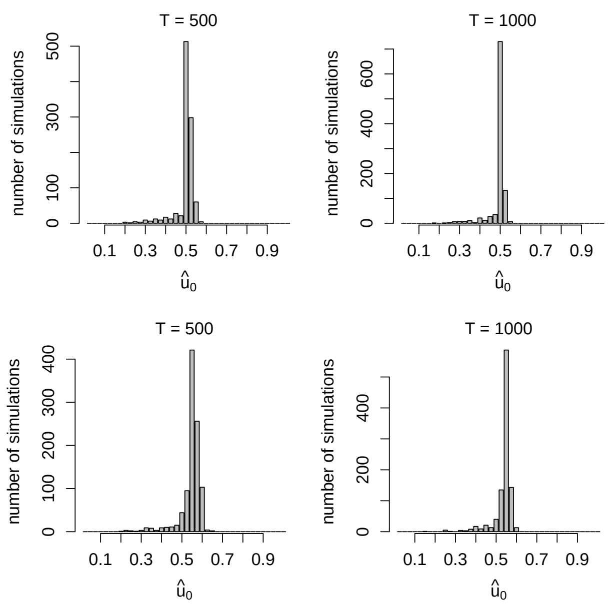

Fig 1: Simulation results for model (S.1) with the volatility function σ_1 (upper panel) and the function σ_2 (lower panel).

We implement our method as described in Setting II of Section 6 in the paper and let $\alpha = 0.1$, $h = 0.2$ as well as $b = 0$, exploiting the fact that the simulated data are independent. The resulting estimator is denoted by

$\hat{u}_0$. For each model specification, we draw $N = 1000$ samples of length $T \in \{500, 1000\}$ and compute the estimate of u_0 for each draw. The results are presented by means of histograms in the same way as in Section 7 of the paper.

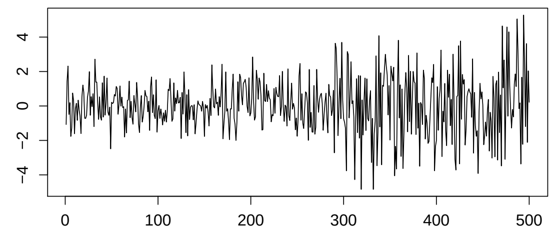

Fig 2: A typical sample path of length $T = 500$ for model (S.1) with σ_2.

We first discuss the results on the univariate model (S.1). The upper panel of Figure 1 presents the histograms for the design with σ_1, the lower panel those for the design with σ_2. The results are fairly similar to those from the time-varying mean setting: Our method is again able to detect the point u_0 quite precisely in the jump design with σ_1. The histograms in the setup with σ_2 are a bit more dispersed, reflecting the fact that it is harder to localize a gradual change than a jump.

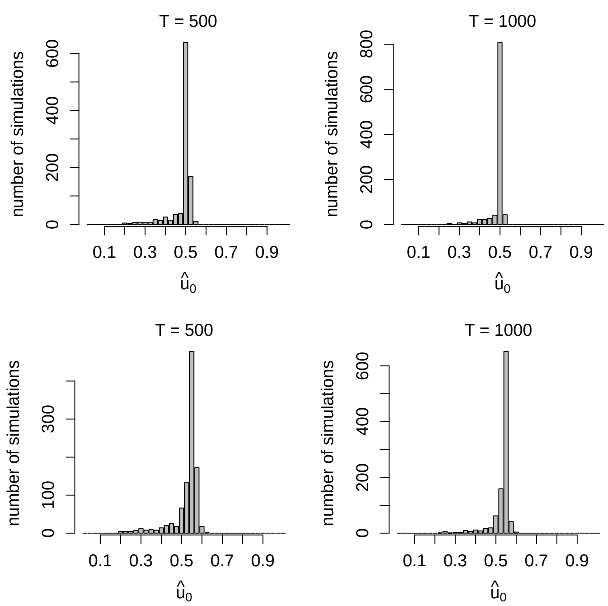

Fig 3: Simulation results for model (S.2) with the volatility matrix Σ_1 (upper panel) and the matrix Σ_2 (lower panel).

Figure 2 shows a typical sample path of length $T = 500$ for the design (S.1) with σ_2. As can be seen, the increase in the volatility level is hardly visible close to $u_0 = 0.5$ and only becomes apparent with some delay. It is thus natural that our procedure detects the time-variation in the volatility level only with a bit of delay. This produces the upward bias in the histograms which becomes less pronounced in larger samples.

We next turn to the results for the bivariate model (S.2). The histograms for the model with Σ_1 are displayed in the upper panel of Figure 3, those for the design with Σ_2 in the lower panel. Overall, the estimates give a good approximation to the true value u_0, those in the jump design with Σ_1 being a bit more precise than those in the gradual change design. Moreover, the histograms again make visible an upward bias which is comparable in size to that in the univariate setting.

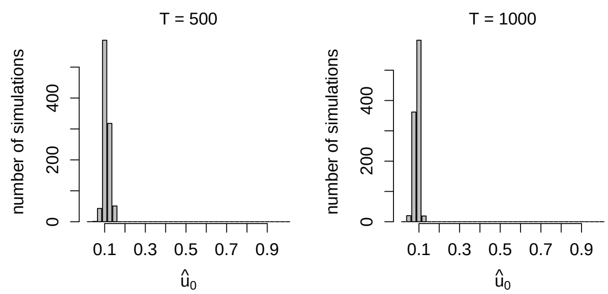

Fig 4: Simulation results produced by our method in model (7.1) with the mean function μ_3 defined in (S.3).

We finally investigate the performance of our procedure when the smooth change point u_0 occurs very early in the sample. In particular, we examine the extreme case that $u_0 = 0$. To do so, we go back to the time-varying mean setting (7.1) from Subsection 7.1 of the paper and consider the function

$$\mu_3(u) = 10u \cdot 1(0 \leq u < 0.2) + \{2 - 2.5(u - 0.2)\} \cdot 1(u \geq 0.2). \tag{S.3}$$

The simulation results for this design are depicted in Figure 4 and show that our method detects the time-variation rather quickly. Of course, it is only able to detect it with some delay which becomes smaller when moving to the larger sample size $T = 1000$.

2. Technical details. We now prove the main theoretical results of the paper. Throughout the section, the symbol C denotes a generic constant which may take a different value on each occurrence. Moreover, the expression $\|X\|_p = (\mathbb{E}|X|^p)^{1/p}$ is used to denote the L_p-norm of a real-valued random variable X.

Auxiliary results. Before we turn to the proofs of the main theorems, we derive some technical lemmas which are needed later on. To formulate them, we introduce some additional notation. To start with, partition the observations $\{X_{t,T} : t = 1, \dots, T\}$ into blocks of size q, where the r-th block spans the observations from time point $(r-1)q+1$ to rq and we set $q = CT^b$ for some small $b > 0$ (in particular $b < \frac{1}{4}$). Now define

$$W_T(k, k') = \sup_{f \in \mathcal{F}} \Big| \sum_{r=k}^{k'} Q_{r,T}(f) \Big|$$

along with

$$Q_{r,T}(f) = \frac{1}{\sqrt{(k' - k + 1)q}} \sum_{t=(2r-2)q+1}^{(2r-1)q \wedge T} \big(f(X_{t,T}) - \mathbb{E} f(X_{t,T}) \big).$$

The terms $Q_{r,T}(f)$ are scaled sums of the variables $f(X_{t,T}) - \mathbb{E}f(X_{t,T})$, the summation running over the observations of the $(2r-1)$-th block. The expression $W_T(k, k')$ sums up the terms $Q_{k,T}(f), \dots, Q_{k',T}(f)$ which correspond to the odd blocks $(2k-1), (2k+1), (2k+3), \dots, (2k'-1)$. The next two lemmas provide a bound on the L_p-norm of $W_T(k, k')$.

LEMMA A.1. *Let assumptions (C1) and (C2) be satisfied and let $f_0 \in \mathcal{F}$ have the property that $\mathbb{E}|f_0(X_{t,T})|^{(1+\delta)p} \le C$ for some even $p \in \mathbb{N}$ and a small $\delta > 0$. Then*

$$\Big\| \sum_{r=k}^{k'} Q_{r,T}(f_0) \Big\|_p \le C$$

for some sufficiently large constant C.

PROOF OF LEMMA A.1. To shorten notation, write $w_{t,T} = f_0(X_{t,T}) - \mathbb{E}f_0(X_{t,T})$ and bound the term

$$V_T = V_T(k, k') = \mathbb{E}\Big[\Big(\sum_{r=k}^{k'} Q_{r,T}(f_0) \Big)^p \Big]$$

by

$$V_T \le \frac{1}{((k'-k+1)q)^{p/2}} \times \sum_{r_1,\ldots,r_p=k}^{k'} \sum_{t_1=(2r_1-2)q+1}^{(2r_1-1)q\wedge T} \cdots \sum_{t_p=(2r_p-2)q+1}^{(2r_p-1)q\wedge T} \big|\mathbb{E}[w_{t_1,T}\ldots w_{t_p,T}]\big|$$
$$\le \frac{p!}{((k'-k+1)q)^{p/2}} \sum_{\substack{t_1,\ldots,t_p=(2k-2)q+1 \\ t_1\le\ldots\le t_p}}^{(2k'-1)q\wedge T} \big|\mathbb{E}[w_{t_1,T}\ldots w_{t_p,T}]\big|.$$

Let $(t_1,\ldots,t_p)$ be a tuple of ordered indices, that is, $t_1 \le \ldots \le t_p$. We say that the index t_i has a neighbour if $|t_i - t_{i-1}| \le C^* \log T$ or $|t_i - t_{i+1}| \le C^* \log T$ for some large constant C^* to be specified later on. Moreover, t_i is said to have exactly one neighbour if either $|t_i - t_{i-1}| \le C^* \log T$ and $|t_i - t_{i+1}| > C^* \log T$ or vice versa. Finally, we call (t_{i-1}, t_i) a pair of neighbours if $|t_i - t_{i-1}| \le C^* \log T$. Now let $S_\le$ denote the set of ordered tuples $(t_1,\ldots,t_p) \in \{(2k-2)q+1,\ldots,(2k'-1)q\wedge T\}^p$ such that each index t_i has a neighbour. In addition, let $S_>$ be the set of tuples such that at least one index does not have a neighbour. With this notation at hand, we can write $V_T \le V_T^{\le} + V_T^{>}$, where for $\ell \in \{\le, >\}$,

$$V_T^{\ell} = \frac{p!}{((k'-k+1)q)^{p/2}} \sum_{(t_1,\ldots,t_p)\in S_\ell} \big|\mathbb{E}[w_{t_1,T}\ldots w_{t_p,T}]\big|.$$

We now analyze the two terms $V_T^{\le}$ and $V_T^{>}$ separately. For the investigation of $V_T^{\le}$, define

$$S_{\le,a} = \big\{(t_1,\ldots,t_p) \in S_\le \mid \text{each index } t_i \text{ has exactly one neighbour}\big\}$$

together with $S_{\le,b} = S_\le \setminus S_{\le,a}$. First suppose that $(t_1,\ldots,t_p) \in S_{\le,a}$. In this case, there are exactly $p/2$ pairs (t_{2i-1}, t_{2i}) of neighbours (recalling that p is even by assumption). Using Davydov's inequality (see e.g. Corollary 1.1 in Bosq (1996)) to bound the covariances of the mixing variables $w_{t,T}$ and exploiting the fact that the mixing coefficients are decaying exponentially fast, we obtain that

$$\begin{aligned}
&\big|\mathbb{E}[w_{t_1,T}\ldots w_{t_p,T}]\big| \\
&\le \big|\mathbb{E}[w_{t_1,T}w_{t_2,T}]\mathbb{E}[w_{t_3,T}\ldots w_{t_p,T}]\big| + \big|\mathrm{Cov}(w_{t_1,T}w_{t_2,T}, w_{t_3,T}\ldots w_{t_p,T})\big| \\
&= \big|\mathbb{E}[w_{t_1,T}w_{t_2,T}]\mathbb{E}[w_{t_3,T}\ldots w_{t_p,T}]\big| + O\big(\alpha(C^*\log T)^{\xi}\big)
\end{aligned}$$

$$
\begin{aligned}
&= \big|\mathrm{Cov}(w_{t_1,T}, w_{t_2,T})\mathbb{E}[w_{t_3,T} \dots w_{t_p,T}]\big| + O\big(\alpha(C^* \log T)^{\xi}\big) \\
&\ \vdots \\
&\leq \Big|\prod_{i=1}^{p/2} \mathrm{Cov}(w_{t_{2i-1},T}, w_{t_{2i},T})\Big| + O(T^{-\nu}),
\end{aligned}
$$

where $\xi > 0$ is a sufficiently small number and the constant $\nu > 0$ can be made arbitrarily large (by choosing the constant C^* sufficiently large). This implies that

$$
\begin{aligned}
V_T^{\leq,a} &= \frac{p!}{((k'-k+1)q)^{p/2}} \sum_{(t_1,\dots,t_p)\in S_{\leq,a}} \big|\mathbb{E}[w_{t_1,T} \dots w_{t_p,T}]\big| \\
&\leq \frac{p!}{((k'-k+1)q)^{p/2}} \sum_{(t_1,\dots,t_p)\in S_{\leq,a}} \Big|\prod_{i=1}^{p/2} \mathrm{Cov}(w_{t_{2i-1},T}, w_{t_{2i},T})\Big| + o(1) \\
&\leq \frac{p!}{((k'-k+1)q)^{p/2}} \\
&\quad \times \prod_{i=1}^{p/2}\Big(\sum_{\ell=0}^{\lceil C^* \log T\rceil} \sum_{t_{2i-1}=(2k-2)q+1}^{(2k'-1)q\wedge T} \big|\mathrm{Cov}(w_{t_{2i-1},T}, w_{t_{2i-1}+\ell,T})\big|\Big) + o(1) \\
&\leq C\frac{p!}{((k'-k+1)q)^{p/2}} \\
&\quad \times ((k'-k+1)q)^{p/2}\Big(\sum_{\ell=0}^{\lceil C^* \log T\rceil} \alpha(\ell)^{\xi}\Big)^{p/2} + o(1) \leq C
\end{aligned}
$$

for some sufficiently large constant C, where the last line again uses Davydov's inequality to bound the covariance expressions in the formula.

Next consider the sum $V_T^{\leq,b}$ corresponding to indices in the set $S_{\leq,b}$. The cardinality of this set is bounded by $C((k'-k+1)q)^{\frac{p}{2}-1}(\log T)^{\frac{p}{2}+1}$, which implies

$$
\begin{aligned}
V_T^{\leq,b} &= \frac{p!}{((k'-k+1)q)^{p/2}} \sum_{(t_1,\dots,t_p)\in S_{\leq,b}} \big|\mathbb{E}[w_{t_1,T} \dots w_{t_p,T}]\big| \\
&\leq C\frac{(\log T)^{p/2+1}}{(k'-k+1)q} = o(1)
\end{aligned}
$$

(noting that $q = T^b$). This shows that the term $V_T^{\leq}$ is bounded.

Finally, we examine the term $V_T^{>}$ corresponding to the index set $S_{>}$. By definition, the tuples contained in this set have at least one element, say t_i, without a neighbour, that is, $|t_i - t_{i+1}| > C^* \log T$ and $|t_i - t_{i-1}| > C^* \log T$. Exploiting the mixing conditions on the model variables in a similar way as above, we obtain that

$$\begin{aligned}
&\mathbb{E}[w_{t_1,T} \dots w_{t_p,T}] \\
&= \mathbb{E}[w_{t_1,T} \dots w_{t_{i-1},T}]\mathbb{E}[w_{t_i,T} \dots w_{t_p,T}] + \mathrm{Cov}(w_{t_1,T} \dots w_{t_{i-1},T}, w_{t_i,T} \dots w_{t_p,T}) \\
&= \mathbb{E}[w_{t_1,T} \dots w_{t_{i-1},T}]\mathrm{Cov}(w_{t_i,T}, w_{t_{i+1},T} \dots w_{t_p,T}) + O(T^{-\nu}) = O(T^{-\nu}),
\end{aligned}$$

where ν can be chosen arbitrarily large (if C^* is chosen large enough). Recalling the definition of $V_T^>$, this yields that $V_T^> = o(1)$. Putting everything together, the quantity V_T is seen to be bounded. This completes the proof. □

Lemma A.2. *Let (C1) and (C2) be satisfied. Moreover, assume that for some even $p \in \mathbb{N}$ and some small $\delta > 0$,*

$$\mathbb{E}\Big[\Big|\frac{f(X_{t,T}) - f'(X_{t,T})}{d_{\mathcal{F}}(f,f')}\Big|^{(1+\delta)p}\Big] \le C$$

for all functions $f, f' \in \mathcal{F}$. Then for any $f_0 \in \mathcal{F}$,

$$\big\|W_T(k,k')\big\|_p \le C\Big(\Big\|\sum_{r=k}^{k'} Q_{r,T}(f_0)\Big\|_p + \int_0^{\mathrm{diam}(\mathcal{F})} \mathcal{N}(w/2, \mathcal{F}, d_{\mathcal{F}})^{1/p} dw\Big),$$

where $\mathcal{N}(w, \mathcal{F}, d_{\mathcal{F}})$ denotes the covering number of $(\mathcal{F}, d_{\mathcal{F}})$ and $\mathrm{diam}(\mathcal{F}) = \sup_{f,f' \in \mathcal{F}} d_{\mathcal{F}}(f,f')$ *is the diameter of $\mathcal{F}$.*

Proof of Lemma A.2. The claim immediately follows from Theorem 2.2.4 and Corollary 2.2.5 in van der Vaart and Wellner (1996) (see their remark on p.100 before Subsection 2.2.1). It thus suffices to verify the conditions of Theorem 2.2.4. In particular, we have to show that

$$\mathbb{E}\Big[\Big|\sum_{r=k}^{k'} Q_{r,T}(f) - \sum_{r=k}^{k'} Q_{r,T}(f')\Big|^p\Big] \le C d_{\mathcal{F}}(f,f')^p$$

for some sufficiently large constant C. To prove this, we introduce the notation

$$w_{t,T} = \frac{f(X_{t,T}) - f'(X_{t,T})}{d_{\mathcal{F}}(f,f')} - \mathbb{E}\Big[\frac{f(X_{t,T}) - f'(X_{t,T})}{d_{\mathcal{F}}(f,f')}\Big]$$

and consider

$$\begin{aligned}
V_T = V_T(k,k') &= \mathbb{E}\Big[\Big|\sum_{r=k}^{k'} \frac{Q_{r,T}(f) - Q_{r,T}(f')}{d_{\mathcal{F}}(f,f')}\Big|^p\Big] \\
&\leq \frac{1}{((k'-k+1)q)^{p/2}} \\
&\quad \times \sum_{r_1,\dots,r_p=k}^{k'} \sum_{t_1=(2r_1-2)q+1}^{(2r_1-1)q\wedge T} \cdots \sum_{t_p=(2r_p-2)q+1}^{(2r_p-1)q\wedge T} \big|\mathbb{E}[w_{t_1,T}\dots w_{t_p,T}]\big| \\
&\leq \frac{p!}{((k'-k+1)q)^{p/2}} \sum_{\substack{t_1,\dots,t_p=(2k-2)q+1 \\ t_1\leq\dots\leq t_p}}^{(2k'-1)q\wedge T} \big|\mathbb{E}[w_{t_1,T}\dots w_{t_p,T}]\big|.
\end{aligned}$$

Repeating the arguments from Lemma A.1, we can show that V_T is bounded, thus completing the proof. □

Proof of Theorem 5.1. To show that $\hat{H}_T = \sqrt{T}[\hat{D}_T - D]$ weakly converges to H, it suffices to prove that

$$\hat{H}_T^c := \sqrt{T}\big[\hat{D}_T - \mathbb{E}\hat{D}_T\big] \rightsquigarrow H \tag{S.4}$$

together with

$$\sqrt{T} \sup_{(u,v,f)\in\Delta\times\mathcal{F}} |\mathbb{E}\hat{D}_T - D| = o(1), \tag{S.5}$$

where $\hat{H}_T^c$ is the centred version of $\hat{H}_T$. We start with the proof of (S.5). Making use of condition (C4), we obtain that

$$\begin{aligned}
\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor uT\rfloor} \mathbb{E}\big[f(X_{t,T})\big] &= \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor uT\rfloor} \mathbb{E}\Big[f\Big(X_t\Big(\frac{t}{T}\Big)\Big)\Big] + o(1) \\
&= \sqrt{T} \sum_{t=1}^{\lfloor uT\rfloor} \int_{\frac{t-1}{T}}^{\frac{t}{T}} \mathbb{E}\big[f(X_t(w))\big]dw + o(1) \\
&= \sqrt{T} \int_0^u \mathbb{E}\big[f(X_t(w))\big]dw + o(1)
\end{aligned}$$

uniformly with respect to $u \in [0,1]$ and $f \in \mathcal{F}$. From this, (S.5) immediately follows. To verify (S.4), we show weak convergence of the finite dimensional distributions of $\hat{H}_T^c$ as well as stochastic equicontinuity of $\hat{H}_T^c$. In particular, we derive the following two results.

Proposition A.1. *For any finite number of points (u_i, v_i, f_i) with $1 \leq i \leq n$, it holds that*

$$\big(\hat{H}_T^c(u_1, v_1, f_1), \ldots, \hat{H}_T^c(u_n, v_n, f_n)\big)^\intercal \xrightarrow{d} N(0, \Sigma),$$

where $\Sigma = (\Sigma_{ij})_{1 \leq i,j \leq n}$ and $\Sigma_{ij} = \mathrm{Cov}(H(u_i, v_i, f_i), H(u_j, v_j, f_j))$.

Proposition A.2. *The sequence of processes $\hat{H}_T^c$ is asymptotically stochastically equicontinuous, that is, for any $\varepsilon > 0$,*

$$\lim_{\delta \searrow 0} \limsup_{T \to \infty} \mathbb{P}\Big(\sup_{\substack{|u-u'|+|v-v'| \\ +d_{\mathcal{F}}(f,f') \leq \delta}} \big| \hat{H}_T^c(u, v, f) - \hat{H}_T^c(u', v', f') \big| > \varepsilon \Big) = 0. \tag{S.6}$$

To prove these two results, we make use of the notation

$$\hat{H}_T^c(u, v, f) = \hat{G}_T(v, f) - \left(\frac{v}{u}\right) \hat{G}_T(u, f),$$

where

$$\hat{G}_T(u, f) = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor uT \rfloor} \big(f(X_{t,T}) - \mathbb{E} f(X_{t,T}) \big).$$

Combining Propositions A.1 and A.2, the statement (S.4) follows from a standard functional central limit theorem (see van der Vaart and Wellner (1996)).

Proof of Proposition A.1. We first calculate the asymptotic expectation and covariances of the process $\hat{H}_T^c$. As the process is centred, it holds that $\mathbb{E}[\hat{H}_T^c(u, v, f)] = 0$. Moreover,

$$\begin{aligned} \mathrm{Cov}\big(\hat{H}_T^c(u_1, v_1, f_1), \hat{H}_T^c(u_2, v_2, f_2)\big) = {} & \frac{v_1 v_2}{u_1 u_2} \, \mathbb{E}\big[\hat{G}_T(u_1, f_1) \hat{G}_T(u_2, f_2)\big] \\ & - \frac{v_2}{u_2} \, \mathbb{E}\big[\hat{G}_T(v_1, f_1) \hat{G}_T(u_2, f_2)\big] \\ & - \frac{v_1}{u_1} \, \mathbb{E}\big[\hat{G}_T(u_1, f_1) \hat{G}_T(v_2, f_2)\big] \\ & + \mathbb{E}\big[\hat{G}_T(v_1, f_1) \hat{G}_T(v_2, f_2)\big]. \end{aligned} \tag{S.7}$$

In what follows, we show that

$$\mathbb{E}\big[\hat{G}_T(u_1, f_1) \hat{G}_T(u_2, f_2)\big] = \sum_{\ell=-\infty}^{\infty} \int_0^{\min\{u_1, u_2\}} c_\ell(w) dw + o(1) \tag{S.8}$$

with $c_\ell(w) = c_\ell(w, f_1, f_2) = \mathrm{Cov}(f_1(X_0(w)), f_2(X_\ell(w)))$. Plugging (S.8) into (S.7) yields

$$\begin{aligned}\mathrm{Cov}\big(&\hat{H}_T^c(u_1, v_1, f_1), \hat{H}_T^c(u_2, v_2, f_2)\big) \\ &= \mathrm{Cov}\big(H(u_1, v_1, f_1), H(u_2, v_2, f_2)\big) + o(1).\end{aligned}$$

Hence, the covariances of $\hat{H}_T^c$ converge to those of the Gaussian process H.

To show (S.8), we assume without loss of generality that $u_1 \le u_2$. Exploiting the mixing condition (C2) by means of Davydov's inequality, it can be seen that $\mathrm{Cov}\big(f_1(X_{t,T}), f_2(X_{s,T})\big) \le C\alpha(|s-t|)^\xi \le C a^{\xi|s-t|}$ for some $a < 1$, a sufficiently small $\xi > 0$ and a large enough constant C. We thus obtain that

$$\begin{aligned}&\mathbb{E}\big[\hat{G}_T(u_1, f_1)\hat{G}_T(u_2, f_2)\big] \\ &= \frac{1}{T}\sum_{t=1}^{\lfloor u_1 T\rfloor}\sum_{s=1}^{\lfloor u_2 T\rfloor} \mathrm{Cov}\big(f_1(X_{t,T}), f_2(X_{s,T})\big) \\ &= \frac{1}{T}\sum_{t=1}^{\lfloor u_1 T\rfloor}\sum_{s=1}^{\lfloor u_2 T\rfloor} 1\{|s-t| \le C^* \log T\}\mathrm{Cov}\big(f_1(X_{t,T}), f_2(X_{s,T})\big) + o(1) \\ &=: Q_T^{(1)} + Q_T^{(2)} + Q_T^{(3)} + o(1)\end{aligned}$$

for some sufficiently large constant C^*, where the random variables $Q_T^{(j)}$ ($j = 1, 2, 3$) are defined by

$$\begin{aligned}Q_T^{(1)} &= \frac{1}{T}\sum_{\ell=1}^{\lceil C^* \log T\rceil}\sum_{t=1}^{T-\ell} 1\big\{t \le \lfloor u_1 T\rfloor, t+\ell \le \lfloor u_2 T\rfloor\big\} \\ &\qquad\qquad \times \mathrm{Cov}\big(f_1(X_{t,T}), f_2(X_{t+\ell,T})\big) \\ Q_T^{(2)} &= \frac{1}{T}\sum_{t=1}^{\lfloor u_1 T\rfloor} \mathrm{Cov}\big(f_1(X_{t,T}), f_2(X_{t,T})\big) \\ Q_T^{(3)} &= \frac{1}{T}\sum_{\ell=1}^{\lceil C^* \log T\rceil}\sum_{t=\ell+1}^{T} 1\big\{t \le \lfloor u_1 T\rfloor, t-\ell \le \lfloor u_2 T\rfloor\big\} \\ &\qquad\qquad \times \mathrm{Cov}\big(f_1(X_{t,T}), f_2(X_{t-\ell,T})\big).\end{aligned}$$

By assumption (C4), it follows for $\ell \le \lceil C^* \log T\rceil$ and any w with $|w - \frac{t}{T}| \le \frac{1}{T}$

that

$$\begin{aligned}
c_{t,T,\ell} &:= \mathrm{Cov}\big(f_1(X_{t,T}), f_2(X_{t+\ell,T})\big) \\
&= \mathrm{Cov}\Big(f_1\Big(X_t\Big(\frac{t}{T}\Big)\Big), f_2\Big(X_{t+\ell}\Big(\frac{t+\ell}{T}\Big)\Big)\Big) + O\Big(\frac{1}{\sqrt{T}}\Big) \\
&= \mathrm{Cov}\Big(f_1\Big(X_t\Big(\frac{t}{T}\Big)\Big), f_2\Big(X_{t+\ell}\Big(\frac{t}{T}\Big)\Big)\Big) + O\Big(\sqrt{\frac{\log T}{T}}\Big) \\
&= \mathrm{Cov}\big(f_1(X_0(w)), f_2(X_\ell(w))\big) + O\Big(\sqrt{\frac{\log T}{T}}\Big) \\
&=: c_\ell(w) + O\Big(\sqrt{\frac{\log T}{T}}\Big),
\end{aligned}$$

the last line defining $c_\ell(w)$ in an obvious manner. From this, it is easy to see that

$$\begin{aligned}
\frac{1}{T}\sum_{\ell=1}^{\lceil C^*\log T\rceil}\sum_{t=1}^{T-\ell}|c_{t,T,\ell}| &= \sum_{\ell=1}^{\lceil C^*\log T\rceil}\sum_{t=1}^{T-\ell}\int_{\frac{t-1}{T}}^{\frac{t}{T}}|c_\ell(w)|dw + O\Big(\log T\sqrt{\frac{\log T}{T}}\Big) \\
&= \sum_{\ell=1}^{\lceil C^*\log T\rceil}\int_0^1 |c_\ell(w)|dw + O\Big(\log T\sqrt{\frac{\log T}{T}}\Big).
\end{aligned}$$

Because of the mixing assumption (C2), the left-hand side of this equation is bounded as $T\to\infty$ and consequently $\sum_{\ell=1}^\infty\int_0^1 c_\ell(w)dw$ is absolutely convergent. Therefore we obtain for the term $Q_T^{(1)}$ as $T\to\infty$ (recall that $u_1\le u_2$)

$$\begin{aligned}
Q_T^{(1)} &= \sum_{\ell=1}^{\lceil C^*\log T\rceil}\sum_{t=1}^{\lfloor u_1T\rfloor-\ell}\int_{\frac{t-1}{T}}^{\frac{t}{T}} c_\ell(w)dw + O\Big(\log T\sqrt{\frac{\log T}{T}}\Big) \\
&= \sum_{\ell=1}^{\infty}\int_0^{u_1} c_\ell(w)dw + O\Big(\log T\sqrt{\frac{\log T}{T}}\Big)
\end{aligned}$$

and similarly

$$\begin{aligned}
Q_T^{(2)} &= \int_0^{u_1} c_0(w)dw + O\Big(\log T\sqrt{\frac{\log T}{T}}\Big) \\
Q_T^{(3)} &= \sum_{\ell=1}^{\infty}\int_0^{u_1} c_{-\ell}(w)dw + O\Big(\log T\sqrt{\frac{\log T}{T}}\Big).
\end{aligned}$$

Putting everything together, we arrive at (S.8).

Having calculated the asymptotic covariance structure of $\hat{H}_T^c$, we now apply a central limit theorem for mixing arrays of random variables (see e.g. Liebscher (1996)) together with the Cramér-Wold device to obtain weak convergence of the finite dimensional distributions. □

Proof of Proposition A.2. Straightforward calculations show that

$$\begin{aligned}\sup_{\substack{|u-u'|+|v-v'|\\+d_{\mathcal{F}}(f,f')\le\delta}} \left|\hat{H}_T^c(u,v,f)-\hat{H}_T^c(u',v',f')\right| &\le 2\sup_{\substack{|u-u'|\le\delta\\ f\in\mathcal{F}}}\left|\hat{G}_T(u,f)-\hat{G}_T(u',f)\right| \\ &\quad + 2\sup_{\substack{d_{\mathcal{F}}(f,f')\le\delta\\ u\in[0,1]}}\left|\hat{G}_T(u,f)-\hat{G}_T(u,f')\right| \\ &\quad + 2\sup_{\substack{u\in[0,1]\\ f\in\mathcal{F}}}\left|\delta^{\frac{1}{2}-\eta}\ \hat{G}_T(u,f)\right| + 2\sup_{\substack{u\in[0,\delta^{1/2+\eta}]\\ f\in\mathcal{F}}}\left|\hat{G}_T(u,f)\right|\end{aligned}$$

for some small $\eta>0$. Therefore, stochastic equicontinuity follows from the statements

$$\lim_{\delta\searrow 0}\limsup_{T\to\infty}\mathbb{P}\Big(\sup_{\substack{|u-u'|\le\delta\\ f\in\mathcal{F}}}\left|\hat{G}_T(u,f)-\hat{G}_T(u',f)\right|>\varepsilon\Big)=0 \tag{S.9}$$

$$\lim_{\delta\searrow 0}\limsup_{T\to\infty}\mathbb{P}\Big(\sup_{\substack{d_{\mathcal{F}}(f,f')\le\delta\\ u\in[0,1]}}\left|\hat{G}_T(u,f)-\hat{G}_T(u,f')\right|>\varepsilon\Big)=0 \tag{S.10}$$

$$\lim_{\delta\searrow 0}\limsup_{T\to\infty}\mathbb{P}\Big(\sup_{\substack{u\in[0,1]\\ f\in\mathcal{F}}}\left|\delta^{\frac{1}{2}-\eta}\ \hat{G}_T(u,f)\right|>\varepsilon\Big)=0 \tag{S.11}$$

$$\lim_{\delta\searrow 0}\limsup_{T\to\infty}\mathbb{P}\Big(\sup_{\substack{u\in[0,\delta^{1/2+\eta}]\\ f\in\mathcal{F}}}\left|\hat{G}_T(u,f)\right|>\varepsilon\Big)=0. \tag{S.12}$$

(S.9)–(S.12) can be shown by very similar arguments. We thus restrict ourselves to the proof of (S.9).

First of all, observe that for any function $g:[0,1]\to\mathbb{R}$, the inequality

$$\begin{aligned}\sup_{\substack{|u-u'|\le\delta\\ u,u'\in[0,1]}}|g(u)-g(u')| &\le \max_{j=1,\dots,\lceil 1/\delta\rceil}\sup_{u\in[u_{j-1},u_j]}|g(u)-g(u_j)| \\ &\quad + \max_{j=1,\dots,\lceil 1/\delta\rceil}\sup_{u'\in[u_{j-2},u_{j+1}]}|g(u')-g(u_j)|\end{aligned}$$

holds, where $u_{-1}=u_0=0$, $u_j=j\delta$ ($j=1,\dots,\lceil 1/\delta\rceil-1$) and $u_{\lceil 1/\delta\rceil}=$

$u_{\lceil 1/\delta \rceil + 1} = 1$. From this, it is easily seen that (S.9) is a consequence of

$$\lim_{\delta \searrow 0} \limsup_{T \to \infty} \mathbb{P}\Big(\max_{j=1,\ldots,\lceil 1/\delta \rceil} \sup_{u \in [u_{j-1}, u_j]} \sup_{f \in \mathcal{F}} \Big| \hat{G}_T(u,f) - \hat{G}_T(j\delta, f) \Big| > \varepsilon \Big) = 0. \tag{S.13}$$

In the sequel, we derive a suitable bound for the probability

$$P_T(\delta,\varepsilon) = \mathbb{P}\Big(\max_{j=1,\ldots,\lceil 1/\delta \rceil} \sup_{u \in [u_{j-1}, u_j]} \sup_{f \in \mathcal{F}} \Big| \hat{G}_T(u,f) - \hat{G}_T(j\delta, f) \Big| > \varepsilon \Big)$$

in (S.13). To start with, we crudely bound this probability by $P_T(\delta,\varepsilon) \le \sum_{j=1}^{\lceil 1/\delta \rceil} P_{T,j}(\delta,\varepsilon)$, where

$$\begin{aligned} P_{T,j}(\delta,\varepsilon) &= \mathbb{P}\Big(\sup_{u \in [u_{j-1}, u_j]} \sup_{f \in \mathcal{F}} \Big| \hat{G}_T(u,f) - \hat{G}_T(j\delta, f) \Big| > \varepsilon \Big) \\ &= \mathbb{P}\Big(\max_{\lfloor (j-1)\delta T \rfloor \le \ell \le \lfloor j\delta T \rfloor} \sup_{f \in \mathcal{F}} \Big| \hat{G}_T\Big(\frac{\ell}{T}, f\Big) - \hat{G}_T(j\delta, f) \Big| > \varepsilon \Big). \end{aligned}$$

To bound the probabilities $P_{T,j}(\delta,\varepsilon)$, we write

$$\hat{G}_T(j\delta, f) - \hat{G}_T\Big(\frac{\ell}{T}, f\Big) = B_T^{\ell+}(f) + \sum_{r=\lceil \frac{\ell}{q} \rceil + 1}^{\lfloor \frac{j\delta T}{q} \rfloor} B_{r,T}(f) + B_T^{j-}(f).$$

Here, $B_{r,T}(f)$ are blocks of length q given by

$$B_{r,T}(f) = \frac{1}{\sqrt{T}} \sum_{t=(r-1)q+1}^{rq} \big(f(X_{t,T}) - \mathbb{E} f(X_{t,T}) \big),$$

where we set $q = CT^b$ for some small $b > 0$ (specifically, $b < \frac{1}{4}$) as in the discussion of the auxiliary results. In addition,

$$\begin{aligned} B_T^{\ell+}(f) &= \frac{1}{\sqrt{T}} \sum_{t=\ell+1}^{\lceil \frac{\ell}{q} \rceil q} \big(f(X_{t,T}) - \mathbb{E} f(X_{t,T}) \big) \\ B_T^{j-}(f) &= \frac{1}{\sqrt{T}} \sum_{t=\lfloor \frac{j\delta T}{q} \rfloor q + 1}^{\lfloor j\delta T \rfloor} \big(f(X_{t,T}) - \mathbb{E} f(X_{t,T}) \big) \end{aligned}$$

denote the first and the last block, respectively. With this notation at hand, we obtain

$$
\begin{aligned}
P_{T,j}(\delta, 6\varepsilon) \leq\; & \mathbb{P}\Big(\max_{\lfloor (j-1)\delta T \rfloor \leq \ell \leq \lfloor j\delta T \rfloor} \sup_{f\in\mathcal{F}} \Big| \sum_{r=\lceil \frac{\ell}{q} \rceil + 1}^{\lfloor \frac{j\delta T}{q} \rfloor} B_{r,T}(f) \Big| > 4\varepsilon \Big) \\
& + \mathbb{P}\Big(\max_{\lfloor (j-1)\delta T \rfloor \leq \ell \leq \lfloor j\delta T \rfloor} \sup_{f\in\mathcal{F}} |B_T^{\ell +}(f)| > \varepsilon \Big) \\
& + \mathbb{P}\Big(\sup_{f\in\mathcal{F}} |B_T^{j-}(f)| > \varepsilon \Big) \\
=:\; & P_{T,j,1}(\delta, 4\varepsilon) + P_{T,j,2}(\delta, \varepsilon) + P_{T,j,3}(\delta, \varepsilon).
\end{aligned}
$$

The terms $P_{T,j,2}$ and $P_{T,j,3}$ can be bounded by fairly straightforward arguments: Applying a maximal inequality (see e.g. Section 2.1.3 in van der Vaart and Wellner (1996)), we get that

$$
\begin{aligned}
& \Big\| \max_{\lfloor (j-1)\delta T \rfloor \leq \ell \leq \lfloor j\delta T \rfloor} \sup_{f\in\mathcal{F}} |B_T^{\ell +}(f)| \Big\|_p \\
& \qquad \leq C(\delta T)^{1/p} \max_{\lfloor (j-1)\delta T \rfloor \leq \ell \leq \lfloor j\delta T \rfloor} \big\| \sup_{f\in\mathcal{F}} |B_T^{\ell +}(f)| \big\|_p .
\end{aligned}
$$

Moreover,

$$
\sup_{f\in\mathcal{F}} |B_T^{\ell +}(f)| \leq \frac{2}{\sqrt{T}} \sum_{t=\ell+1}^{\lceil \frac{\ell}{q} \rceil q} F(X_{t,T})
$$

and by the moment conditions on the envelope F in (C3), it holds that $\| \sup_{f\in\mathcal{F}} |B_T^{\ell +}(f)| \|_p \leq Cq/\sqrt{T}$. Hence, by Markov's inequality,

$$
\begin{aligned}
P_{T,j,2}(\delta, \varepsilon) & \leq \varepsilon^{-p} \Big\| \max_{\lfloor (j-1)\delta T \rfloor \leq \ell \leq \lfloor j\delta T \rfloor} \sup_{f\in\mathcal{F}} |B_T^{\ell +}(f)| \Big\|_p^p \\
& \leq C\delta T \Big(\frac{q}{\varepsilon \sqrt{T}} \Big)^p = o(1)
\end{aligned}
$$

for $T \to \infty$ given that $q = T^b$ with $b < \frac{1}{4}$. By analogous considerations, $P_{T,j,3}(\delta, \varepsilon)$ can be bounded in the same way. To deal with $P_{T,j,1}$, we split it up into two parts:

$$
P_{T,j,1}(\delta, 4\varepsilon) \leq \Delta_T^{(0)} + \Delta_T^{(1)}
$$

with

$$\Delta_T^{(0)} = \mathbb{P}\Big(\max_{\lfloor \frac{(j-1)\delta T}{2q} \rfloor \le k \le \lceil \frac{j\delta T}{2q} \rceil} \sup_{f \in \mathcal{F}} \Big| \sum_{r=k}^{\lfloor \frac{j\delta T}{2q} \rfloor} B_{2r,T}(f) \Big| > 2\varepsilon \Big)$$

$$\Delta_T^{(1)} = \mathbb{P}\Big(\max_{\lfloor \frac{(j-1)\delta T}{2q} \rfloor \le k \le \lceil \frac{j\delta T}{2q} \rceil} \sup_{f \in \mathcal{F}} \Big| \sum_{r=k}^{\lceil \frac{j\delta T}{2q} \rceil} B_{2r-1,T}(f) \Big| > 2\varepsilon \Big).$$

As the two terms can be treated in the same way, we restrict ourselves to $\Delta_T^{(1)}$. Applying a version of Ottaviani's inequality for α-mixing processes (which has the form stated in Chapter 10.2 of Lin and Bai (2010) and can be proven by the arguments therein), we obtain that

$$\text{(S.14)} \quad \Delta_T^{(1)} \le \frac{\mathbb{P}\Big(\sup_{f \in \mathcal{F}} \Big| \sum_{r=\lfloor \frac{(j-1)\delta T}{2q} \rfloor}^{\lceil \frac{j\delta T}{2q} \rceil} B_{2r-1,T}(f) \Big| > \varepsilon \Big) + \frac{\delta T}{2q} \alpha(q)}{1 - \max_{\lfloor \frac{(j-1)\delta T}{2q} \rfloor \le k \le \lceil \frac{j\delta T}{2q} \rceil} \mathbb{P}\Big(\sup_{f \in \mathcal{F}} \Big| \sum_{r=\lfloor \frac{(j-1)\delta T}{2q} \rfloor}^{k} B_{2r-1,T}(f) \Big| > \varepsilon \Big)}.$$

In order to bound the right-hand side of (S.14), we make use of the random variables

$$Q_{r,T}(f) = \frac{1}{\sqrt{(k'-k+1)q}} \sum_{t=(2r-2)q+1}^{(2r-1)q \wedge T} \big(f(X_{t,T}) - \mathbb{E} f(X_{t,T}) \big)$$

and $W_T(k,k') = \sup_{f \in \mathcal{F}} | \sum_{r=k}^{k'} Q_{r,T}(f)|$, which we have introduced at the beginning of the section when discussing the auxiliary results. Combining Lemmas A.1 and A.2 and noting that $\int_0^{\text{diam}(\mathcal{F})} \mathcal{N}(w/2, \mathcal{F}, d)^{1/p} dw$ is finite by assumption (C3), we get that $\mathbb{E}\big[|W_T(k,k')|^p\big] \le C < \infty$ for some sufficiently large constant C. This implies that

$$\begin{aligned} \mathbb{P}\Big(\sup_{f \in \mathcal{F}} \Big| \sum_{r=k}^{k'} B_{2r-1,T}(f) \Big| > \varepsilon \Big) &= \mathbb{P}\Big(W_T(k,k') > \frac{\varepsilon \sqrt{T}}{\sqrt{(k'-k+1)q}} \Big) \\ &\le \mathbb{E}\big[|W_T(k,k')|^p\big] \Big(\frac{(k'-k+1)q}{\varepsilon^2 T} \Big)^{p/2} \\ &\le C \Big(\frac{(k'-k+1)q}{\varepsilon^2 T} \Big)^{p/2}. \end{aligned}$$

Specifically, whenever $(k - k' + 1)q \leq \delta T$,

$$\mathbb{P}\Big(\sup_{f \in \mathcal{F}}\Big|\sum_{r=k}^{k'} B_{2r-1,T}(f)\Big| > \varepsilon\Big) \leq C\frac{\delta^{p/2}}{\varepsilon^p}. \tag{S.15}$$

With (S.15), it is easy to see that the denominator in (S.14) is bounded away from zero for δ sufficiently small and to infer that

$$\Delta_T^{(1)} \leq C\Big(\frac{\delta^{p/2}}{\varepsilon^p} + \frac{\delta T}{2q}\alpha(q)\Big).$$

Using an analogous bound for the term $\Delta_T^{(0)}$, it follows that

$$P_T(\delta,\varepsilon) \leq \sum_{j=1}^{\lceil 1/\delta \rceil} P_{T,j}(\delta,\varepsilon) \leq C\Big\lceil\frac{1}{\delta}\Big\rceil\Big(\frac{\delta^{p/2}}{\varepsilon^p} + \frac{\delta T}{2q}\alpha(q) + \delta T\Big(\frac{q}{\varepsilon\sqrt{T}}\Big)^p\Big).$$

This yields that $\lim_{\delta\searrow 0}\limsup_{T\to\infty} P_T(\delta,\varepsilon) = 0$ and the assertion (S.13) follows. By the discussion at the beginning of this proof we obtain (S.9), which implies stochastic equicontinuity. □

Proof of Theorem 5.3. The proof is an immediate consequence of the following two statements:

$$\mathbb{P}\big(\hat{u}_0(\tau_T) < u_0\big) = o(1) \tag{S.16}$$

$$\mathbb{P}\big(\hat{u}_0(\tau_T) > u_0 + K\gamma_T\big) = o(1) \tag{S.17}$$

for some sufficiently large constant $K > 0$.

Proof of (S.16). It holds that

$$\begin{aligned}\mathbb{P}\big(\hat{u}_0(\tau_T) < u_0\big) &\leq \mathbb{P}\Big(\sqrt{T}\hat{\mathcal{D}}_T(u) > \tau_T \text{ for some } u < u_0\Big)\\ &\leq \mathbb{P}\Big(\sqrt{T}\mathcal{D}(u) + \hat{\mathcal{H}}_T(u) > \tau_T \text{ for some } u < u_0\Big)\\ &\leq \mathbb{P}\Big(\sup_{u\in[0,1]}\hat{\mathcal{H}}_T(u) > \tau_T\Big),\end{aligned}$$

where the second inequality follows from the fact that $\sqrt{T}\hat{\mathcal{D}}_T(u) \leq \sqrt{T}\mathcal{D}(u) + \hat{\mathcal{H}}_T(u)$ and the third one exploits the fact that $\mathcal{D}(u) = 0$ at points $u < u_0$. From Corollary 5.2, we know that $\sup_{u\in[0,1]}\hat{\mathcal{H}}_T(u) = \hat{\mathbb{H}}_T(1)$ converges in distribution to $\mathbb{H}(1)$. Moreover, the distribution function F of $\mathbb{H}(1)$ is continuous on $[0,\infty)$ by the results of Section 3 in Lifshits (1982). We can thus infer that the distribution function F_T of $\hat{\mathbb{H}}_T(1)$ uniformly converges to F on $[0,\infty)$. As a result, we obtain that

$$\mathbb{P}\big(\hat{\mathbb{H}}_T(1) > \tau_T\big) = 1 - F_T(\tau_T) = [1 - F(\tau_T)] + [F(\tau_T) - F_T(\tau_T)] = o(1),$$

which in turn yields (S.16). □

Proof of (S.17). Similarly as above, we can write

$$\begin{aligned}
&\mathbb{P}\big(\hat{u}_0(\tau_T) > u_0 + K\gamma_T\big) \\
&\le \mathbb{P}\Big(\sqrt{T}\hat{\mathcal{D}}_T(u) \le \tau_T \text{ for some } u > u_0 + K\gamma_T\Big) \\
&\le \mathbb{P}\Big(\sqrt{T}\mathcal{D}(u) - \hat{\mathcal{H}}_T(u) \le \tau_T \text{ for some } u > u_0 + K\gamma_T\Big),
\end{aligned}$$

the last line following from the fact that $\sqrt{T}\mathcal{D}(u) - \hat{\mathcal{H}}_T(u) \le \sqrt{T}\hat{\mathcal{D}}_T(u)$. Next notice that

$$\min_{u\in[u_0+K\gamma_T,1]} \mathcal{D}(u) \ge \frac{c_\kappa(K\gamma_T)^\kappa}{2}$$

for sufficiently large T, which follows upon inspection of (5.1). This allows us to infer that

$$\begin{aligned}
&\mathbb{P}\Big(\sqrt{T}\mathcal{D}(u) - \hat{\mathcal{H}}_T(u) \le \tau_T \text{ for some } u > u_0 + K\gamma_T\Big) \\
&\le \mathbb{P}\Big(\frac{\sqrt{T}c_\kappa(K\gamma_T)^\kappa}{2} - \hat{\mathbb{H}}_T(1) \le \tau_T\Big) \le P_1 + P_2,
\end{aligned}$$

where

$$\begin{aligned}
P_1 &= \mathbb{P}\Big(\frac{\sqrt{T}c_\kappa(K\gamma_T)^\kappa}{2} - \hat{\mathbb{H}}_T(1) \le \tau_T, \hat{\mathbb{H}}_T(1) \le b_T\Big) \\
P_2 &= \mathbb{P}\Big(\hat{\mathbb{H}}_T(1) > b_T\Big)
\end{aligned}$$

and b_T is some diverging sequence of positive numbers satisfying $b_T/\tau_T \to 0$. As already seen in the proof of (S.16), it holds that $P_2 = o(1)$. Moreover, $P_1 = 0$ for sufficiently large T if we set $\gamma_T = (\tau_T/\sqrt{T})^{1/\kappa}$ and choose K to be sufficiently large. This shows (S.17). □

Proof of Theorem 5.4. We first derive (5.12) which says that

$$\mathbb{P}\big(\hat{u}_0(\tau_\alpha) < u_0\big) \le \alpha + o(1).$$

It holds that

$$\begin{aligned}
\mathbb{P}\big(\hat{u}_0(\tau_\alpha) < u_0\big) &\le \mathbb{P}\big(\sqrt{T}\hat{\mathcal{D}}_T(u) > \tau_\alpha \text{ for some } u < u_0\big) \\
&\le \mathbb{P}\big(\sqrt{T}\hat{\mathbb{D}}_T(u_0) > \tau_\alpha\big) \\
&= \mathbb{P}\big(\hat{\mathbb{H}}_T(u_0) > \tau_\alpha\big).
\end{aligned}$$

We now make use of the following fact which is a direct consequence of the results from Section 3 in Lifshits (1982):

($*$) For each u, the random variable

$$\mathbb{H}(u) = \sup_{f \in \mathcal{F}} \sup_{0 \le w \le v \le u} |H(v, w, f)|$$

has a distribution function which is continuous on $[0, \infty)$.

By ($*$), we obtain that

$$\begin{aligned} \mathbb{P}\big(\hat{\mathbb{H}}_T(u_0) > \tau_\alpha\big) &= \mathbb{P}\big(\mathbb{H}(u_0) > \tau_\alpha\big) + \Big[\mathbb{P}\big(\hat{\mathbb{H}}_T(u_0) > \tau_\alpha\big) - \mathbb{P}\big(\mathbb{H}(u_0) > \tau_\alpha\big)\Big] \\ &= \mathbb{P}\big(\mathbb{H}(u_0) > \tau_\alpha\big) + o(1) = \alpha + o(1), \end{aligned}$$

where the last equality is due to the fact that $\tau_\alpha = q_\alpha(u_0)$ is the $(1-\alpha)$-quantile of $\mathbb{H}(u_0)$. From this, (5.12) immediately follows. The statement (5.13) can be proven by the same arguments as for (S.17) in the proof of Theorem 5.3. □

Proof of Corollary 5.5. Let $q_\alpha(u_n)$ be the $(1-\alpha)$-quantile of $\mathbb{H}(u_n)$ and $q_\alpha(u)$ the corresponding quantile of $\mathbb{H}(u)$. We first show that for any α with $0 < \alpha < 1$,

$$q_\alpha(u_n) \to q_\alpha(u) \tag{S.18}$$

as $u_n \to u$. To do so, let d denote the natural semimetric on $\Delta \times \mathcal{F}$ introduced in Subsection 5.2. Moreover, let $\mathcal{C}_u(\Delta \times \mathcal{F}, d)$ be the space of uniformly continuous functions on $(\Delta \times \mathcal{F}, d)$ and define the functionals

$$\begin{aligned} M_n(x) = M_{u_n}(x) &= \sup_{f \in \mathcal{F}} \sup_{0 \le w \le v \le u_n} |x(v, w, f)| \\ M(x) = M_u(x) &= \sup_{f \in \mathcal{F}} \sup_{0 \le w \le v \le u} |x(v, w, f)| \end{aligned}$$

for $x \in \mathcal{C}_u(\Delta \times \mathcal{F}, d)$. Elementary arguments show that

$$M(x) = \lim_{n \to \infty, y \to x} M_n(y),$$

where x and y are elements of $\mathcal{C}_u(\Delta \times \mathcal{F}, d)$. Using this together with the extended continuous mapping theorem (see e.g. Theorem 1.11.1 in van der Vaart and Wellner (1996)), we obtain that

$$M_n(H) \xrightarrow{d} M(H),$$

or put differently,

$$\mathbb{H}(u_n) \xrightarrow{d} \mathbb{H}(u),$$

since $M_n(H) = \mathbb{H}(u_n)$ and $M(H) = \mathbb{H}(u)$. From this, we can infer that the quantile functions converge as well, thus arriving at (S.18).

Next let $\tilde{u}_0$ be a consistent estimator of u_0. By (S.18), the quantile function $q_\alpha(\cdot)$ is continuous at each point u, in particular at u_0. Hence,

$$\hat{\tau}_\alpha = q_\alpha(\tilde{u}_0) \xrightarrow{P} \tau_\alpha = q_\alpha(u_0). \tag{S.19}$$

Moreover,

$$\begin{aligned}
\mathbb{P}\big(\hat{u}_0(\hat{\tau}_\alpha) < u_0\big) &\le \mathbb{P}\big(\sqrt{T}\hat{\mathcal{D}}_T(u) > \hat{\tau}_\alpha \text{ for some } u < u_0\big) \\
&\le \mathbb{P}\big(\sqrt{T}\hat{\mathbb{D}}_T(u_0) > \hat{\tau}_\alpha\big) \\
&= \mathbb{P}\big(\hat{\mathbb{H}}_T(u_0) > \hat{\tau}_\alpha\big).
\end{aligned}$$

Since $\hat{\mathbb{H}}_T(u_0) \xrightarrow{d} \mathbb{H}(u_0)$ and the distribution function of $\mathbb{H}(u_0)$ is continuous on $[0, \infty)$ by $(*)$, the distribution function of $\hat{\mathbb{H}}_T(u_0)$ uniformly converges to that of $\mathbb{H}(u_0)$ on $[0, \infty)$. Hence,

$$\begin{aligned}
\mathbb{P}\big(\hat{\mathbb{H}}_T(u_0) > \hat{\tau}_\alpha\big) &= \mathbb{P}\big(\mathbb{H}(u_0) > \hat{\tau}_\alpha\big) + \Big[\mathbb{P}\big(\hat{\mathbb{H}}_T(u_0) > \hat{\tau}_\alpha\big) - \mathbb{P}\big(\mathbb{H}(u_0) > \hat{\tau}_\alpha\big)\Big] \\
&= \mathbb{P}\big(\mathbb{H}(u_0) > \hat{\tau}_\alpha\big) + o(1).
\end{aligned}$$

Finally, as $\hat{\tau}_\alpha = \tau_\alpha + o_p(1)$ and the distribution function of $\mathbb{H}(u_0)$ is continuous by $(*)$, we obtain that

$$\mathbb{P}\big(\mathbb{H}(u_0) > \hat{\tau}_\alpha\big) = \mathbb{P}\big(\mathbb{H}(u_0) > \tau_\alpha\big) + o(1) = \alpha + o(1).$$

This completes the proof of (5.14). The statement (5.15) can again be shown by the same arguments as for (S.17) in the proof of Theorem 5.3. □

Proof of (5.11). To start with, write

$$\begin{aligned}
\mathrm{MSE}_1(\tau_T) = \mathbb{E}\Big[\Big\{ &\int_{u_0}^1 1(\sqrt{T}\hat{\mathcal{D}}_T(u) \le \tau_T) du \\
&- \int_0^{u_0} \big[1 - 1(\sqrt{T}\hat{\mathcal{D}}_T(u) \le \tau_T)\big] du\Big\}^2 1(\hat{\mathbb{H}}_T(1) \le b_T)\Big]
\end{aligned}$$

and note that

$$\sqrt{T}\mathcal{D}(u) - \hat{\mathbb{H}}_T(1) \le \sqrt{T}\hat{\mathcal{D}}_T(u) \le \sqrt{T}\mathcal{D}(u) + \hat{\mathbb{H}}_T(1). \tag{S.20}$$

On the event that $\hat{\mathbb{H}}_T(1) \le b_T$, it holds that

$$0 \le \int_0^{u_0} \big[1 - 1(\sqrt{T}\hat{\mathcal{D}}_T(u) \le \tau_T)\big] du \le \int_0^{u_0} \big[1 - 1(\sqrt{T}\mathcal{D}(u) + b_T \le \tau_T)\big] du = \int_0^{u_0} \big[1 - 1(b_T \le \tau_T)\big] du = 0 \tag{S.21}$$

for sufficiently large T, where the second line follows by (S.20) and the third one uses the fact that $\mathcal{D}(u) = 0$ for $u \le u_0$. (S.21) immediately implies that

$$\text{MSE}_1(\tau_T) = \mathbb{E}\Big[\Big\{\int_{u_0}^1 1(\sqrt{T}\hat{\mathcal{D}}_T(u) \le \tau_T) du\Big\}^2 1(\hat{\mathbb{H}}_T(1) \le b_T)\Big] \tag{S.22}$$

for sufficiently large T. We now derive an upper and lower bound for the right-hand side of (S.22): Using (S.20) and borrowing some arguments from the proof of Theorem 5.3, we obtain that

$$\text{MSE}_1(\tau_T) \le \Big\{\int_{u_0}^1 1(\sqrt{T}\mathcal{D}(u) \le \tau_T + b_T) du\Big\}^2 = \Big\{\int_{u_0}^{u_0 + \overline{C}\gamma_T} 1(\sqrt{T}\mathcal{D}(u) \le \tau_T + b_T) du\Big\}^2 \le (\overline{C}\gamma_T)^2$$

for some sufficiently large constant $\overline{C}$ and large enough sample sizes T. Similarly,

$$\text{MSE}_1(\tau_T) \ge \Big\{\int_{u_0}^{u_0 + \underline{C}\gamma_T} 1(\sqrt{T}\mathcal{D}(u) \le \tau_T - b_T) du\Big\}^2 \ge (\underline{C}\gamma_T)^2$$

for a sufficiently small constant $\underline{C}$ and large enough T, since by (5.1), $\max_{u \in [u_0, u_0 + \underline{C}\gamma_T]} \mathcal{D}(u) \le 2c_\kappa(\underline{C}\gamma_T)^\kappa$ and thus for $u \in [u_0, u_0 + \underline{C}\gamma_T]$,

$$1(\sqrt{T}\mathcal{D}(u) \le \tau_T - b_T) \ge 1(2c_\kappa \underline{C}^\kappa \tau_T \le \tau_T - b_T) = 1$$

for large enough T and sufficiently small $\underline{C}$. □

REFERENCES

BOSQ, D. (1996). *Nonparametric statistics for stochastic processes.* New York, Springer.

LIEBSCHER, E. (1996). Central limit theorems for sums of α-mixing random variables. *Stochastics and Stochastic Reports* **59** 241-258.

LIFSHITS, M. A. (1982). On the absolute continuity of distributions of functionals of random processes. *Theory of Probability & Its Applications* **27** 600-607.

LIN, Z. and BAI, Z. (2010). *Probability inequalities.* New York, Springer.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak convergence and empirical processes.* New York, Springer.

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITÄT KONSTANZ
78457 KONSTANZ, GERMANY
E-MAIL: michael.vogt@uni-konstanz.de

DEPARTMENT OF MATHEMATICS
RUHR-UNIVERSITÄT BOCHUM
44780 BOCHUM, GERMANY
E-MAIL: holger.dette@ruhr-uni-bochum.de