0.1 Theoretical properties of the test

To start with, we introduce the auxiliary statistic

$$\widehat{\Phi}_{n,T} = \max_{1 \le i < j \le n} \widehat{\Phi}_{ij,T}, \tag{0.1}$$

where

$$\widehat{\Phi}_{ij,T} = \max_{(u,h) \in \mathcal{G}_T} \left\{ \left| \frac{\widehat{\phi}_{ij,T}(u,h)}{\{\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\}^{1/2}} \right| - \lambda(h) \right\}$$

and $\widehat{\phi}_{ij,T}(u,h) = \sum_{t=1}^{T} w_{t,T}(u,h) \{ (\varepsilon_{it} - \bar{\varepsilon}_i) + \boldsymbol{\beta}_i^\top (\mathbf{X}_{it} - \bar{\mathbf{X}}_i) - (\varepsilon_{jt} - \bar{\varepsilon}_j) - \boldsymbol{\beta}_j^\top (\mathbf{X}_{jt} - \bar{\mathbf{X}}_j) \}$ with $\bar{\varepsilon}_i = \bar{\varepsilon}_{i,T} = T^{-1} \sum_{t=1}^{T} \varepsilon_{it}$ and $\bar{\mathbf{X}}_i = \bar{\mathbf{X}}_{i,T} = T^{-1} \sum_{t=1}^{T} \mathbf{X}_{it}$ respectively. Our first theoretical result characterizes the asymptotic behaviour of the statistic $\widehat{\Phi}_{n,T}$.

Theorem 0.1. *Suppose that the error processes $\mathcal{E}_i = \{\varepsilon_{it} : 1 \le t \le T\}$ are independent across i and satisfy* **??**–**??** *for each i. Moreover, let* **??**–**??** *be fulfilled and assume that $\widehat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$ with $\rho_T = o(1/\log T)$ for each i. Then*

$$\mathbb{P}\big(\widehat{\Phi}_{n,T} \le q_{n,T}(\alpha) | \{\mathbf{X}_{it} : 1 \le t \le T, 1 \le i \le n\}\big) = (1-\alpha) + o(1) \ a.s.$$

1 Proof of the Theorem 0.1

In this section, we prove the theoretical results from Section **??**. We use the following notation: The symbol C denotes a universal real constant which may take a different value on each occurrence. For $a, b \in \mathbb{R}$, we write $a_+ = \max\{0, a\}$ and $a \vee b = \max\{a, b\}$. For any set A, the symbol $|A|$ denotes the cardinality of A. The notation $X \overset{\mathcal{D}}{=} Y$ means that the two random variables X and Y have the same distribution. Finally, $f_0(\cdot)$ and $F_0(\cdot)$ denote the density and distribution function of the standard normal distribution, respectively.

Auxiliary results using strong approximation theory

The main purpose of this section is to prove that there is a version of the multiscale statistic $\widehat{\Phi}_{n,T}$ defined in (0.1) which is close to a Gaussian statistic whose distribution is known. More specifically, we prove the following result.

Proposition A.1. *Under the conditions of Theorem 0.1, there exist statistics $\widetilde{\Phi}_{n,T}$ for $T = 1, 2, \ldots$ with the following two properties: (i) $\widetilde{\Phi}_{n,T}$ has the same distribution as $\widehat{\Phi}_{n,T}$ for any T, and (ii)*

$$\big|\widetilde{\Phi}_{n,T} - \Phi_{n,T}\big| = o_p\Big(\frac{T^{1/q}}{\sqrt{Th_{\min}}} + \rho_T \sqrt{\log T} + \rho_T \max_{1 \le i \le n} \Big| \sum_{s=1}^{T} \widehat{\boldsymbol{\beta}}_i^\top (\mathbf{X}_{is} - \bar{\mathbf{X}}_i) \Big| +$$

$$+ \frac{T^{-1/2}}{\sqrt{Th_{min}}} \max_{1 \le i \le n} \max_{1 \le t \le T} \Big| \sum_{s=1}^{t} (\mathbf{X}_{is} - \bar{\mathbf{X}}_i) \Big| \Big),$$

*where $\Phi_{n,T}$ is a Gaussian statistic as defined in (**??**).*

Proof of Proposition A.1. For the proof, we draw on strong approximation theory for each stationary process $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$ that fulfill the conditions **??**–**??**. By Theorem 2.1 and Corollary 2.1 in Berkes et al. (2014), the following strong approximation result holds true: On a richer probability space, there exist a standard Brownian motion $\mathbb{B}$ and a sequence $\{\widetilde{\varepsilon}_t : t \in \mathbb{N}\}$ such that $[\widetilde{\varepsilon}_1, \ldots, \widetilde{\varepsilon}_T] \overset{\mathcal{D}}{=} [\varepsilon_1, \ldots, \varepsilon_T]$ for each T and

$$\max_{1 \leq t \leq T} \Big| \sum_{s=1}^{t} \widetilde{\varepsilon}_s - \sigma \mathbb{B}(t) \Big| = o\big(T^{1/q}\big) \quad \text{a.s.,} \tag{1.1}$$

where $\sigma^2 = \sum_{k \in \mathbb{Z}} \text{Cov}(\varepsilon_0, \varepsilon_k)$ denotes the long-run error variance.

To apply this result, we define

$$\widetilde{\Phi}_{n,T} = \max_{1 \leq i < j \leq n} \widetilde{\Phi}_{ij,T}, \tag{1.2}$$

where

$$\widetilde{\Phi}_{ij,T} = \max_{(u,h) \in \mathcal{G}_T} \left\{ \left| \frac{\widetilde{\phi}_{ij,T}(u,h)}{\{\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\}^{1/2}} \right| - \lambda(h) \right\},$$

where $\widetilde{\phi}_{ij,T}(u,h) = \sum_{t=1}^{T} w_{t,T}(u,h)\big\{(\widetilde{\varepsilon}_{it} - \bar{\widetilde{\varepsilon}}_i) + \boldsymbol{\beta}_i^\top(\mathbf{X}_{it} - \bar{\mathbf{X}}_i) - (\widetilde{\varepsilon}_{jt} - \bar{\widetilde{\varepsilon}}_j) - \boldsymbol{\beta}_j^\top(\mathbf{X}_{jt} - \bar{\mathbf{X}}_j)\big\}$ and $\widetilde{\sigma}_i^2$ are the same estimators as $\widehat{\sigma}_i^2$ with $Y_{it} = (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top \mathbf{X}_{it} + m_i(t/T) + (\alpha_i - \widehat{\alpha}_i) + \varepsilon_{it}$ replaced by $\widetilde{Y}_{t,T} = (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top \mathbf{X}_{it} + m_i(t/T) + (\alpha_i - \widehat{\alpha}_i) + \widetilde{\varepsilon}_{it}$ for $1 \leq t \leq T$. In addition, we let

$$\Phi_{n,T} = \max_{1 \leq i < j \leq n} \Phi_{ij,T} = \max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \left\{ \left| \frac{\phi_{ij,T}(u,h)}{\{\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\}^{1/2}} \right| - \lambda(h) \right\}$$

$$\Phi_{n,T}^\diamond = \max_{1 \leq i < j \leq n} \Phi_{ij,T}^\diamond = \max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \left\{ \left| \frac{\phi_{ij,T}(u,h)}{\{\widetilde{\sigma}_i^2 + \widetilde{\sigma}_j^2\}^{1/2}} \right| - \lambda(h) \right\}$$

with

$$\phi_{ij,T}(u,h) = \sum_{t=1}^{T} w_{t,T}(u,h)\big\{\widehat{\sigma}_i(Z_{it} - \bar{Z}_i) + \widehat{\boldsymbol{\beta}}_i^\top(\mathbf{X}_{it} - \bar{\mathbf{X}}_i) - \widehat{\sigma}_j(Z_{jt} - \bar{Z}_j) - \widehat{\boldsymbol{\beta}}_j^\top(\mathbf{X}_{jt} - \bar{\mathbf{X}}_j)\big\}$$

and $Z_{it} = \mathbb{B}_i(t) - \mathbb{B}_i(t-1)$. With this notation, we can write

$$\big|\widetilde{\Phi}_{n,T} - \Phi_{n,T}\big| \leq \big|\widetilde{\Phi}_{n,T} - \Phi_{n,T}^\diamond\big| + \big|\Phi_{n,T}^\diamond - \Phi_{n,T}\big|. \tag{1.3}$$

First consider $\big|\widetilde{\Phi}_{n,T} - \Phi_{n,T}^\diamond\big|$. Straightforward calculations yield that

$$\big|\widetilde{\Phi}_{n,T} - \Phi_{n,T}^\diamond\big| \leq \{\widetilde{\sigma}_i^2 + \widetilde{\sigma}_j^2\}^{-1/2} \max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \big|\widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)\big|. \tag{1.4}$$

Using summation by parts, $(\sum_{i=1}^{n} a_i b_i = \sum_{i=1}^{n-1} A_i(b_i - b_{i+1}) + A_n b_n$ with $A_j = \sum_{j=1}^{i} a_j)$ we further obtain that

$$\big|\widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)\big| =$$

$$= \Big| \sum_{t=1}^{T} w_{t,T}(u,h)\big\{(\widetilde{\varepsilon}_{it} - \bar{\widetilde{\varepsilon}}_i) + \boldsymbol{\beta}_i^\top(\mathbf{X}_{it} - \bar{\mathbf{X}}_i) - (\widetilde{\varepsilon}_{jt} - \bar{\widetilde{\varepsilon}}_j) - \boldsymbol{\beta}_j^\top(\mathbf{X}_{jt} - \bar{\mathbf{X}}_j) -$$

$$- \widehat{\sigma}_i(Z_{it} - \bar{Z}_i) - \widehat{\boldsymbol{\beta}}_i^\top(\mathbf{X}_{it} - \bar{\mathbf{X}}_i) - \widehat{\sigma}_j(Z_{jt} - \bar{Z}_j) + \widehat{\boldsymbol{\beta}}_j^\top(\mathbf{X}_{jt} - \bar{\mathbf{X}}_j)\big\} \Big| =$$

$$= \Big| \sum_{t=1}^{T-1} A_{ij,t}\big(w_{t,T}(u,h) - w_{t+1,T}(u,h)\big) + A_{ij,T} w_{t,T}(u,h) \Big|,$$

where

$$A_{ij,t} = \sum_{s=1}^{t} \left\{ (\widetilde{\varepsilon}_{is} - \bar{\widetilde{\varepsilon}}_i) + \boldsymbol{\beta}_i^\top (\mathbf{X}_{is} - \bar{\mathbf{X}}_i) - (\widetilde{\varepsilon}_{js} - \bar{\widetilde{\varepsilon}}_j) - \boldsymbol{\beta}_j^\top (\mathbf{X}_{js} - \bar{\mathbf{X}}_j) - \right.$$
$$\left. - \widehat{\sigma}_i(Z_{it} - \bar{Z}_i) - \widehat{\boldsymbol{\beta}}_i^\top (\mathbf{X}_{it} - \bar{\mathbf{X}}_i) - \widehat{\sigma}_j(Z_{jt} - \bar{Z}_j) + \widehat{\boldsymbol{\beta}}_j^\top (\mathbf{X}_{jt} - \bar{\mathbf{X}}_j) \right\}.$$

Note that by construction $A_{ij,T} = 0$ for all pairs (i,j). Denoting

$$W_T(u, h) = \sum_{t=1}^{T-1} |w_{t+1,T}(u, h) - w_{t,T}(u, h)|,$$

we have

$$\left| \widetilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h) \right| = \left| \sum_{t=1}^{T-1} A_{ij,t} \big(w_{t,T}(u,h) - w_{t+1,T}(u,h) \big) \right| \leq W_T(u,h) \max_{1 \leq t \leq T} |A_{ij,t}|.$$
$$(1.5)$$

Now consider $\max_{1 \leq t \leq T} |A_{ij,t}|$:

$$\max_{1 \leq t \leq T} |A_{ij,t}| \leq \max_{1 \leq t \leq T} \left| \sum_{s=1}^{t} (\widetilde{\varepsilon}_{is} - \bar{\widetilde{\varepsilon}}_i) - \sigma_i \sum_{s=1}^{t} \left\{ Z_{is} - \bar{Z}_i \right\} \right| + \max_{1 \leq t \leq T} \left| \sum_{s=1}^{t} (\widetilde{\varepsilon}_{js} - \bar{\widetilde{\varepsilon}}_j) - \sigma_j \sum_{s=1}^{t} \left\{ Z_{js} - \bar{Z}_j \right\} \right| +$$

$$+ \max_{1 \leq t \leq T} \left| \sum_{s=1}^{t} (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top (\mathbf{X}_{is} - \bar{\mathbf{X}}_i) \right| + \max_{1 \leq t \leq T} \left| \sum_{s=1}^{t} (\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j)^\top (\mathbf{X}_{js} - \bar{\mathbf{X}}_j) \right| \leq$$

$$\leq \max_{1 \leq t \leq T} \left| \sum_{s=1}^{t} \widetilde{\varepsilon}_{is} - \sigma_i \sum_{s=1}^{t} Z_{is} \right| + \max_{1 \leq t \leq T} \left| \sum_{s=1}^{t} \widetilde{\varepsilon}_{js} - \sigma_j \sum_{s=1}^{t} Z_{js} \right| +$$

$$+ \max_{1 \leq t \leq T} \left| t(\bar{\widetilde{\varepsilon}}_i - \sigma_i \bar{Z}_i) \right| + \max_{1 \leq t \leq T} \left| t(\bar{\widetilde{\varepsilon}}_j - \sigma_j \bar{Z}_j) \right| +$$

$$+ \max_{1 \leq t \leq T} \left| \sum_{s=1}^{t} (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top (\mathbf{X}_{is} - \bar{\mathbf{X}}_i) \right| + \max_{1 \leq t \leq T} \left| \sum_{s=1}^{t} (\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j)^\top (\mathbf{X}_{js} - \bar{\mathbf{X}}_j) \right| \leq$$

$$\leq 2 \max_{1 \leq t \leq T} \left| \sum_{s=1}^{t} \widetilde{\varepsilon}_{is} - \sigma_i \sum_{s=1}^{t} Z_{is} \right| + 2 \max_{1 \leq t \leq T} \left| \sum_{s=1}^{t} \widetilde{\varepsilon}_{js} - \sigma_j \sum_{s=1}^{t} Z_{js} \right| +$$

$$+ \max_{1 \leq t \leq T} \left| \sum_{s=1}^{t} (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top (\mathbf{X}_{is} - \bar{\mathbf{X}}_i) \right| + \max_{1 \leq t \leq T} \left| \sum_{s=1}^{t} (\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j)^\top (\mathbf{X}_{js} - \bar{\mathbf{X}}_j) \right| =$$

$$= 2 \max_{1 \leq t \leq T} \left| \sum_{s=1}^{t} \widetilde{\varepsilon}_{is} - \sigma_i \sum_{s=1}^{t} \big(\mathbb{B}_i(s) - \mathbb{B}_i(s-1) \big) \right| + 2 \max_{1 \leq t \leq T} \left| \sum_{s=1}^{t} \widetilde{\varepsilon}_{js} - \sigma_j \sum_{s=1}^{t} \big(\mathbb{B}_j(s) - \mathbb{B}_j(s-1) \big) \right|$$

$$+ \max_{1 \leq t \leq T} \left| \sum_{s=1}^{t} (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top (\mathbf{X}_{is} - \bar{\mathbf{X}}_i) \right| + \max_{1 \leq t \leq T} \left| \sum_{s=1}^{t} (\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j)^\top (\mathbf{X}_{js} - \bar{\mathbf{X}}_j) \right| =$$

$$= 2 \max_{1 \leq t \leq T} \left| \sum_{s=1}^{t} \widetilde{\varepsilon}_{is} - \sigma_i \mathbb{B}_i(t) \right| + 2 \max_{1 \leq t \leq T} \left| \sum_{s=1}^{t} \widetilde{\varepsilon}_{js} - \sigma_j \mathbb{B}_j(t) \right| +$$

$$+ \max_{1 \leq t \leq T} \left| (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top \sum_{s=1}^{t} (\mathbf{X}_{is} - \bar{\mathbf{X}}_i) \right| + \max_{1 \leq t \leq T} \left| (\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j)^\top \sum_{s=1}^{t} (\mathbf{X}_{js} - \bar{\mathbf{X}}_j) \right|$$

3

Now suppose that for all $1 \leq i \leq n$ we have $\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i = o_P(T^{-1/2})$ almost surely. Then applying the strong approximation result (1.1), we can infer that

$$\max_{1 \leq t \leq T} |A_{ij,t}| = o(T^{1/q}) + o_P(T^{-1/2}) \max_{1 \leq t \leq T} \Big| \sum_{s=1}^{t} (\mathbf{X}_{is} - \bar{\mathbf{X}}_i) \Big| + o_P(T^{-1/2}) \max_{1 \leq t \leq T} \Big| \sum_{s=1}^{t} (\mathbf{X}_{js} - \bar{\mathbf{X}}_j) \Big| \tag{1.6}$$

Standard arguments show that $\max_{(u,h) \in \mathcal{G}_T} W_T(u,h) = O(1/\sqrt{Th_{\min}})$. Plugging (1.6) in (1.5) and then in (1.4), we can thus infer that

$$\big| \widetilde{\Phi}_{n,T} - \Phi_{n,T}^{\diamond} \big| \leq \{ \widetilde{\sigma}_i^2 + \widetilde{\sigma}_j^2 \}^{-1/2} \max_{(u,h) \in \mathcal{G}_T} W_T(u,h) \max_{1 \leq i < j \leq n} \max_{1 \leq t \leq T} |A_{ij,t}| \leq$$

$$\leq o\Big(\frac{T^{1/q}}{\sqrt{Th_{\min}}} \Big) + o_P\Big(\frac{T^{-1/2}}{\sqrt{Th_{\min}}} \Big) \max_{1 \leq i \leq n} \max_{1 \leq t \leq T} \Big| \sum_{s=1}^{t} (\mathbf{X}_{is} - \bar{\mathbf{X}}_i) \Big|. \tag{1.7}$$

Now consider $|\Phi_{n,T}^{\diamond} - \Phi_{n,T}|$. Since $\phi_{ij,T}(u,h)$ conditionally on $\{\mathbf{X}_{it}\}$ is distributed as $N(\widehat{\boldsymbol{\beta}}_i^\top (\mathbf{X}_{it} - \bar{\mathbf{X}}_i) - \widehat{\boldsymbol{\beta}}_j^\top (\mathbf{X}_{jt} - \bar{\mathbf{X}}_j), \sigma_i^2 + \sigma_j^2)$ for all $(u,h) \in \mathcal{G}_T$ and all $1 \leq i < j \leq n$, $|\mathcal{G}_T| = O(T^\theta)$ for some large but fixed constant θ by Assumption (??), n is fixed and $\widehat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$ as well as $\widetilde{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$, we can establish that

$$\big| \Phi_{n,T}^{\diamond} - \Phi_{n,T} \big| \leq \max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \Big| \frac{\phi_{ij,T}(u,h)}{\{ \widetilde{\sigma}_i^2 + \widetilde{\sigma}_j^2 \}^{1/2}} - \frac{\phi_{ij,T}(u,h)}{\{ \widehat{\sigma}_i^2 + \widehat{\sigma}_j^2 \}^{1/2}} \Big| =$$

$$= o_P(\rho_T \sqrt{\log T}) + o_P(\rho_T) \max_{1 \leq i \leq n} \Big| \sum_{s=1}^{T} \widehat{\boldsymbol{\beta}}_i^\top (\mathbf{X}_{is} - \bar{\mathbf{X}}_i) \Big|. \tag{1.8}$$

Plugging (1.7) and (1.8) in (1.3) completes the proof. $\qquad \square$

Auxiliary results using anti-concentration bounds

In this section, we establish some properties of the Gaussian statistic $\Phi_{n,T}$ defined in (??). We in particular show that $\Phi_{n,T}$ does not concentrate too strongly in small regions of the form $[x - \delta_T, x + \delta_T]$ with δ_T converging to zero.

Proposition A.2. *Under the conditions of Theorem 0.1, it holds that*

$$\sup_{x \in \mathbb{R}} \mathbb{P}\Big(|\Phi_{n,T} - x| \leq \delta_T \big| \{ \mathbf{X}_{it} : 1 \leq i \leq n, 1 \leq t \leq T \} \Big) = o(1) \ a.s.,$$

where $\delta_T = T^{1/q}/\sqrt{Th_{\min}} + \rho_T \sqrt{\log T}$.

Proof of Proposition A.2. The main technical tool for proving Proposition A.2 are anti-concentration bounds for Gaussian random vectors. The following proposition slightly generalizes anti-concentration results derived in Chernozhukov et al. (2015), in particular Theorem 3 therein.

Proposition A.3. *Let $(X_1, \ldots, X_p)^\top$ be a Gaussian random vector in $\mathbb{R}^p$ with $\mathbb{E}[X_j] = \mu_j$ and $\mathrm{Var}(X_j) = \sigma_j^2 > 0$ for $1 \leq j \leq p$. Define $\overline{\mu} = \max_{1 \leq j \leq p} |\mu_j|$ together with $\underline{\sigma} = \min_{1 \leq j \leq p} \sigma_j$ and $\overline{\sigma} = \max_{1 \leq j \leq p} \sigma_j$. Moreover, set $a_p = \mathbb{E}[\max_{1 \leq j \leq p}(X_j - \mu_j)/\sigma_j]$ and $b_p = \mathbb{E}[\max_{1 \leq j \leq p}(X_j - \mu_j)]$. For every $\delta > 0$, it holds that*

$$\sup_{x \in \mathbb{R}} \mathbb{P}\Big(\big| \max_{1 \leq j \leq p} X_j - x \big| \leq \delta \Big) \leq C\delta \{ \overline{\mu} + a_p + b_p + \sqrt{1 \vee \log(\underline{\sigma}/\delta)} \},$$

where $C > 0$ depends only on $\underline{\sigma}$ and $\overline{\sigma}$.

The proof of Proposition A.3 is provided in Khismatullina and Vogt (2018). To apply Proposition A.3 to our setting at hand, we introduce the following notation: We write $x = (u, h)$ along with $\mathcal{G}_T = \{x : x \in \mathcal{G}_T\} = \{x_1, \ldots, x_p\}$, where $p := |\mathcal{G}_T| \leq O(T^\theta)$ for some large but fixed $\theta > 0$ by our assumptions. Moreover, for $j = 1, \ldots, p$, we set

$$X_{2j-1} = \frac{\phi_T(x_{j1}, x_{j2})}{\sigma} - \lambda(x_{j2})$$

$$X_{2j} = -\frac{\phi_T(x_{j1}, x_{j2})}{\sigma} - \lambda(x_{j2})$$

with $x_j = (x_{j1}, x_{j2})$. This notation allows us to write

$$\Phi_T = \max_{1 \leq j \leq 2p} X_j,$$

where $(X_1, \ldots, X_{2p})^\top$ is a Gaussian random vector with the following properties: (i) $\mu_j := \mathbb{E}[X_j] = -\lambda(x_{j2})$ and thus $\overline{\mu} = \max_{1 \leq j \leq 2p} |\mu_j| \leq C\sqrt{\log T}$, and (ii) $\sigma_j^2 := \mathrm{Var}(X_j) = 1$ for all j. Since $\sigma_j = 1$ for all j, it holds that $a_{2p} = b_{2p}$. Moreover, as the variables $(X_j - \mu_j)/\sigma_j$ are standard normal, we have that $a_{2p} = b_{2p} \leq \sqrt{2\log(2p)} \leq C\sqrt{\log T}$. With this notation at hand, we can apply Proposition A.3 to obtain that

$$\sup_{x \in \mathbb{R}} \mathbb{P}\Big(\big|\Phi_T - x\big| \leq \delta_T\Big) \leq C\delta_T\Big[\sqrt{\log T} + \sqrt{\log(1/\delta_T)}\Big] = o(1)$$

with $\delta_T = T^{1/q}/\sqrt{Th_{\min}} + \rho_T\sqrt{\log T}$, which is the statement of Proposition **??**. $\qquad\square$

Proof of Theorem 0.1

To prove Theorem 0.1, we make use of the two auxiliary results derived above. By Proposition A.1, there exist statistics $\widetilde{\Phi}_{n,T}$ for $T = 1, 2, \ldots$ which are distributed as $\widehat{\Phi}_T$ for any $T \geq 1$ and which have the property that

$$\big|\widetilde{\Phi}_{n,T} - \Phi_{n,T}\big| = o_p\Big(\frac{T^{1/q}}{\sqrt{Th_{\min}}} + \rho_T\sqrt{\log T} + \rho_T \max_{1 \leq i \leq n}\Big|\sum_{s=1}^{T}\widehat{\boldsymbol{\beta}}_i^\top(\mathbf{X}_{is} - \bar{\mathbf{X}}_i)\Big| +$$

$$+ \frac{T^{-1/2}}{\sqrt{Th_{min}}} \max_{1 \leq i \leq n} \max_{1 \leq t \leq T}\Big|\sum_{s=1}^{t}(\mathbf{X}_{is} - \bar{\mathbf{X}}_i)\Big|\Big), \qquad (1.9)$$

where $\Phi_{n,T}$ is a Gaussian statistic as defined in (**??**). The approximation result (1.9) allows us to replace the multiscale statistic $\widehat{\Phi}_{n,T}$ by an identically distributed version $\widetilde{\Phi}_{n,T}$ which is close to the Gaussian statistic $\Phi_{n,T}$. In the next step, we show that

$$\sup_{x \in \mathbb{R}}\big|\mathbb{P}(\widetilde{\Phi}_T \leq x) - \mathbb{P}(\Phi_T \leq x)\big| = o(1), \qquad (1.10)$$

which immediately implies the statement of Theorem 0.1. For the proof of (1.10), we use the following simple lemma:

Lemma A.4. *Let V_T and W_T be real-valued random variables for $T = 1, 2, \ldots$ such that $V_T - W_T = o_p(\delta_T)$ with some $\delta_T = o(1)$. If*

$$\sup_{x \in \mathbb{R}} \mathbb{P}(|V_T - x| \leq \delta_T) = o(1), \qquad (1.11)$$

then

$$\sup_{x \in \mathbb{R}}\big|\mathbb{P}(V_T \leq x) - \mathbb{P}(W_T \leq x)\big| = o(1). \qquad (1.12)$$

5

The statement of Lemma A.4 can be summarized as follows: If W_T can be approximated by V_T in the sense that $V_T - W_T = o_p(\delta_T)$ and if V_T does not concentrate too strongly in small regions of the form $[x - \delta_T, x + \delta_T]$ as assumed in (1.11), then the distribution of W_T can be approximated by that of V_T in the sense of (1.12). We now apply this lemma with $V_T = \Phi_T$, $W_T = \widetilde{\Phi}_T$ and $\delta_T = T^{1/q}/\sqrt{Th_{\min}} + \rho_T\sqrt{\log T}$: From (**??**), we already know that $\widetilde{\Phi}_T - \Phi_T = o_p(\delta_T)$. Moreover, by Proposition **??**, it holds that

$$\sup_{x \in \mathbb{R}} \mathbb{P}\Big(|\Phi_T - x| \leq \delta_T\Big) = o(1). \tag{1.13}$$

Hence, the conditions of Lemma A.4 are satisfied. Applying the lemma, we obtain (**??**), which completes the proof of Theorem 0.1.

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.

KHISMATULLINA, M. and VOGT, M. (2018). Multiscale inference and long-run variance estimation in nonparametric regression with time series errors. *Preprint*.