

# Conditional Extreme Value Estimation for Dependent Time Series

MARTIN BLADT<sup>1,a</sup>, LAURITS GLARGAARD<sup>1,b</sup> and THEODOR HENNINGSEN  
(corresponding author) <sup>1,c</sup>

<sup>1</sup>*Department of Mathematical Sciences, University of Copenhagen, Denmark,*  
<sup>a</sup>[martinbladt@math.ku.dk](mailto:martinbladt@math.ku.dk), <sup>b</sup>[lauritsglargaard@mail.dk](mailto:lauritsglargaard@mail.dk), <sup>c</sup>[th@math.ku.dk](mailto:th@math.ku.dk)

We study the consistency and weak convergence of the conditional tail function and conditional Hill estimators under broad dependence assumptions for a heavy-tailed response sequence and a covariate sequence. Consistency is established under  $\alpha$ -mixing, while asymptotic normality follows from  $\beta$ -mixing and second-order conditions. A key aspect of our approach is its versatile functional formulation in terms of the conditional tail process. Simulations demonstrate its performance across dependence scenarios. We apply our method to extreme event modelling in the oil industry, revealing distinct tail behaviours under varying conditioning values.

*MSC2020 subject classifications:* Primary 62G32; secondary 62E20, 62M10

*Keywords:* Extreme values; Time series; Conditional regular variation; Weak convergence

## Contents

1	Introduction . . . . .	2
2	The estimators . . . . .	3
2.1	Notation and setting . . . . .	3
2.2	Conditional tail function and conditional Hill estimators . . . . .	4
3	Main results . . . . .	6
3.1	Consistency . . . . .	6
3.2	Central limit theorems . . . . .	7
4	Second order behaviour and specification of sequences . . . . .	9
4.1	Consistency . . . . .	9
4.2	Asymptotic normality . . . . .	9
5	Finite-sample behaviour . . . . .	12
5.1	Bandwidth selection . . . . .	12
5.2	Simulation study . . . . .	13
6	Real data Analysis . . . . .	15
6.1	Description of dataset . . . . .	15
6.2	Analysis of crude oil spot price . . . . .	16
7	Conclusion and outlook . . . . .	18
	References . . . . .	21
A	Proofs of main results from Section 3 . . . . .	23
A.1	Proof of Theorem 1 (Tail consistency, $m$ -dependence) . . . . .	23
A.2	Proof of Theorem 2 (Tail consistency, $\alpha$ -mixing) . . . . .	25
A.3	Proof of Theorem 3 (Hill consistency) . . . . .	26
A.4	Proof of Theorem 4 (Functional CLT) . . . . .	27
A.5	Proof of Theorem 5 (Hill CLT) . . . . .	37
B	Proofs of results from Section 4 . . . . .	40
C	Anti-clustering conditions for simulations . . . . .	47
C.1	Conditional Fréchet and Pareto . . . . .	47
C.2	Conditional Max-Stable process . . . . .	48

D Kernel density estimation for mixing sequences . . . . .	49
E Background on fundamental limit theorems . . . . .	54

## 1. Introduction

The analysis of extremal events in the presence of covariates is a central problem in modern risk management. A persistent challenge in this setting is the temporal dependence inherent in both the covariate process and the process of interest, which complicates statistical inference. This paper focuses on the estimators for a heavy-tailed mixing time series given another mixing covariate time series.

Functional limit theorems for the tail empirical process under dependence conditions have been studied in the literature. Early results were established by [Rootzén \(1995\)](#) for  $\alpha$ -mixing sequences and later extended by [Rootzén \(2009\)](#) to more general dependence structures. In [Drees \(2000\)](#) the asymptotic normality of the POT estimator was shown under  $\beta$ -mixing conditions. Among the most widely used estimators for heavy-tailed data is the Hill estimator, originally introduced by [Hill \(1975\)](#). Its consistency in dependent settings was first investigated by [Resnick and Stărică \(1995\)](#), who analysed its behaviour under various dependence structures. Further work by [Resnick and Stărică \(1997\)](#) established the consistency of the tail empirical measure for autoregressive processes with heavy-tailed innovations. The asymptotic normality of the Hill estimator for i.i.d. data has been well documented, with significant contributions from [Hall \(1982\)](#), [Davis and Resnick \(1984\)](#), [Haeusler and Teugels \(1985\)](#), and [Csörgő and Mason \(1985\)](#). Extending these results to stationary sequences, [Hsing \(1991\)](#) proved a central limit theorem for the Hill estimator under  $\beta$ -mixing conditions, and a modern treatment can be found in [Kulik and Soulier \(2020\)](#).

In this paper, we investigate the consistency and weak convergence of the conditional kernel Hill estimator under general assumptions on both the response and covariate sequences. Specifically, we establish consistency under  $\alpha$ -mixing conditions and derive asymptotic normality under  $\beta$ -mixing conditions, with the latter results obtained under two general second-order conditions. A key feature of our approach is its functional formulation, which naturally leads to limit theorems for the conditional tail process. Through simulation studies, we investigate our method across different scenarios. Notably, a distinguishing advantage of our estimator is its simple (and easy to plug-in) variance structure, in contrast to the often intricate variance expressions encountered in the unconditional case. We apply our method to extreme event modelling in the oil industry where, interestingly, distinct tail behaviours arise under different conditioning values.

Other work is related to ours. In the presence of covariates, the asymptotic properties of extreme value estimators have been investigated using regression-based approaches. In [Daouia et al. \(2010\)](#) tail index estimation is studied under  $\alpha$ -mixing conditions using a smoothed Pickands estimator, a technique later generalized by [Daouia, Gardes and Girard \(2013\)](#), [Daouia, Stupfler and Usseglio-Carlevé \(2023\)](#) to regression models within the general max-domain of attraction. Furthermore, tail index regression in a random covariate setting has been analysed using a maximum likelihood framework, as explored by [Wang and Tsai \(2009\)](#). Finally, asymptotic normality of the conditional kernel Hill estimator has been explored through different techniques and for a special second-order regularly varying class in the preprint [Goegebeur and Guillou \(2024\)](#).

The remainder of the paper is structured as follows. Section 2 introduces the mathematical framework for heavy-tailed analysis in dependent settings, defining key concepts and estimators. Section 3 establishes sufficient conditions for the consistency of the estimators under  $\alpha$ -mixing dependence, as well as conditions for their asymptotic normality under  $\beta$ -mixing conditions. Section 4 provides specific second-order conditions and explicit sequences which imply our results. Section 5 evaluates the finite-sample performance and robustness of the estimator through simulation studies, while Section 6

applies our methodology to extreme event modelling in the oil industry. Finally, Section 7 concludes. Proofs of all results are provided in a series of appendices, as well as a additional results regarding anti-clustering conditions for certain models, kernel density estimation for dependent data, and mixing and functional limit theorems.

## 2. The estimators

In this section we introduce estimators for the conditional tail function and for the conditional Hill estimator, the latter having the same structure as in independent settings. It is only when proving its asymptotics that the temporal dependence of the data comes into play.

### 2.1. Notation and setting

Let  $\{(X_j, Y_j), j \in \mathbb{Z}\}$  be a stationary time series defined on the probability space  $(\Omega, \Pr, \mathcal{F})$ . Here,  $\{Y_j, j \in \mathbb{Z}\}$  is a positive process of interest and  $\{X_j, j \in \mathbb{Z}\}$  is a process of covariates, where  $X_j, Y_j \in \mathbb{R}$ , for each  $j \in \mathbb{Z}$ . Let the distribution function of  $Y_0$  be denoted by  $F$ , the density of  $X_0$  be denoted by  $g$ , which is assumed twice continuously differentiable, the joint density  $g_{X_i, X_j}$  of  $(X_i, X_j)$  exists, and is bounded for all  $i, j \in \mathbb{Z}$ , and let the distribution function of  $Y_0$  given  $X_0 = x$  be denoted by  $F^x$ .

Let  $x \in \mathbb{R}$  with  $g(x) > 0$  be given. Our key assumption is that the conditional distribution of  $Y_0$ , given  $X_0 = x$ , is regularly varying with tail index  $1/\gamma(x) > 0$ , and that this holds in a neighbourhood around  $x$ , that is

$$\bar{F}^u(y) = \Pr(Y_0 > y | X_0 = u) = L^u(y) y^{-\frac{1}{\gamma(u)}}, \quad y > 0, \quad (1)$$

for all  $u$  in an open set  $U \subseteq \mathbb{R}$  containing  $x$ , and where  $L^u$  is slowly varying at infinity, that is  $L^u(tv)/L^u(v) \rightarrow 1$  as  $v \rightarrow \infty$  for any  $t > 0$ .

Let  $K$  be a bounded density, denoted a kernel, supported on  $[-1, 1]$  and denote by  $\{h_n, n \in \mathbb{N}\}$  the corresponding bandwidth (scale) parameter. Let  $\{u_n^x, n \in \mathbb{N}\}$  be a scaling sequence diverging to infinity. The conditional survival function  $\bar{F}^x$  is estimated with the Nadaraya–Watson pathwise estimator as

$$\bar{F}_n^x(y) = \sum_{j=1}^n \frac{K\left(\frac{x-X_j}{h_n}\right)}{\sum_{j=1}^n K\left(\frac{x-X_j}{h_n}\right)} 1_{\{Y_j > y\}}, \quad y > 0.$$

Let the corresponding tail empirical distribution function, for deterministic threshold, be given by

$$\tilde{T}_n^x(s) = \frac{\bar{F}_n^x(su_n^x)}{\bar{F}_n^x(u_n^x)} = \frac{\frac{1}{nh_n \bar{F}^x(u_n^x)} \sum_{j=1}^n K\left(\frac{x-X_j}{h_n}\right) 1_{\{Y_j > su_n^x\}}}{\frac{1}{nh_n} \sum_{j=1}^n K\left(\frac{x-X_j}{h_n}\right)} =: \frac{\tilde{V}_n^x(s)}{g_n(x)},$$

for  $s > 0$ . The target of  $\tilde{T}_n^x$  is the conditional tail function  $s \mapsto T^x(s) = s^{-1/\gamma(x)}$ . Indeed, this convergence, both in probability and weakly, will be shown under suitable conditions in this paper. However, the deterministic thresholds  $u_n^x$  are not observable. Below we provide asymptotic results also for when these thresholds are estimated from data.

## 2.2. Conditional tail function and conditional Hill estimators

To transfer the tail function behaviour to the conditional tail index, the quantity of interest, the following identity is key:

$$\gamma(x) = \int_1^\infty s^{-1/\gamma(x)-1} ds = \int_1^\infty \frac{T^x(s)}{s} ds.$$

In this paper, we assume that an *intermediate sequence* is given, that is a sequence  $k_n \leq n$  such that

$$k_n \rightarrow \infty, \quad k_n/n \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

Let  $u_n^x = F^{\leftarrow, x}(1 - k_n/n)$  be an unobserved sequence of deterministic thresholds, diverging to infinity, where  $F^{\leftarrow, x}$  denotes the generalised inverse of  $F^x$ . We may clearly estimate this sequence by its empirical counterpart  $\{q_{n, k_n}^x, n \in \mathbb{N}\}$  given by

$$q_{n, k_n}^x = F_n^{\leftarrow, x}(1 - k_n/n), \quad n \in \mathbb{N}.$$

The conditional tail function is thus estimated by

$$\widehat{T}_n^x(s) = \frac{\overline{F}_n^x(sq_{n, k_n}^x)}{\overline{F}_n^x(u_n^x)} = \widetilde{T}_n^x\left(s \frac{q_{n, k_n}^x}{u_n^x}\right),$$

which no longer depends on the sequence  $u_n^x$ , which can be seen from the definition of  $\widetilde{T}_n^x$  and from the fact that  $\overline{F}_n^x(u_n^x) = k_n/n$ .

Furthermore, we construct a conditional version of the Hill Estimator, given by

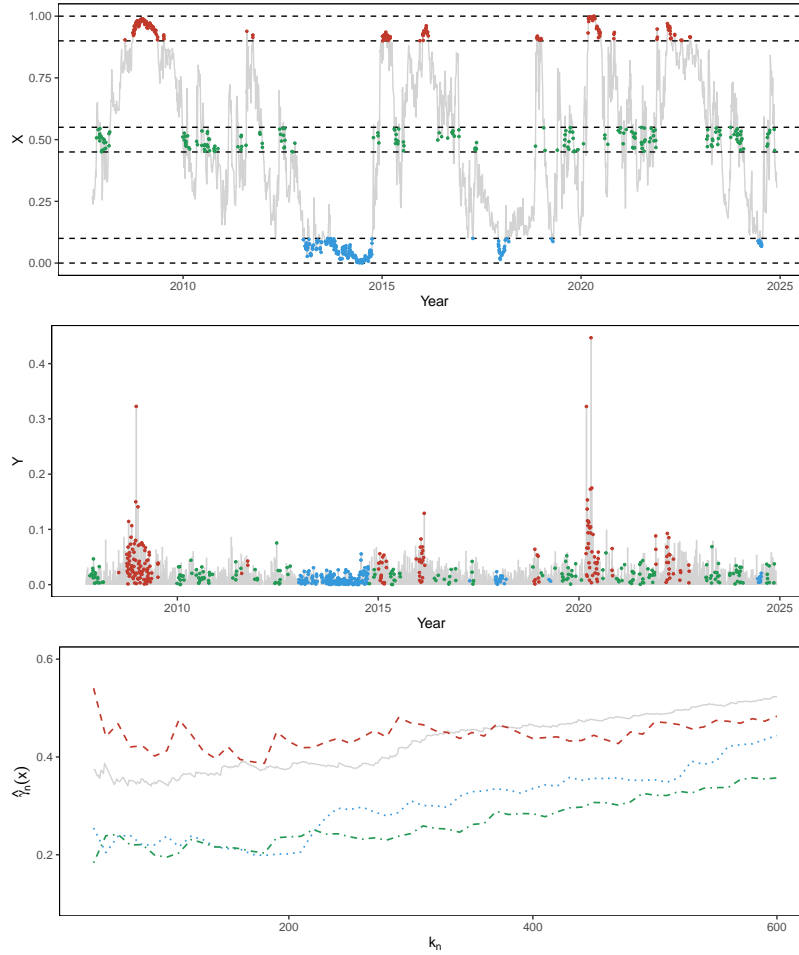
$$\widehat{\gamma}_n(x) = \int_1^\infty \frac{\widehat{T}_n^x(s)}{s} ds = - \int_1^\infty \log(s) \widehat{T}_n^x(ds) = \frac{n}{k_n} \frac{\sum_{j=1}^n K\left(\frac{x-X_j}{h_n}\right) \log_+\left(\frac{Y_j}{q_{n, k_n}^x}\right)}{\sum_{j=1}^n K\left(\frac{x-X_j}{h_n}\right)}.$$

It is worth noticing that the effective sample size used in this estimator is proportional to  $h_n k_n$  instead of  $k_n$ , which is also the rate for weak convergence. Consequently, the parameter  $k_n$  has a different interpretation than in the unconditional case. In particular,  $\widehat{\gamma}_n(x)$  can be plotted for the entire range  $k_n \in \{2, 3, \dots, n\}$ .

**Example.** In case of  $K$  being the uniform density on  $[-1, 1]$ , then  $\widehat{\gamma}_n(x)$  reduces to the Hill estimator applied to the subsample of the data having  $X_i \in [x - h_n, x + h_n]$ . Notice that the  $k_n$  sequence no longer corresponds to the number of top order statistics in the subsample. Instead, let that number be defined as  $\tilde{k}_n$ . They are linked through the following relation:

$$k_n \sim \frac{n\tilde{k}_n}{\#\{X_i \in [x - h_n, x + h_n]\}}$$

Before delving into the mathematical treatment of our estimators, we provide an example of a conditional Hill plot. In Figure 1 we have taken  $\{Y_j\}$  to be the (absolute value of) the negative log-returns of the West Texas Intermediate (WTI) grade crude oil spot prices from 2007-2024. As a covariate  $\{X_j\}$ , we take the Crude Oil Volatility Index (COVE OVX), transformed to have uniform marginals. For



**Figure 1.** Top panel: covariate process  $\{X_j\}$  given by the Crude Oil Volatility Index (CBOE OVX). The effective sample is calculated using a uniform kernel centred at  $x = 0.05, 0.50, 0.95$  (highlighted in red, green and blue, respectively) and with  $h_n = 0.05$ . Middle panel: target process  $\{Y_j\}$  given by the absolute value of the negative log-returns of the West Texas Intermediate (WTI) grade crude oil spot prices with highlighted subsamples. Bottom panel: conditional Hill estimates  $\hat{\gamma}_n(x)$  in dotted blue, dashdotted green and dashed red, respectively; in solid grey we have the unconditional Hill estimator of the entire sample.

large  $x$ , we clearly observe a different, heavier tail regime. In contrast, for medium and small  $x$ , the tail regime is similar and lighter. Below, it's shown that the choice of kernel enters the asymptotic variance of the conditional Hill estimator through a certain integral. Transforming the covariate process to have uniform marginals is not necessary. It's nonetheless done here, in the simulations and real data analysis below to avoid sparse regions of data to ensure good practical performance.

### 3. Main results

In this section we provide consistency and weak convergence results for the estimators of the conditional tail function and conditional tail index under suitable conditions. Most conditions are versions of their most general unconditional counterparts. Notably, only  $\alpha$ -mixing is required for consistency, while  $\beta$ -mixing is required for normality. Furthermore, second order conditions are formulated in a very general and abstract form. It is later seen that broad classes of regularly-varying tails satisfy them. Unless otherwise specified, all statements below hold for every  $x \in U$ .

#### 3.1. Consistency

A fundamental underlying result is the uniform consistency of the conditional tail empirical function, for deterministic thresholds. For  $m$ -dependent sequences, this can be established under rather lax conditions, while for  $\alpha$ -mixing sequences we require additional dependence and speed conditions.

**Condition 1.** *The conditions for consistency are as follows.*

- 1.1 *The map  $z \mapsto 1/\gamma(z)$  is  $c_\gamma$ -Lipschitz continuous,*
- 1.2  $\lim_{n \rightarrow \infty} h_n \log(u_n^x) = 0$ ,
- 1.3 *There exists  $y_0 \in \mathbb{R}$  and  $c_L > 0$  such that*

$$\sup_{y \geq y_0} \left| \frac{\log(L^x(y)) - \log(L^u(y))}{\log(y)} \right| \leq c_L |x - u|,$$

*uniformly in  $x, u \in U$ ,*

- 1.4  $\lim_{n \rightarrow \infty} nh_n \bar{F}^x(u_n^x) = \infty$ ,
- 1.5 *It holds  $g_n(x) \xrightarrow{\text{Pr}} g(x)$ , see Lemma 12, 13 or 14,*
- 1.6 *For any  $s, t > 0$  there exists a finite constant  $M^x(s, t)$  not depending on  $j$  such that*

$$\frac{\mathbb{E} \left[ 1_{\{Y_0 > su_n^x\}} 1_{\{Y_j > tu_n^x\}} \mid X_0 = x + h_n w_1, X_j = x + h_n w_2 \right]}{\bar{F}^x(u_n^x)} \rightarrow M^x(s, t),$$

*uniformly in  $w_1, w_2 \in [-1, 1]$ .*

**Theorem 1.** *Assume (1) and that conditions 1.1-1.5 hold and that the time series  $\{(X_j, Y_j), j \in \mathbb{Z}\}$  is stationary and  $m$ -dependent. Then*

$$\tilde{T}_n^x(s) \xrightarrow{\text{Pr}} T^x(s) = s^{-1/\gamma(x)}, \quad s > 0, \quad (2)$$

*the convergence being uniform on compacts bounded away from zero.*

**Theorem 2.** *Assume (1), condition 1 and that the time series  $\{(X_j, Y_j), j \in \mathbb{Z}\}$  is stationary and  $\alpha$ -mixing with  $\alpha(j) = \mathcal{O}(j^{-\eta})$ , for  $\eta > 2$ .*

*If there exists  $c_x \in (0, \infty)$  such that*

$$\frac{\bar{F}^x(u_n^x) + h_n^{c_x \eta - 2}}{nh_n^{c_x} (\bar{F}^x(u_n^x))^2} \rightarrow 0, \quad (3)$$

*then (2) holds uniformly on compacts bounded away from zero.*

As a first application we get that consistency of the conditional tail empirical function implies that the intermediate empirical quantile  $q_{n,k_n}^x$  is equivalent to the deterministic  $u_n^x$ . This is of practical importance, since the former is estimable from data. Consistency of the conditional Hill estimator then directly follows.

**Proposition 1.** *If (2) holds, then  $q_{n,k_n}^x/u_n^x \xrightarrow{\text{Pr}} 1$ .*

**Theorem 3.** *Assume conditions 1.1-1.3, 1.5 and (2). Then  $\widehat{\gamma}_n(x) \xrightarrow{\text{Pr}} \gamma(x)$ .*

### 3.2. Central limit theorems

We now consider the univariate conditional tail empirical process for deterministic thresholds

$$\widetilde{\mathbb{T}}_n^x(s) = \sqrt{k_n h_n} \{ \widetilde{T}_n^x(s) - T^x(s) \}, \quad s > 0,$$

for which a functional central limit is established in the case of  $\beta$ -mixing. The fundamental tool to achieve this goal is the blocking method. Let  $\{r_n, n \in \mathbb{N}\}$  be an intermediate sequence of block sizes and set  $m_n = \lfloor n/r_n \rfloor$  which is the number of complete blocks in the sample.

Notice that the speed of convergence is determined by the effective sample size  $nh_n \overline{F}^x(u_n^x) = h_n k_n$ , which is slower than in the unconditional case, which is simply  $k_n$ . This price to pay, however, results in a variance which agrees with the i.i.d. conditional case, in contrast to the more involved expression for unconditional time series provided as an infinite sum, e.g. in [Kulik and Soulier \(2020, Eq. \(9.5.4\)\)](#).

For  $s_0 \in (0, 1)$ , define

$$B_n^x(s_0) = \sup_{t \geq s_0} \left| \mathbb{E} \left[ \widetilde{V}_n^x(t) \right] - t^{-1/\gamma(x)} g(x) \right|.$$

**Condition 2.** *The conditions for normality are as follows.*

2.1  $\lim_{n \rightarrow \infty} r_n h_n \overline{F}^x(u_n^x) = 0,$

2.2 *For all  $s, t > 0$ ,*

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \sum_{j=m}^{r_n} \frac{\mathbb{E} \left[ K \left( \frac{x-X_0}{h_n} \right) 1_{\{Y_0 > s u_n^x\}} K \left( \frac{x-X_j}{h_n} \right) 1_{\{Y_j > t u_n^x\}} \right]}{h_n \overline{F}^x(u_n^x)} = 0,$$

2.3 *For the stationary time series  $\{(X_j, Y_j), j \in \mathbb{Z}\}$  with  $\beta$ -mixing coefficients  $\{\beta_n\}$ , there exists an intermediate sequence  $\{\ell_n\}$  such that*

$$\lim_{n \rightarrow \infty} \frac{1}{\ell_n} = \lim_{n \rightarrow \infty} \frac{\ell_n}{r_n} = \lim_{n \rightarrow \infty} \frac{r_n}{n} = \lim_{n \rightarrow \infty} \frac{n}{r_n} \beta_{\ell_n} = 0,$$

2.4 *It holds that  $\sqrt{k_n h_n} \{g(x) - g_n(x)\} \xrightarrow{\text{Pr}} 0$ , see [Lemma 15](#) or [16](#),*

2.5 *There exists  $s_0 \in (0, 1)$ , such that  $\sqrt{k_n h_n} B_n^x(s_0) \rightarrow 0$ ,*

2.6 *For all  $A > 0$ ,*

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \sum_{j=m}^{r_n} \frac{\mathbb{E} \left[ K \left( \frac{x-X_0}{h_n} \right) K \left( \frac{x-X_j}{h_n} \right) \log_+ \left( \frac{Y_0}{u_n^x A} \right) \log_+ \left( \frac{Y_j}{u_n^x A} \right) \right]}{h_n \overline{F}^x(u_n^x)} = 0,$$

2.7 It holds that

$$\lim_{n \rightarrow \infty} \sqrt{k_n h_n} \int_1^\infty \frac{\mathbb{E} [\tilde{V}_n^x(t)] - t^{-1/\gamma(x)} g(x)}{t} dt = 0.$$

**Proposition 2.** Assume (1), conditions 1.1-1.5 and 2.1-2.5. Then it holds that  $\tilde{\mathbb{T}}_n^x$  converges weakly to a centred Gaussian process  $\mathbb{T}^x$  in  $\mathbb{D}([s_0, \infty))$  endowed with the  $J_1$ -topology, with covariance function

$$\Gamma(s, t) = \frac{(s \vee t)^{-1/\gamma(x)}}{g(x)} \int K^2(u) du,$$

and the process  $\mathbb{T}^x$  has an almost surely continuous version.

**Proposition 3.** Assume (1), conditions 1.1-1.5 and 2.1-2.5. Then

$$\sqrt{h_n k_n} \left\{ \frac{q_{n, k_n}^x}{u_n^x} - 1 \right\} \xrightarrow{d} \gamma(x) \mathbb{T}^x(1),$$

where

$$\gamma(x) \mathbb{T}^x(1) \stackrel{d}{=} N\left(0, \frac{\gamma(x)^2}{g(x)} \int K^2(u) du\right).$$

The sequence  $\{u_n^x\}$  is unobserved and we may estimate it with  $\{q_{n, k_n}^x\}$ . Let

$$\widehat{\mathbb{T}}_n^x(s) = \sqrt{k_n h_n} \left\{ \widehat{T}_n^x(s) - T^x(s) \right\},$$

which we refer to as the tail empirical process with random levels.

**Theorem 4.** Assume (1), conditions 1.1-1.5 and 2.1-2.5. Then

$$\widehat{\mathbb{T}}_n^x \xrightarrow{w} \mathbb{T}^x - \mathbb{T}^x(1) \times T^x,$$

in  $\mathbb{D}([s_0, \infty))$  endowed with the  $J_1$  topology.

It turns out that Theorem 4 is not required to establish convergence of the conditional Hill estimator. However, it is an interesting result in its own right, since the empirical tail process can be used to measure how closely the top order statistics behave like a strict Pareto tail. Generalisations to multiple dimensions are also possible through the same proof techniques.

**Theorem 5.** Assume (1), that conditions 1.1-1.5 and 2 hold. Then it holds that

$$\sqrt{k_n h_n} (\widehat{\gamma}_n(x) - \gamma(x)) \xrightarrow{d} N\left(0, \frac{\gamma(x)^2}{g(x)} \int K^2(u) du\right).$$

**Remark.** Examining closely the proof techniques of Theorems 3 and 5, we see that our results can be extended to functionals of the tail empirical process of the form

$$\widehat{\Phi}_n(x) = \int \phi(s) \widehat{T}_n^x(ds),$$

provided conditions involving the logarithmic function are replaced with suitably modified versions involving  $\phi$ . For instance, the choice  $\phi(s) = \log(s)^2$ , yields an estimator of  $2\gamma(x)^2$ . In case the functional  $z \mapsto \int \phi dy$ , or any other functional is defined on  $\mathbb{D}([A, B])$  and is sup-norm continuous or Hadamard differentiable, consistency and a central limit theorem for the associated estimator follow directly from Theorem 4, together with the continuous mapping theorem or functional delta-method, respectively.

**Remark.** The bias of the conditional Hill estimator vanishes due to condition 2.7. If this condition is relaxed to

$$\lim_{n \rightarrow \infty} \sqrt{k_n h_n} \int_1^\infty \frac{\mathbb{E} [\widetilde{V}_n^x(t)] - t^{-1/\gamma(x)} g(x)}{t} dt = B \in \mathbb{R},$$

the asymptotic variance in Theorem 5 remains similar, while the bias becomes  $B/g(x)$ .

## 4. Second order behaviour and specification of sequences

In this section we investigate the fulfilment of the abstract second order conditions by considering certain classes of regularly-varying tails. Moreover, there are several deterministic sequences involved in the above results, which have to satisfy speed assumptions. We also verify that it is always possible to find such sequences for the proposed classes.

### 4.1. Consistency

In subsection 3.1 we presented two theorems that yield consistency of  $\widetilde{T}_n^x$ , the first one under  $m$ -dependence, and the second one assuming  $\alpha$ -mixing and additionally (3). In the latter case there are three limit assumptions related to the speed of deterministic sequences needed for consistency, which are given by

1.  $h_n \log(u_n^x) = h_n \log(F^{c \cdot x}(1 - k_n/n)) \rightarrow 0$ , (1.2)
2.  $nh_n \overline{F}^x(u_n^x) = k_n h_n \rightarrow \infty$ , (1.4)
3.  $(\overline{F}^x(u_n^x) + h_n^{c_x \eta - 2}) / (nh_n^{c_x} (\overline{F}^x(u_n^x))^2) = (k_n + nh_n^{c_x \eta - 2}) / (h_n^{c_x} k_n^2) \rightarrow 0$ , (3).

Notice that, in the case of  $m$ -dependence, only the first two conditions are required. In any case, the conditions are directly on the conditional distribution, so the results below do not make use of the time-series structure of the data, but only on the (first-order) regular variation property of the conditional tail.

In what follows, we use the notation  $f_n \asymp g_n$ , when  $f_n \sim c g_n$ , for some constant  $c > 0$ .

**Theorem 6.** Assume (1). If  $k_n \asymp n^\delta$  for  $\delta \in ((\eta + 1)^{-1}, 1)$ ,  $h_n \asymp \overline{F}^x(u_n^x)^e$  for  $e \in (0, \delta/(1 - \delta) \wedge (\delta(\eta + 1) - 1)/(2(1 - \delta)))$ , there exists a  $c_x \in (0, \infty)$  such that the above limit conditions are satisfied.

**Theorem 7.** Assume (1). If  $h_n \asymp \log(n)^{-e}$  for  $e \in (4/3, 2)$  and  $k_n \asymp \log(n)^2$ , then the first two above limit conditions are satisfied.

### 4.2. Asymptotic normality

Asymptotic normality of the tail empirical process  $\widetilde{\mathbb{T}}_n^x$  and of the conditional Hill estimator  $\widehat{\gamma}_n(x)$  rely on two crucial, yet abstract conditions 2.5 and 2.7 respectively, related to the regular variation bias.

In this subsection we propose quite general second-order-type conditions, under which we can provide examples of sequences that satisfy the speed restrictions required for asymptotic normality.

**Lemma 1.** *Assume conditions 1.1-1.4, that  $F^x$  is a continuous distribution and that there exist  $c_x, \gamma(x) > 0$ , and  $\zeta_x \gamma(x) > 1$  such that*

$$\sup_{t \geq 1} t^{\zeta_x} \left| \overline{F}^x(t) - c_x t^{-1/\gamma(x)} \right| < \infty. \quad (4)$$

*If there exist an intermediate sequence  $\{k_n\}$  and a bandwidth sequence  $\{h_n\}$  such that*

$$\lim_{n \rightarrow \infty} \sqrt{k_n h_n^3} \log(n/k_n) = \lim_{n \rightarrow \infty} \sqrt{h_n} k_n^{\zeta_x \gamma(x) - \frac{1}{2}} n^{1 - \zeta_x \gamma(x)} = 0, \quad (5)$$

*then conditions 2.5 and 2.7 hold.*

**Lemma 2.** *Assume conditions 1.1-1.4, that  $F^x$  is a continuous distribution and that there exist  $c_x, \gamma(x) > 0$ , and  $\zeta_x \gamma(x) > 1$  such that*

$$\sup_{t \geq 1} t^{\zeta_x} \left| \overline{F}^x(t) - c_x \log(t) t^{-\gamma(x)} \right| < \infty. \quad (6)$$

*If there exist an intermediate sequence  $\{k_n\}$  and a bandwidth sequence  $\{h_n\}$  such that*

$$\lim_{n \rightarrow \infty} \sqrt{k_n h_n^3} \log(n/k_n) = \lim_{n \rightarrow \infty} \frac{\sqrt{k_n h_n}}{\log(n/k_n)} = 0, \quad (7)$$

*then conditions 2.5 and 2.7 hold.*

**Lemma 3.** *Assume conditions 1.1-1.4, 1.6, 2.1, the time series  $\{(X_j, Y_j), j \in \mathbb{Z}\}$  is stationary and  $\alpha$ -mixing with  $\alpha(j) = O(j^{-\eta})$ , for  $\eta > 2$ , and that for  $A > 0$  there exists a finite constant  $\widetilde{M}^x$  not depending on  $j$  such that*

$$\frac{\mathbb{E} \left[ \log_+ \left( \frac{Y_0}{u_n^x A} \right) \log_+ \left( \frac{Y_j}{u_n^x A} \right) \mid X_0 = x + h_n w_1, X_j = x + h_n w_2 \right]}{\overline{F}^x(u_n^x)} \rightarrow \widetilde{M}^x, \quad (8)$$

*uniformly in  $w_1, w_2 \in [-1, 1]$ . If there exist a bandwidth sequence  $\{h_n\}$  and constants  $c_x, c'_x > 0$  and  $p > 1$ , such that*

$$\lim_{n \rightarrow \infty} \left( h_n^{1-c_x} + h_n^{c_x(\eta-1)-1} \overline{F}^x(u_n^x)^{-1} \right) = \lim_{n \rightarrow \infty} \left( h_n^{1-c'_x} + h_n^{-1/p+c'_x(\eta/p-1)} \overline{F}^x(u_n^x)^{-1/p} \right) = 0, \quad (9)$$

*then the anti-clustering conditions 2.2 and 2.6 hold.*

In essence, Lemmas 1 and 2 transfer the asymptotic bias conditions into additional speed conditions on the  $h_n$  and  $k_n$  sequences. Standard second-order conditions in extremes are often formulated in terms of limits which imply (4); in particular the broad Hall class (Hall and Welsh, 1985) falls into this framework. Second-order conditions akin to (2) are not often encountered in the literature, and our result is inspired by (Kulik and Soulier, 2020, Eq. (9.6.3)). The result of Lemma 3 provides an alternative way of verifying the anti-clustering conditions. For convenience, we now restate the main speed conditions for asymptotic normality of the conditional Hill estimator to hold.

In addition to the assumptions needed for consistency, we require sequences  $\{(k_n, h_n, r_n, \ell_n), n \in \mathbb{N}\}$  such that the following conditions are met:

1.  $r_n h_n \overline{F}^x(u_n^x) = r_n h_n k_n / n \rightarrow 0$ , (2.1)
2.  $\lim_{n \rightarrow \infty} 1/\ell_n = \lim_{n \rightarrow \infty} \ell_n / r_n = \lim_{n \rightarrow \infty} r_n / n = \lim_{n \rightarrow \infty} (n/r_n) \beta_{\ell_n} = 0$ , (2.3)
3.  $\sqrt{k_n h_n} \{g(x) - g_n(x)\} \xrightarrow{\text{Pr}} 0$ , (2.4)
4.  $\sqrt{k_n h_n} B_n^x(s_0) \rightarrow 0$ , (2.5)
5.  $\sqrt{k_n h_n} \int_1^\infty (\mathbb{E}[\tilde{V}_n^x(t)] - t^{-1/\gamma(x)} g(x)) / t \, dt \rightarrow 0$ , (2.7).

**Theorem 8.** Assume (1) and that  $\{(X_j, Y_j), j \in \mathbb{Z}\}$  is a stationary,  $\beta$ -mixing time series, with  $\alpha$ -mixing coefficients  $\alpha(j) = O(j^{-\eta})$ , for  $\eta > 2$ ,  $\beta$ -mixing coefficients  $\beta(j) = O(j^{-\nu})$  for  $\nu > 0$ , and that  $F^x$  is continuous. Assume conditions 1.1, 1.3, 1.6, (8), and the second-order condition (4). Then there exist sequences  $\{(k_n, h_n, r_n, \ell_n), n \in \mathbb{N}\}$  such that the first two conditions for consistency, all the above conditions for normality and anti-clustering conditions 2.2 and 2.6 are satisfied.

**Theorem 9.** Assume (1) and that  $\{(X_j, Y_j), j \in \mathbb{Z}\}$  is a stationary,  $\beta$ -mixing time series, with  $\alpha$ -mixing coefficients  $\alpha(j) = O(j^{-\eta})$ , for  $\eta > 2$ ,  $\beta$ -mixing coefficients  $\beta(j) = O(j^{-\nu})$  for  $\nu > 0$ , and that  $F^x$  is continuous. Assume conditions 1.1 and 1.3 and the second-order condition (6). Then there exist sequences  $\{(k_n, h_n, r_n, \ell_n), n \in \mathbb{N}\}$  such that the first two conditions for consistency and all the above conditions for normality are satisfied.

Note that in both of the above theorems, only the first two conditions for consistency are required, by  $\beta$ -mixing. In the simulation study below, we present three non-trivial examples of processes where the above speed conditions and anti-clustering conditions are satisfied.

**Remark.** In the proof of Theorem 8, we let  $k_n \asymp n^\delta$ , where

$$\delta \in \left(0, \frac{3}{2} \frac{\nu}{1+\nu} \wedge \frac{2(\zeta_x - 1/\gamma(x))}{2\zeta_x - 1/\gamma(x)}\right).$$

The restriction  $\delta < 3\nu/(2(1+\nu))$  is a consequence of allowing the  $\beta$ -mixing coefficient to decay polynomially of any order. The more common and stronger assumption of exponential decay i.e.  $\beta(j) = O(\rho^j)$  for  $\rho \in (0, 1)$  removes this restriction. On the other hand, the restriction  $\delta < 2(\zeta_x - 1/\gamma(x))/(2\zeta_x - 1/\gamma(x))$  is a direct consequence of the required rates in Lemma 1 that arise from the second order assumption (4).

Also notice that the rate of convergence in Theorem 5 is  $k_n h_n$ . In the setting of Theorem 8, this rate reduces to

$$k_n h_n \asymp n^{\delta - (1-\delta)e},$$

where  $\delta$  is in the above interval and  $e$  is bounded from below by  $\delta/(3(1-\delta))$ . To optimize this rate,  $\delta$  should be picked as large as possible and  $e$  should be as small as possible. For  $e$  close to its lower bound,  $k_n h_n$  behaves like  $n^{2\delta/3}$ . If we consider only the case of exponential decay of the  $\beta$ -mixing coefficients, the optimal rate is then

$$k_n h_n = o\left(n^{\frac{2}{3} \frac{2(\zeta_x - 1/\gamma(x))}{2\zeta_x - 1/\gamma(x)}}\right).$$

It resembles the rate achieved in the unconditional case under a corresponding second order assumption,  $k_n = o(n^\rho)$  where  $\rho = 2(\zeta_x - 1/\gamma(x))/(2\zeta_x - 1/\gamma(x))$  (see [Kulik and Soulier \(2020\)](#), problem 9.6), however the fraction  $2/3$  in the exponent is caused by the bandwidth sequence slowing the speed of convergence.

The preprint [Goegebeur and Guillou \(2024\)](#) obtains similar asymptotic rates for the conditional Hill estimator. We now compare their main assumptions. Their kernel assumptions ( $\mathcal{K}$ ) are similar to ours; they also cover the multidimensional covariate case. They throughout assume a second order condition ( $\mathcal{D}$ ), which is seen to be a special case of equation (4); they do not cover the second order condition of equation (6). Their assumption ( $\mathcal{J}$ ) implies our anti-clustering condition 2.2. Their strong condition ( $\mathcal{X}$ ) on the joint behaviour of  $(X_0, X_i)$  is here replaced by simply  $g_{X_0, X_i}$  being bounded for every  $i$ . Their condition (A) is similar to our beta-mixing condition 2.3 but it depends on an involved probability sequence  $\{\nu_n\}$ . Their condition (B) implies our anti-clustering condition 2.6. In their Theorem 3.2 and 3.3, the first bandwidth condition corresponds to condition 1.2, the second to equation (16), the third and fourth to the first and second part of equation (5).

## 5. Finite-sample behaviour

In this section we investigate the finite-sample behaviour of the conditional Hill estimator. We begin by discussing the selection of the bandwidth parameter for a fixed sample size. Then we perform some simulation studies where we examine how different intertemporal dependence structures affect the performance of the estimator.

### 5.1. Bandwidth selection

In this section we present methods for bandwidth selection, which is one of the main parameters above subject to tuning. The choice of bandwidth should, according to the theorems above, obey certain speed constraints, and the remark in the former section yields that it should depend on the threshold sequence  $\{k_n\}$ . Below we present four selection approaches, two of which are  $k_n$ -dependent and two of which are not. The proposed techniques are assessed in the subsequent subsection.

1. Sheather-Jones bandwidth fitted only on the covariate process (see [Sheather and Jones \(1991\)](#)). It's a plug-in method that provides a global bandwidth, independent of  $k_n$ .
2. Sheather-Jones bandwidth fitted only on the top  $k_n$  values of  $\{X_j\}$ , sorted according to the order statistics of  $\{Y_j\}$ .
3. Various approaches concerning non-parametric regression (see [Daouia et al. \(2010\)](#), [Gannoun et al. \(2002\)](#)) have proposed a purely data-driven, cross-validation bandwidth selection method

$$\hat{h} = \arg \min_{h>0} \sum_{i=1}^n \sum_{j=1}^n \left( 1_{\{Y_i > Y_j\}} - \bar{F}_{n,-i}^{X_i}(Y_j) \right)^2, \quad (10)$$

where  $\bar{F}_{n,-i}^x$  is the leave-one-out Nadaraya–Watson estimator of the conditional survival function given by

$$\bar{F}_{n,-i}^x(y) = \frac{\sum_{j=1, j \neq i}^n K\left(\frac{x-X_j}{h_n}\right) 1_{\{Y_j > y\}}}{\sum_{j=1, j \neq i}^n K\left(\frac{x-X_j}{h_n}\right)}.$$

Alternatively, as proposed by Li and Racine (2008), a leave-one-out doubly-smoothed Nadaraya–Watson estimator of the conditional survival function, where also  $\{Y_j\}$  is smoothed with distribution function  $G$ , using bandwidth  $b_n$  is given by

$$\bar{F}_{n,-i}^x(y) = \frac{\sum_{j=1, j \neq i}^n K\left(\frac{x-X_j}{h_n}\right) G\left(\frac{y-Y_j}{b_n}\right)}{\sum_{j=1, j \neq i}^n K\left(\frac{x-X_j}{h_n}\right)}, \quad (11)$$

where  $G(v) = \int_{(-\infty, v]} w(u) du$ , for symmetric, bounded and compactly supported density  $w$ . In the numeric studies below the kernels are taken to be Gaussian, which deviates from the compact support assumption; this deviation does not pose any practical issues. The asymptotic properties of the bandwidth selection through such leave-one-out procedure for the empirical conditional survival function have, to our knowledge, not been examined in the case of mixing data and are thus material for future work. Presently, we simply mention that this approach provides a global bandwidth independent of both  $k_n$  and focuses not only on  $x$ , but on all covariate values at the same time.

4. Finally consider the data-independent case of picking  $h_n = \sqrt{\log(k_n)/n}$ , which falls into the framework of conditions 1.2, 1.4, 2.1, and equations (5), (16) for certain  $\{k_n\}$  and appropriately chosen  $\{r_n\}$ .

## 5.2. Simulation study

In this section we examine the finite-sample behaviour of the conditional Hill Estimator  $\hat{\gamma}_n(x)$  through a simulation study. To this end, we compare the bias and the mean squared error in four settings, where in each setting we have two degrees of dependence in the data (low and high intertemporal dependence) and for a fixed  $x$ , arising from the true tail index function  $\gamma(x) = 3x(x-1) + 1$ . More concretely, the settings considered are as follows.

- a) Conditional Fréchet distribution: The covariate process  $\{X_j\}$  is simulated from an autoregressive (AR) process of order 1 (with coefficient 0.1 and 0.9 in the low and high dependence cases respectively) transformed to have uniform marginals, similarly for the uniform process  $\{U_j\}$  that is then transformed into  $\{Y_j\}$  via the inverse Fréchet distribution function using the chosen  $\gamma(X_n)$  as tail index. Here we use the Sheather–Jones optimal bandwidth, fitted for each  $k_n$  on the  $k_n$  concomitant values of  $\{X_j\}$ , i.e. selected to correspond to the top  $k_n$  order statistics of  $\{Y_j\}$ . To evaluate the sample size, we consider  $n = 10^3, 10^4$ .
- b) Conditional Pareto distribution: Similar construction as the conditional Fréchet, using instead the Pareto distribution function. For a fixed sample size of  $n = 500$ , we compare the (global) optimal Sheather–Jones bandwidth fitted on the entire covariate process, fitted on the top  $k_n$  values of  $\{X_j\}$ , sorted according to the order statistics of  $\{Y_j\}$ , the method of equation (10) using equation (11) as conditional tail estimate and the choice  $h_n = \sqrt{\log(k_n)/n}$ .
- c) Consider a centred Gaussian process  $\mathbf{W}$  with  $\text{var}(W_j) = \sigma_j^2$  and Gaussian covariance kernel, and a covariate process  $\{X_j\}$ . Let  $\{P_i^z\}$  be the points of a Poisson point process on  $(0, \infty)$  having mean measure  $\nu_{1/\gamma(z)}$ , where  $\mathbb{R} \ni z \mapsto 1/\gamma(z) \in (0, \infty)$  is Lipschitz, and let  $\mathbf{W}^{(i)}, i \geq 1$  be i.i.d. copies of  $\mathbf{W}$ . Then the process  $\{Y_j\}$  defined by

$$Y_j = \bigvee_{i=1}^{\infty} P_i^{X_j} e^{W_j^{(i)} - (1/\gamma(X_j)) \sigma_j^2/2}, \quad (12)$$

is marginally regularly varying with tail index  $1/\gamma(x)$  on  $(X_j = x)$  and its construction is shown in Figure 7. We refer to this construction as a Conditional Subordinated Gaussian Max-Stable (CSGMS) process. In the simulation study we let  $\{X_j\}$  be i.i.d. standard uniform. To evaluate the sample size, we consider  $n = 10^3, 10^4$ , and length-scale  $l = 0.5, 2$  for all  $j$ , for the low and high dependence cases respectively.

- d) Conditional Fréchet with long memory: Similar construction as the conditional Fréchet, where the auto-regressive processes are replaced by auto-regressive fractional integrated moving averages (ARFIMA) with AR and moving average (MA) orders of 1, with AR coefficient 0.5, MA coefficient 0.2, i.e.

$$\left(1 - \frac{1}{2}B\right)(1 - B)^d X_t = \left(1 + \frac{1}{5}B\right)\varepsilon_t,$$

where  $B$  is the backshift operator, and  $\{\varepsilon_t\}$  are iid standard Gaussian noise terms. The fractional coefficient  $d$  is chosen, such that the decay of  $\beta$ -mixing coefficient is of polynomial order as opposed to exponential. To evaluate the sample size, we consider  $n = 10^3, 10^4$ .

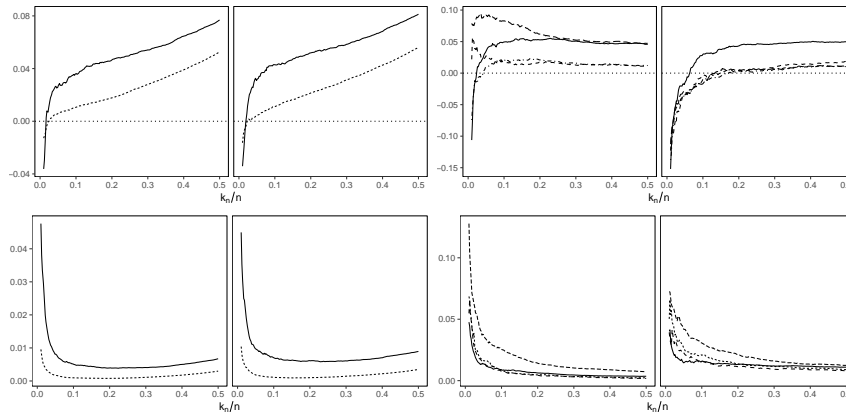
The conditional Fréchet and Pareto are known to satisfy the second-order condition of equation (4). By the remarks on page 224 in Mikosch and Wintenberger (2024), it holds that  $V_j^{(i)} := e^{W_j^{(i)} - (1/\gamma(X_j))\sigma_j^2/2}$  in equation (12) can be represented by  $(f_j(U_i))_{j=1}^\infty$  for suitable non-negative measurable functions  $f_j$  on  $(0, 1)$ , and  $\{U_i\}$  are i.i.d. uniform variables on  $(0, 1)$ . By eq. (5.5.7) therein, the CSGMS process above has conditional Fréchet marginals. In particular it also satisfies the condition of equation (4) and we note that all the three constructions satisfy equation (1). The mapping  $z \mapsto 1/\gamma(z) = 1/(3z(z-1) + 1)$  is clearly Lipschitz continuous and hence condition 1.1 holds. Condition 1.6 and equation (8) are verified for the conditional Fréchet and Pareto in Lemma 9, Appendix C, and anti-clustering conditions for the CSGMS hold due to Lemma 10, Appendix C. In the first two cases, the underlying processes are simulated from AR(1)-processes with auto-regressive coefficients absolutely bounded by 1, which implies  $\beta$ -mixing with mixing coefficients of exponential decay (see Ibragimov (1962)). When considering the unconditional version of the CSGMS-process, it is possible to show  $\beta$ -mixing decay of exponential order (see Kulik and Soulier (2020), Example 13.5.4). In equation (12) we extend the construction by bundling with the covariate sequence, which itself is i.i.d and in particular of exponential  $\beta$ -mixing decay. Since  $z \mapsto 1/\gamma(z)$  is Lipschitz, then also the bivariate sequence  $\{X_j, Y_j\}$  is of exponential  $\beta$ -mixing decay. Hence, since the conditional slowly varying functions in equation (1) are seen to satisfy condition 1.3 for both Pareto and Fréchet marginals, we may invoke Theorem 8 to conclude that there exists sequences  $\{(k_n, h_n, r_n, \ell_n)\}$  such that all conditions required in Theorem 5 are met in the above simulation studies.

Note that in the case when the conditional Fréchet process is constructed from an ARFIMA(1,  $d$ , 1)-process with  $d = 0.1, 0.45$  in the low and high dependence respectively, both cases imply that the  $\beta$ -mixing coefficients decay polynomially, still falling in the framework of Theorem 8 and 9. The results of the simulation study for  $k_n \in \{n/100, \dots, n/2\}$  are depicted in Figures 2 and 3, where we plot the bias and Mean Squared Error (MSE) of the estimator, evaluated in  $x = 0.6$ , as a function of  $k_n/n$ , averaged over 500 simulations.

For the Fréchet setting a), we observe that increasing the dependence structure in the underlying data, results in a similar finite-sample bias and MSE. In addition the bias exhibits unstable behaviour when  $k_n/n$  approaches zero, for the smaller sample size. This highlights the double sub-sampling used in conditional extreme estimators.

For the Pareto setting b), we focus on comparing the bandwidth selection techniques, as this distribution has no second-order component (no bias term). For the both dependence structure, the optimal

bandwidth on the upper  $k_n$  concomitants of  $X$ , yields slightly lower bias when  $k_n/n$  is away from zero. For  $k_n/n$  lower than approximately 0.1, the data-independent bandwidth  $h_n = \sqrt{\log(k_n)/n}$  results in the lowest mean squared error. We observe that neither of these two techniques are superior to another, and hence use the technique of lowest computational complexity, which is  $h_n = \sqrt{\log(k_n)/n}$  for the remaining simulations and real-data analysis.



**Figure 2.** Top panels: Bias. Bottom panels: MSE. We consider the conditional Fréchet model (low dependence (far left) and high dependence (center left))  $n = 10^3, 10^4$  in solid and dashed, respectively; number of simulations is  $N = 500$ . We also consider the conditional Pareto model (low dependence (center right) and high dependence (far right)) with  $n = 500, N = 200$ , and optimal bandwidth of  $\{X_j, Y_j\}$  using least-squares cross validation (long dash), optimal bandwidth of  $X$  (dot dash), optimal bandwidth of upper  $k_n$  concomitants of  $X$  (dash) and  $h_n = \sqrt{\log(k_n)/n}$  (solid).

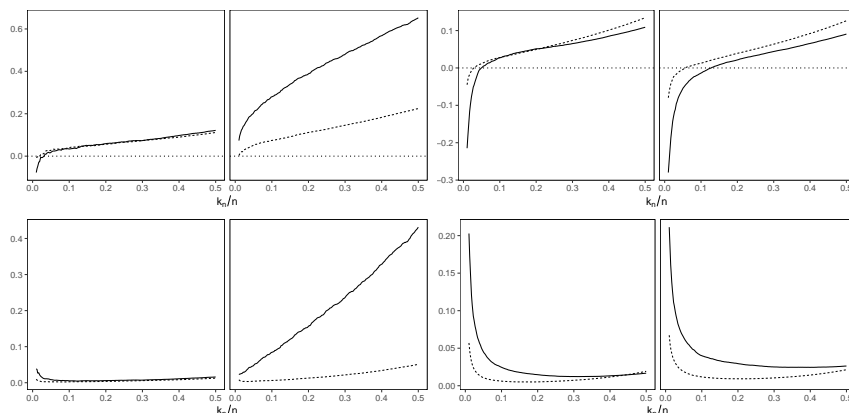
For the CSGMS setting c), we observe a strong positive bias for  $k_n/n$  away from zero for high dependence and low sample size. Increasing the sample size dramatically reduces the bias and MSE for the high degree of dependence, to the point where the high dependence scenario catches up to the performance of the low dependence scenario.

Finally, for the long memory setting d), which also falls in the scope of our asymptotic results, the estimator performs rather well with a negative bias for high quantiles, and positive bias for lower quantiles. In both cases the MSE stabilises as the sample size increases, and the negative bias at top quantiles diminishes. This suggests that the conditional Hill estimator can possibly be studied for time series with even stronger temporal dependence, albeit with possibly other techniques than the ones proposed in this paper.

## 6. Real data Analysis

### 6.1. Description of dataset

The West Texas Intermediate (WTI) is a grade of crude oil and is considered one of the main benchmarks in oil pricing. We are interested to examine whether the spot price of WTI is conditionally regularly varying given the Crude Oil Volatility Index (CBOE OVX) and, if this is the case, how the tail index is influenced by different levels of implied volatility. The CBOE OVX is constructed by applying



**Figure 3.** Top panels: Bias. Bottom panels: MSE. Sample sizes  $n = 10^3, 10^4$  in solid and dashed, respectively; number of simulations is  $N = 500$ . We consider the CSGMS model (low dependence (far left) and high dependence (center left)), and the conditional Fréchet model constructed from an underlying ARFIMA process (low dependence (center right) and high dependence (far right)).

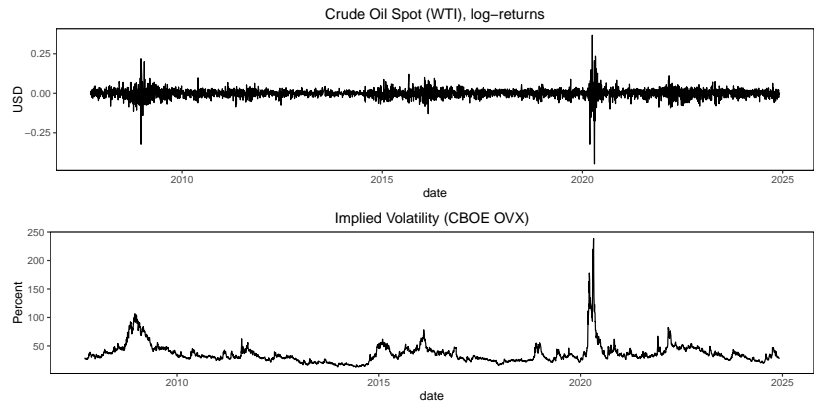
the VIX (Chicago Board Options Exchange’s Volatility Index) methodology (see [Whaley \(2000\)](#)) to options on the United States Oil Fund (USO), using a weighted average of mid-quote prices from calls and puts across a range of strike prices. This aggregated implied volatility is extrapolated to a 30-day horizon to provide a forward-looking measure of the expectation of the market of price fluctuations and is expressed in percent.

Now, as USO is designed to track the WTI price using future contracts, the OVX provides a forward looking market expectation of the WTI price volatility. Thus, in the case of regularly varying marginal distributions of the WTI, we expect that the tail index varies as a function of OVX. While the OVX is mean-reverting and can be considered stationary (see [Chen, Liu and Li \(2018\)](#)), oil prices are generally not stationary, and are consequently often transformed to log-returns (see [Hamilton \(1983\)](#), [Sadorsky \(1999\)](#)) to achieve stationarity. We follow this approach. The corresponding time series are depicted in [Figure 4](#). Note how high activity periods such as the financial crisis (2008), oil price collapse due to US and OPEC oil war (2015), covid period (2020-2021), and the Russian and Ukrainian conflict (2022), are visually observable. Remark that the WTI price attained a negative value on the 20th of April 2020, however this observation has been removed in order to transform with logarithms.

## 6.2. Analysis of crude oil spot price

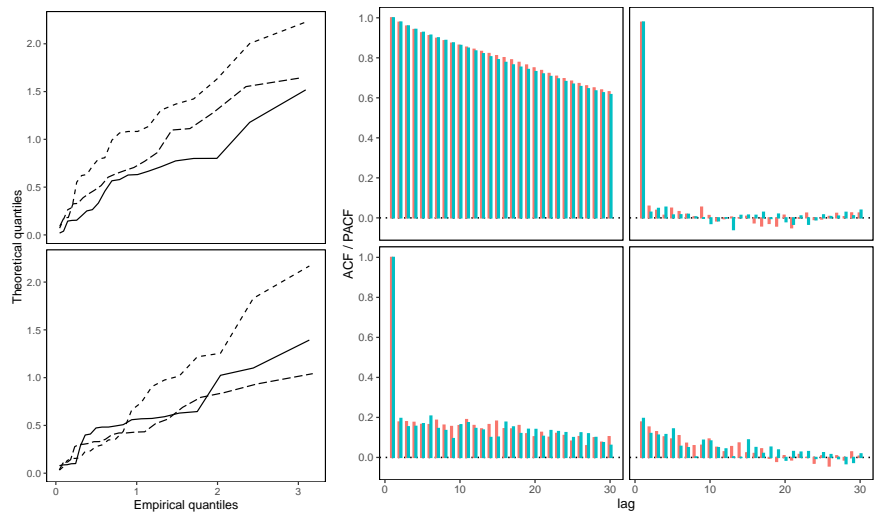
As seen from [Figure 4](#), the WTI log-returns attain both positive and negative values. By construction, the conditional Hill estimator is applicable only for positive  $\{Y_j\}$ . Splitting the bivariate time series  $\{X_j, Y_j\}$  according to where  $\{Y_j\}$  is positive and negative respectively, and switching the sign of the negative parts, results in two bivariate time series with positive  $\{Y_j\}$  components. The two time series are analysed separately and are seen to exhibit different tail behaviour, which allows for a more detailed analysis.

The covariate process OVX is, as in the simulation study, transformed to have uniform margins. Prior to employing our proposed conditional Hill estimator, we preliminarily check that the bivariate time series (positive or negative side) is in fact conditionally regularly varying. For this purpose, we



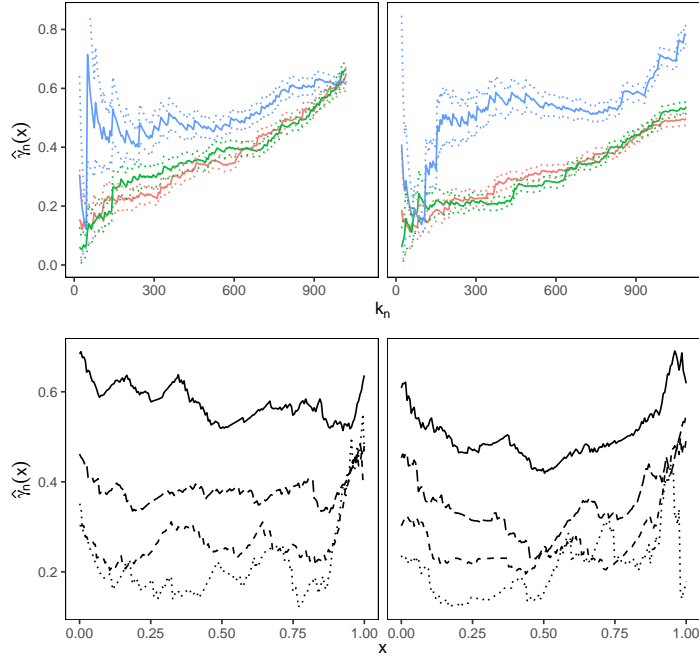
**Figure 4.** Log-returns of the WTI spot price (top) and oil volatility index CBOE OVX (bottom) from 2007 to end of 2024.

subsample the WTI log-returns for different values of (the uniformly transformed) OVX. This then results in the Pareto QQ-plots depicted in Figure 5, where the largest  $k_n = n/2$  observations are used. We clearly observe linearity in the QQ-plots, which suggests conditional regular variation. We can also already note that for both the positive and negative sides of the WTI log-returns, the slope is steepest for large implied volatility values.



**Figure 5.** Left panels: Pareto QQ-plots of the negative (top) and positive (bottom) WTI log-returns subsampled on uniformly transformed values of OVX in the intervals  $\{[0.01, 0.03], [0.49, 0.51], [0.96, 0.98]\}$  in dashed, long-dashed and solid respectively. Four right panels: Auto-correlation plots (left) and partial auto-correlation plots (right) of the negative (red) and positive (blue) of CBOE OVX data (top) and WTI log-returns (bottom).

In Figure 6 we consider the conditional Hill estimator of the WTI log-returns given the uniformly transformed OVX. We observe that the largest  $\hat{\gamma}_n(x)$  value is attained for both negative and positive



**Figure 6.** Plots of  $\hat{\gamma}_n(x)$ , using the negative (left) and positive (right) WTI log-returns as  $(Y_n)$  and the uniformly transformed OVX as  $(X_n)$  (covariate process). In the top panels we depict the conditional Hill estimator, as of function of  $k_n$ , i.e. the number of top order statistics included, for four chosen covariate values  $x \in \{0.231, 0.402, 0.995\}$  (red, green, and blue, respectively). In the bottom panels  $\hat{\gamma}_n(x)$  is plotted as a function of  $x$  for four fixed number of order statistics,  $k_n \in \{80, 241, 552, 959\}$  (dotted, dashed, long-dashed and solid, respectively).

sides of the WTI log-returns, when the implied volatility is large. This is in accordance with the common financial intuition that large implied volatility implies heavier tails (and thus fewer moments) for log-returns of markets on which it is based. Furthermore, when the implied volatility is large, the negative side of the WTI log-returns exhibits slightly heavier tails than the positive side. This has been documented for many assets (see [Christie \(1982\)](#)), where shock effects can create large upswings but even greater downswings. In addition, the confidence intervals that arise from [Theorem 5](#) are depicted. For the both sides, the lower bound of the estimator corresponding to high volatility is separated from the upper bound of the of the estimators, corresponding to low and mid volatility, within the stable regions. This supports the claim the the the WTI log-returns change tail index according the implied volatility. Since the conditional Hill estimator can be viewed as a function of the covariate value  $x$  for a fixed sample fraction  $k_n$  bivariate, we also plot these “risk profiles”, where slightly noisy right-skewed smiles can be seen, which indicates that the tail index is large when volatility is low and even larger when high.

## 7. Conclusion and outlook

We have shown that the conditional Hill estimator can be applied for a variety of time series models, in particular only  $\alpha$ -mixing being required to establish consistency. Numerical studies suggest good performance for both exponentially and polynomially decaying  $\beta$ -mixing coefficients, which is also

supported by the theory. The estimator might also be asymptotically convergent for stronger dependence structures, which is subject of further investigation.

Apart from considering stronger intertemporal dependence structures, our work also allows us to consider extensions in several directions. First, having an estimate of the entire tail at our disposal our results can be used to construct and statistically analyse conditional risk measures such as conditional expected shortfall or conditional tail moments. Secondly, a promising direction is to consider specific models where the same time series can be used as the covariate and response, but at different lags; Markov processes are particularly attractive in this connection. Finally, removing top observations can sometimes lead to the construction of stable estimators that allow transparent threshold selection as in [Bladt, Albrecher and Beirlant \(2020\)](#), which is yet to be explored both for univariate and conditional cases.

### **Funding**

The authors were supported by the Carlsberg Foundation, grant CF23-1096.

### **Data availability statement**

The data were accessed through a subscription-based platform and are not publicly available.

## References

- BILLINGSLEY, P. (2009). *Convergence of probability measures*, 2nd ed. ed. *Wiley series in probability and statistics. Probability and statistics section*. Wiley-Blackwell. <https://doi.org/10.1002/9780470316962>
- BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1987). *Regular Variation*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511721434>
- BLADT, M., ALBRECHER, H. and BEIRLANT, J. (2020). Threshold selection and trimming in extremes. *Extremes* **23** 629–665. <https://doi.org/10.1007/s10687-020-00385-0>
- BRADLEY, R. C. (2005). Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions. *Probability Surveys* **2** 107 – 144. <https://doi.org/10.1214/154957805100000104>
- CHEN, H., LIU, L. and LI, X. (2018). The predictive content of the CBOE crude oil volatility index. *Physica A: Statistical Mechanics and its Applications* **492** 837–850. <https://doi.org/10.1016/j.physa.2017.11.014>
- CHRISTIE, A. A. (1982). The Stochastic Behavior of Common Stock Variances: Value, Leverage and Interest Rate Effects. *Journal of Financial Economics* **10** 407–432. [https://doi.org/10.1016/0304-405X\(82\)90018-6](https://doi.org/10.1016/0304-405X(82)90018-6)
- CSÖRGŐ, S. and MASON, D. M. (1985). Central Limit Theorems for Sums of Extreme Values. *Mathematical Proceedings of the Cambridge Philosophical Society* **98** 547–558. <https://doi.org/10.1017/S03050004100063751>
- DAOUIA, A., GARDES, L. and GIRARD, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli* **19** 2557 – 2589. <https://doi.org/10.3150/12-BEJ466>
- DAOUIA, A., STUPFLER, G. and USSEGLIO-CARLEVÉ, A. (2023). Inference for Extremal Regression with Dependent Heavy-Tailed Data. *The Annals of Statistics* **51** 2040–2066. <https://doi.org/10.1214/23-AOS2320>
- DAOUIA, A., GARDES, L., GIRARD, S. and LEKINA, A. (2010). Kernel estimators of extreme level curves. *TEST* **20** 311–333. <https://doi.org/10.1007/s11749-010-0196-0>
- DAVIS, R. and RESNICK, S. (1984). Tail Estimates Motivated by Extreme Value Theory. *The Annals of Statistics* **12** 1467 – 1487. <https://doi.org/10.1214/aos/1176346804>
- DREES, H. (2000). Weighted approximations of tail processes for  $\beta$ -mixing random variables. *The Annals of Applied Probability* **10** 1274 – 1301. <https://doi.org/10.1214/aoap/1019487617>
- GANNOUN, A., GIRARD, S., GUINOT, C. and SARACCO, J. (2002). Reference curves based on non-parametric quantile regression. *Statistics in Medicine* **21** 3119–3135. <https://doi.org/10.1002/sim.1226>
- GOEGEBEUR, Y. and GUILLOU, A. (2024). Conditional Extreme Quantile Estimation for Time Series. HAL preprint: hal-04673935.
- HAEUSLER, E. and TEUGELS, J. L. (1985). On Asymptotic Normality of Hill’s Estimator for the Exponent of Regular Variation. *The Annals of Statistics* **13** 743 – 756. <https://doi.org/10.1214/aos/1176349551>
- HALL, P. (1982). On Some Simple Estimates of an Exponent of Regular Variation. *Journal of the Royal Statistical Society: Series B (Methodological)* **44** 37–42. <https://doi.org/10.1111/j.2517-6161.1982.tb01183.x>
- HALL, P. and WELSH, A. H. (1985). Adaptive Estimates of Parameters of Regular Variation. *The Annals of Statistics* **13** 331 – 341. <https://doi.org/10.1214/aos/1176346596>
- HAMILTON, J. D. (1983). Oil and the Macroeconomy since World War II. *Journal of Political Economy* **91** 228–248. <https://doi.org/10.1086/261140>
- HILL, B. M. (1975). A Simple General Approach to Inference About the Tail of a Distribution. *The Annals of Statistics* **3** 1163 – 1174. <https://doi.org/10.1214/aos/1176343247>
- HSING, T. (1991). Estimating the parameters of rare events. *Stochastic Processes and their Applications* **37** 117–139. [https://doi.org/10.1016/0304-4149\(91\)90064-J](https://doi.org/10.1016/0304-4149(91)90064-J)
- IBRAGIMOV, I. A. (1962). Some Limit Theorems for Stationary Processes. *Theory of Probability and Its Applications* **7** 349–382. <https://doi.org/10.1137/1107036>
- KULIK, R. and SOULIER, P. (2020). *Heavy-Tailed Time Series*. Springer New York. <https://doi.org/10.1007/978-1-0716-0737-4>
- LI, Q. and RACINE, J. S. (2008). Nonparametric Estimation of Conditional CDF and Quantile Functions with Mixed Categorical and Continuous Data. *Journal of Business & Economic Statistics* **26** 423–434. <https://doi.org/10.1198/073500107000000250>
- MIKOSCH, T. and WINTENBERGER, O. (2024). *Extreme Value Theory for Time Series: Models with Power-Law Tails*. *Springer Series in Operations Research and Financial Engineering*. Springer Nature, Netherlands. <https://doi.org/10.1007/978-3-031-59156-3>

- RESNICK, S. I. and STĂRICĂ, C. (1997). Asymptotic Behavior of Hill's Estimator for Autoregressive Data. *Communications in Statistics: Stochastic Models* **13** 703–721. <https://doi.org/10.1080/15326349708807448>
- RESNICK, S. I. and STĂRICĂ, C. (1995). Consistency of Hill's Estimator for Dependent Data. *Journal of Applied Probability* **32** 139–167. <https://doi.org/10.2307/3214926>
- ROOTZÉN, H. (1995). The Tail Empirical Process for Stationary Sequences, Technical Report, Chalmers University.
- ROOTZÉN, H. (2009). Weak Convergence of the Tail Empirical Process for Dependent Sequences. *Stochastic Processes and their Applications* **119** 468–490. <https://doi.org/10.1016/j.spa.2008.03.003>
- ROSENBLATT, M. (1956). A Central Limit Theorem and a Strong Mixing Condition. *Proceedings of the National Academy of Sciences of the United States of America* **42** 1 43-7. <https://doi.org/10.1073/pnas.42.1.43>
- SADORSKY, P. (1999). Oil Price Shocks and Stock Market Activity. *Energy Economics* **21** 449–469. [https://doi.org/10.1016/S0140-9883\(99\)00020-1](https://doi.org/10.1016/S0140-9883(99)00020-1)
- SHEATHER, S. J. and JONES, M. C. (1991). A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation. *Journal of the Royal Statistical Society. Series B (Methodological)* **53** 683–690.
- VAN DER VAART, A. W. and WELLNER, J. A. (2023). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer International Publishing. <https://doi.org/10.1007/978-3-031-29040-4>
- VOLKONSKII, V. A. and ROZANOV, Y. A. (1959). Some Limit Theorems for Random Functions. I. *Theory of Probability & Its Applications* **4** 178–197. <https://doi.org/10.1137/1104015>
- WANG, H. and TSAI, C.-L. (2009). Tail Index Regression. *Journal of the American Statistical Association* **104** 1233–1240. <https://doi.org/10.1198/jasa.2009.tm08458>
- WHALEY, R. E. (2000). The Investor Fear Gauge. *The Journal of Portfolio Management* **26** 12–17. <https://doi.org/10.3905/jpm.2000.319728>

## Appendix A: Proofs of main results from Section 3

### A.1. Proof of Theorem 1 (Tail consistency, $m$ -dependence)

We begin by proving some preliminary results related to the process  $\tilde{V}_n^x$ .

**Lemma 4.** *Assume conditions 1.1-1.3. Then*

$$\mathbb{E} [\tilde{V}_n^x(s)] \rightarrow s^{-1/\gamma(x)} g(x), \quad s > 0.$$

**Proof.** If  $g(x) > 0$  it holds that

$$\mathbb{E} [\tilde{V}_n^x(s)] = g(x) \frac{\bar{F}^x(su_n^x)}{\bar{F}^x(u_n^x)} \int (su_n^x)^{-(1/\gamma(x-uh_n)-1/\gamma(x))} \frac{L^{x-uh_n}(su_n^x)}{L^x(su_n^x)} \frac{g(x-h_nu)}{g(x)} K(u) du.$$

The expression before the integral converges to the desired limit, and hence it remains to show that the integral converges to 1. We consider each of the three  $n$ -dependent factors in the integral

$$\begin{aligned} (su_n^x)^{-(1/\gamma(x-uh_n)-1/\gamma(x))} &= \exp(-(1/\gamma(x-uh_n) - 1/\gamma(x)) \log(su_n^x)) \\ &\leq \exp(c_\gamma h_n |u| \log(su_n^x)) = \exp(c_\gamma |u| h_n \log(u_n^x)) \exp(c_\gamma h_n |u| \log(s)) \rightarrow 1, \end{aligned}$$

and similarly

$$(su_n^x)^{-(1/\gamma(x-uh_n)-1/\gamma(x))} \geq \exp(-c_\gamma |u| h_n \log(u_n^x)) \exp(-c_\gamma h_n |u| \log(s)) \rightarrow 1,$$

uniformly in  $u \in [-1, 1]$ , using condition 1.1 and  $h_n \log(u_n^x) \rightarrow 0$ . Now, condition 1.3 implies

$$\exp(-\log(su_n^x) h_n c_L |u|) \leq \frac{L^{x-uh_n}(su_n^x)}{L^x(su_n^x)} \leq \exp(\log(su_n^x) h_n c_L |u|)$$

and both sides converge to 1, uniformly in  $u \in [-1, 1]$ . The density  $g$  is assumed twice continuously differentiable which implies Lipschitz continuity, so that

$$1 - \frac{c_g |u| h_n}{g(x)} \leq \frac{g(x-h_nu)}{g(x)} \leq 1 + \frac{c_g |u| h_n}{g(x)}$$

and both sides converge to 1, uniformly in  $u \in [-1, 1]$ .

By the dominated convergence theorem get  $\mathbb{E}[\tilde{V}_n^x(s)] \rightarrow g(x)s^{-1/\gamma(x)}$ . If  $g(x) = 0$ , similar arguments, using Lipschitz continuity of  $g$ , yield the result.  $\square$

**Lemma 5.** *Assume conditions 1.1-1.3. For any  $t, q > 0$  it holds that*

$$\frac{\mathbb{E} \left[ 1_{\{Y > u_n^x t\}} K^q \left( \frac{x-X}{h_n} \right) \right]}{h_n \bar{F}^x(u_n^x)} \rightarrow t^{-1/\gamma(x)} g(x) \int K^q(u) du.$$

**Proof.** Calculations similar to those of Lemma 4 yield the result.  $\square$

**Proof of Theorem 1.** By assumption  $g_n(x) \xrightarrow{\text{Pr}} g(x)$ , and hence, using the relation

$\tilde{T}_n^x(s) = \tilde{V}_n^x(s)/g_n(x)$ , it remains to show that  $\tilde{V}_n^x(s) \xrightarrow{\text{Pr}} g(x)s^{-1/\gamma(x)}$ .

Let  $m \in \mathbb{N}$  be given and split  $\tilde{V}_n^x(\cdot)$  into a main component, which is a sum of independent terms, and a remainder term

$$\tilde{V}_n^x(s) = \frac{1}{m} \sum_{i=1}^m \tilde{V}_n^{x,i}(s) + R_n(s),$$

where

$$\begin{aligned} \tilde{V}_n^{x,i}(s) &= \frac{m}{nh_n \bar{F}^x(u_n^x)} \sum_{j=1}^{\lfloor \frac{n}{m} \rfloor} K\left(\frac{x - X_{(j-1)m+i}}{h_n}\right) 1_{\{Y_{(j-1)m+i} > su_n^x\}}, \\ R_n(s) &= \frac{1}{nh_n \bar{F}^x(u_n^x)} \sum_{j=\lfloor \frac{n}{m} \rfloor m}^n K\left(\frac{x - X_j}{h_n}\right) 1_{\{Y_j > su_n^x\}}. \end{aligned}$$

Stationarity and Lemma 5 yields that

$$\mathbb{E}[\tilde{V}_n^{x,i}(s)] = \frac{m}{n} \left\lfloor \frac{n}{m} \right\rfloor \frac{\mathbb{E}\left[1_{\{Y > u_n^x s\}} K\left(\frac{x-X}{h_n}\right)\right]}{h_n \bar{F}^x(u_n^x)} \rightarrow s^{-1/\gamma(x)} g(x),$$

while  $m$ -dependence implies

$$\text{var}\left(\tilde{V}_n^{x,i}(s)\right) = \frac{\frac{m}{n} \left\lfloor \frac{n}{m} \right\rfloor m}{nh_n \bar{F}^x(u_n^x)} \left[ \frac{\mathbb{E}\left[1_{\{Y > u_n^x s\}} K^2\left(\frac{x-X}{h_n}\right)\right]}{h_n \bar{F}^x(u_n^x)} - \frac{\mathbb{E}\left[1_{\{Y > u_n^x s\}} K\left(\frac{x-X}{h_n}\right)\right]^2}{h_n \bar{F}^x(u_n^x)} \right],$$

which vanishes by Lemma 5. Negligibility of the remainder term follows from

$$\begin{aligned} \mathbb{E}[R_n(s)] &= \left(n - \left\lfloor \frac{n}{m} \right\rfloor m\right) \frac{1}{n} \frac{\mathbb{E}\left[K\left(\frac{x-X}{h_n}\right) 1_{\{Y > su_n^x\}}\right]}{h_n \bar{F}^x(u_n^x)} \rightarrow 0, \\ \mathbb{E}[R_n(s)^2] &= \frac{1}{(nh_n \bar{F}^x(u_n^x))^2} \left( \sum_{j=\lfloor \frac{n}{m} \rfloor m}^n \mathbb{E}\left[K^2\left(\frac{x-X_j}{h_n}\right) 1_{\{Y_j > su_n^x\}}\right] \right. \\ &\quad \left. + 2 \sum_{j=\lfloor \frac{n}{m} \rfloor m}^n \sum_{i=j}^n \mathbb{E}\left[1_{\{Y_i > u_n^x s\}} K\left(\frac{x-X_i}{h_n}\right) 1_{\{Y_j > u_n^x s\}} K\left(\frac{x-X_j}{h_n}\right)\right] \right) \\ &\leq \frac{(n - \lfloor \frac{n}{m} \rfloor m)}{(nh_n \bar{F}^x(u_n^x))^2} \left( \mathbb{E}\left[K^2\left(\frac{x-X}{h_n}\right) 1_{\{Y > su_n^x\}}\right] + 2 \left(n - \left\lfloor \frac{n}{m} \right\rfloor m\right) \sqrt{\mathbb{E}\left[1_{\{Y > u_n^x s\}} K^2\left(\frac{x-X}{h_n}\right)\right]^2} \right) \\ &= \frac{(n - \lfloor \frac{n}{m} \rfloor m)}{n^2 h_n \bar{F}^x(u_n^x)} \frac{\mathbb{E}\left[1_{\{Y > u_n^x t\}} K^2\left(\frac{x-X}{h_n}\right)\right]}{h_n \bar{F}^x(u_n^x)} \left(1 + 2 \left(n - \left\lfloor \frac{n}{m} \right\rfloor m\right)\right) \rightarrow 0, \end{aligned}$$

where we used the Cauchy–Schwarz inequality. Then Dini’s theorem extends the convergence uniformly on compacts bounded away from zero. This finishes the proof.  $\square$

## A.2. Proof of Theorem 2 (Tail consistency, $\alpha$ -mixing)

**Proof of Theorem 2.** By assumption  $g_n(x) \xrightarrow{\text{Pr}} g(x)$  and Lemma 4 yields asymptotic unbiasedness of  $\widetilde{V}_n^x$ , and hence consistency of  $\widetilde{V}_n^x$  follows if the variance is negligible as  $n \rightarrow \infty$ . We have

$$\begin{aligned} \text{var}(\widetilde{V}_n^x(s)) &= \frac{1}{(nh_n \overline{F}^x(u_n^x))^2} \left( \sum_{j=1}^n \mathbb{E} \left[ K^2 \left( \frac{x - X_j}{h_n} \right) 1_{\{Y_j > su_n^x\}} \right] - n^2 \mathbb{E} \left[ K \left( \frac{x - X_0}{h_n} \right) 1_{\{Y_0 > su_n^x\}} \right]^2 \right. \\ &\quad \left. + 2n \sum_{j=1}^n \left( 1 - \frac{j}{n} \right) \mathbb{E} \left[ K \left( \frac{x - X_0}{h_n} \right) 1_{\{Y_0 > su_n^x\}} K \left( \frac{x - X_j}{h_n} \right) 1_{\{Y_j > su_n^x\}} \right] \right) \\ &\leq \frac{1}{(nh_n \overline{F}^x(u_n^x))^2} \left( \sum_{j=1}^n \mathbb{E} \left[ K^2 \left( \frac{x - X_j}{h_n} \right) 1_{\{Y_j > su_n^x\}} \right] \right. \\ &\quad \left. + 2n \sum_{j=1}^n \text{cov} \left( K \left( \frac{x - X_0}{h_n} \right) 1_{\{Y_0 > su_n^x\}}, K \left( \frac{x - X_j}{h_n} \right) 1_{\{Y_j > su_n^x\}} \right) \right), \end{aligned}$$

where the first term vanishes due to Lemma 5 and condition 1.4. The leading term of the covariance behaves as

$$\begin{aligned} &\frac{1}{h_n \overline{F}^x(u_n^x)} \mathbb{E} \left[ K \left( \frac{x - X_0}{h_n} \right) 1_{\{Y_0 > su_n^x\}} K \left( \frac{x - X_j}{h_n} \right) 1_{\{Y_j > su_n^x\}} \right] \\ &= h_n \int_{[-1,1]^2} \left\{ \frac{\mathbb{E} \left[ 1_{\{Y_0 > su_n^x\}} 1_{\{Y_j > su_n^x\}} \mid X_0 = x + w_1 h_n, X_j = x + w_2 h_n \right]}{\overline{F}^x(u_n^x)} \right. \\ &\quad \left. \times g(x + w_1 h_n, x + w_2 h_n) K(w_1) K(w_2) \right\} d(w_1, w_2) = \mathcal{O}(h_n), \end{aligned}$$

where the integral is bounded, for  $n$  large enough, due to condition 1.6 and smoothness of  $g_{X_0, X_j}$ . Now, Billingsley’s inequality with the  $\alpha$ -mixing assumption yields that

$$\begin{aligned} &\left| \text{cov} \left( K \left( \frac{x - X_0}{h_n} \right) 1_{\{Y_0 > su_n^x\}}, K \left( \frac{x - X_j}{h_n} \right) 1_{\{Y_j > su_n^x\}} \right) \right| \\ &\leq 4\alpha(j) \left\| K \left( \frac{x - X_0}{h_n} \right) 1_{\{Y_0 > su_n^x\}} \right\|_{\infty} \left\| K \left( \frac{x - X_j}{h_n} \right) 1_{\{Y_j > su_n^x\}} \right\|_{\infty} = \mathcal{O}(\alpha(j)), \end{aligned}$$

using that the kernel is bounded. Let  $d_n = \lceil h_n^{-c_x} \rceil$  for some  $c_x > 0$ . Combing the above considerations yields that

$$\frac{2n}{(nh_n \overline{F}^x(u_n^x))^2} \left| \sum_{j=1}^n \text{cov} \left( K \left( \frac{x - X_0}{h_n} \right) 1_{\{Y_0 > su_n^x\}}, K \left( \frac{x - X_j}{h_n} \right) 1_{\{Y_j > su_n^x\}} \right) \right|$$

$$\begin{aligned}
&\leq \frac{\text{cst}}{n(h_n \bar{F}^x(u_n^x))^2} \left( \sum_{j=1}^{d_n-1} h_n^2 \bar{F}^x(u_n^x) + \sum_{j=d_n}^{\infty} j^{-\eta} \right) \leq \frac{\text{cst}}{n(h_n \bar{F}^x(u_n^x))^2} \left( d_n h_n^2 \bar{F}^x(u_n^x) + d_n^{-\eta+1} \right) \\
&= \text{cst} \left( \frac{d_n}{n \bar{F}^x(u_n^x)} + \frac{d_n^{-\eta+1}}{n(h_n \bar{F}^x(u_n^x))^2} \right) \sim \text{cst} \left( \frac{1}{n h_n^{c_x} \bar{F}^x(u_n^x)} + \frac{h_n^{c_x(\eta-1)-2}}{n(\bar{F}^x(u_n^x))^2} \right) \rightarrow 0.
\end{aligned}$$

Once again, Dini's theorem extends the convergence uniformly on compacts bounded away from zero.  $\square$

### A.3. Proof of Theorem 3 (Hill consistency)

**Proof of Proposition 1.** Let  $\varepsilon > 0$  be given and assume that  $q_{n,k_n}^x/u_n^x > 1 + \varepsilon$  for all  $n \in \mathbb{N}$ . Since  $\bar{F}_n^x$  is decreasing we have

$$\tilde{T}_n^x(1 + \varepsilon) = \frac{\bar{F}_n^x((1 + \varepsilon)u_n^x)}{\bar{F}^x(u_n^x)} \geq \frac{\bar{F}_n^x(q_{n,k_n}^x)}{\bar{F}^x(u_n^x)} = \frac{\bar{F}_n^x(F_n^{\leftarrow, x}(1 - k_n/n))}{\bar{F}^x(F^{\leftarrow, x}(1 - k_n/n))} = 1.$$

In particular it holds that

$$\Pr \left( \frac{q_{n,k_n}^x}{u_n^x} > 1 + \varepsilon \right) \leq \Pr \left( \tilde{T}_n^x(1 + \varepsilon) > 1 \right).$$

Since (2) holds, then  $\tilde{T}_n^x(1 + \varepsilon) \xrightarrow{\text{Pr}} (1 + \varepsilon)^{-1/\gamma(x)}$  and thus

$$\limsup_{n \rightarrow \infty} \Pr \left( \frac{q_{n,k_n}^x}{u_n^x} > 1 + \varepsilon \right) \leq \lim_{n \rightarrow \infty} \Pr \left( \tilde{T}_n^x(1 + \varepsilon) > 1 \right) = 0.$$

Using similar arguments we obtain that  $\limsup_{n \rightarrow \infty} \Pr(q_{n,k_n}^x/u_n^x < 1 - \varepsilon) = 0$ .  $\square$

**Proof of Theorem 3.** We consider the following decomposition

$$\hat{\gamma}_n(x) = \int_{q_{n,k_n}^x/u_n^x}^1 \frac{\tilde{T}_n^x(s)}{s} ds + \int_1^{\infty} \frac{\tilde{T}_n^x(s)}{s} ds.$$

Proposition 1 yields that  $q_{n,k_n}^x/u_n^x \xrightarrow{\text{Pr}} 1$  and thus the first integral is  $o_{\text{Pr}}(1)$ .

The convergence of equation (2) is uniform on compacts of  $[1, \infty)$  and hence for all  $t > 1$  we have that

$$\int_1^t \frac{\tilde{T}_n^x(s)}{s} ds \xrightarrow{\text{Pr}} \int_1^t \frac{T^x(s)}{s} ds, \quad \text{as } n \rightarrow \infty.$$

Since  $\lim_{t \rightarrow \infty} \int_1^t (T^x(s)/s) ds = \gamma(x)$ , consistency of  $\hat{\gamma}_n$  follows if for every  $\eta > 0$

$$\lim_{t \rightarrow \infty} \limsup_{n \rightarrow \infty} \Pr \left( \int_t^{\infty} \frac{\tilde{T}_n^x(s)}{s} ds > \eta \right) = 0,$$

Now, since

$$\lim_{t \rightarrow \infty} \int_t^\infty \frac{\tilde{T}_n^x(s)}{s} ds = \frac{g(x)}{g_n(x)} \lim_{t \rightarrow \infty} \int_t^\infty \frac{\tilde{V}_n^x(s)}{g(x)s} ds,$$

and  $g(x)/g_n(x) \xrightarrow{\text{Pr}} 1$ , it suffices to show that for every  $\eta > 0$ ,

$$\lim_{t \rightarrow \infty} \limsup_{n \rightarrow \infty} \Pr \left( \int_t^\infty \frac{\tilde{V}_n^x(s)}{g(x)s} ds > \eta \right) = 0.$$

Fix  $\eta > 0$ . We get

$$\begin{aligned} \Pr \left( \int_t^\infty \frac{\tilde{V}_n^x(s)}{g(x)s} ds > \eta \right) &\leq \frac{1}{\eta} \int_t^\infty \frac{\mathbb{E} [\tilde{V}_n^x(s)]}{g(x)s} ds \\ &= \frac{1}{\eta} \int_t^\infty \frac{\bar{F}^x(su_n^x)}{s\bar{F}^x(u_n^x)} \int (su_n^x)^{-(1/\gamma(x)-uh_n)-1/\gamma(x)} \frac{L^{x-uh_n}(su_n^x)}{L^x(su_n^x)} \frac{g(x-h_nu)}{g(x)} K(u) du ds \\ &\leq \frac{1}{\eta} \int_t^\infty \frac{\bar{F}^x(su_n^x)}{s\bar{F}^x(u_n^x)} \sup_{u \in [-1,1]} \left( (su_n^x)^{-(1/\gamma(x)-uh_n)-1/\gamma(x)} \frac{L^{x-uh_n}(su_n^x)}{L^x(su_n^x)} \frac{g(x-h_nu)}{g(x)} \right) ds \\ &\leq \frac{1}{\eta} \int_t^\infty \frac{\bar{F}^x(su_n^x)}{s\bar{F}^x(u_n^x)} \sup_{u \in [-1,1]} \left( \frac{g(x-h_nu)}{g(x)} \right) s^{(c_\gamma+c_L)h_n} (u_n^x)^{(c_\gamma+c_L)h_n} ds \\ &\leq \frac{\text{cst}}{\eta} \sup_{u \in [-1,1]} \left( \frac{g(x-h_nu)}{g(x)} \right) (u_n^x)^{(c_\gamma+c_L)h_n} \int_t^\infty s^{-1/\gamma(x)+\varepsilon-1+(c_\gamma+c_L)h_n} ds, \end{aligned}$$

for  $\varepsilon \in (0, 1/\gamma(x))$ , using Potter's bound. Let  $\varepsilon' \in (0, 1/\gamma(x) - \varepsilon)$  be given. It is possible to find  $n'$  such that  $(c_\gamma + c_L)h_n \leq \varepsilon'$  for all  $n \geq n'$ . Hence for  $n \geq n'$  we have that the integrand  $s \mapsto s^{-1/\gamma(x)+\varepsilon-1+(c_\gamma+c_L)h_n}$  is integrable and consequently

$$\begin{aligned} &\limsup_{n \rightarrow \infty} \Pr \left( \int_t^\infty \frac{\tilde{V}_n^x(s)}{g(x)s} ds > \eta \right) \\ &\leq \limsup_{n \rightarrow \infty} \left\{ \frac{\text{cst}}{\eta} \sup_{u \in [-1,1]} \left( \frac{g(x-h_nu)}{g(x)} \right) (u_n^x)^{(c_\gamma+c_L)h_n} \int_t^\infty s^{-1/\gamma(x)+\varepsilon-1+(c_\gamma+c_L)h_n} ds \right\} \\ &= O \left( t^{-1/\gamma(x)+\varepsilon+\varepsilon'} \right), \end{aligned}$$

which vanishes as  $t \rightarrow \infty$ . □

#### A.4. Proof of Theorem 4 (Functional CLT)

Given the original sample  $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$ , we consider a pseudo-sample

$$\left\{ (X_1^\dagger, Y_1^\dagger), \dots, (X_n^\dagger, Y_n^\dagger) \right\},$$

where the blocks

$$\left\{ \left( X_{(i-1)r_n+1}^\dagger, Y_{(i-1)r_n+1}^\dagger \right), \dots, \left( X_{ir_n}^\dagger, Y_{ir_n}^\dagger \right) \right\}, \quad i = 1, \dots, m_n,$$

are mutually independent and each have the same distribution as the first original sample block  $\{(X_1, Y_1), \dots, (X_{r_n}, Y_{r_n})\}$ .

Define for every  $s > 0$

$$\begin{aligned} \widetilde{V}_n^{\dagger, x}(s) &= \frac{1}{nh_n \overline{F}^x(u_n^x)} \sum_{i=1}^{m_n} \sum_{j=1}^{r_n} K\left(\frac{x - X_{(i-1)r_n+j}^\dagger}{h_n}\right) 1_{\{Y_{(i-1)r_n+j}^\dagger > su_n^x\}}, \\ \widetilde{\mathbb{V}}_n^{\dagger, x}(s) &= \sqrt{k_n h_n} \left\{ \widetilde{V}_n^{\dagger, x}(s) - \mathbb{E} \left[ \widetilde{V}_n^{\dagger, x}(s) \right] \right\}. \end{aligned}$$

We establish weak convergence for  $\widetilde{\mathbb{V}}_n^{\dagger, x}$  and relate it to the original  $\widetilde{\mathbb{V}}_n^x$  and in turn  $\widetilde{\mathbb{T}}_n^x$ . The first step is to establish finite dimensional convergence for  $\widetilde{\mathbb{V}}_n^{\dagger, x}$ .

**Lemma 6.** *Assume conditions 1.1-1.4, and 2.1-2.2. Then for all  $s, t > 0$*

$$\lim_{n \rightarrow \infty} \text{cov} \left( \widetilde{\mathbb{V}}_n^{\dagger, x}(s), \widetilde{\mathbb{V}}_n^{\dagger, x}(t) \right) = (s \vee t)^{-1/\gamma(x)} g(x) \int K^2(u) du. \quad (13)$$

**Proof.** Independence between the blocks of the pseudo-sample and stationarity yield

$$\begin{aligned} \text{cov} \left( \widetilde{\mathbb{V}}_n^{\dagger, x}(s), \widetilde{\mathbb{V}}_n^{\dagger, x}(t) \right) &= \frac{m_n r_n}{nh_n \overline{F}^x(u_n^x)} \left[ \mathbb{E} \left[ 1_{\{Y_0 > (s \vee t) u_n^x\}} K^2\left(\frac{x - X_0}{h_n}\right) \right] \right. \\ &\quad - r_n \mathbb{E} \left[ K\left(\frac{x - X_0}{h_n}\right) 1_{\{Y_0 > su_n^x\}} \right] \mathbb{E} \left[ K\left(\frac{x - X_0}{h_n}\right) 1_{\{Y_0 > tu_n^x\}} \right] \\ &\quad + \sum_{j=1}^{r_n} \left( 1 - \frac{j}{r_n} \right) \left\{ \mathbb{E} \left[ K\left(\frac{x - X_0}{h_n}\right) 1_{\{Y_0 > su_n^x\}} K\left(\frac{x - X_j}{h_n}\right) 1_{\{Y_j > tu_n^x\}} \right] \right. \\ &\quad \left. + \mathbb{E} \left[ K\left(\frac{x - X_0}{h_n}\right) 1_{\{Y_0 > tu_n^x\}} K\left(\frac{x - X_j}{h_n}\right) 1_{\{Y_j > su_n^x\}} \right] \right\} \Bigg]. \end{aligned}$$

Since  $m_n = \lfloor n/r_n \rfloor$  we have that  $\lim_{n \rightarrow \infty} m_n r_n / n = 1$ . The limit of the first term follows from Lemma 5

$$\lim_{n \rightarrow \infty} \frac{m_n r_n}{n} \frac{\mathbb{E} \left[ 1_{\{Y_0 > (s \vee t) u_n^x\}} K^2\left(\frac{x - X_0}{h_n}\right) \right]}{h_n \overline{F}^x(u_n^x)} = (s \vee t)^{-1/\gamma(x)} g(x) \int K^2(u) du,$$

which is the asymptotic covariance. It remains to show that the second and third terms disappear asymptotically. The second term vanishes by the assumption  $\lim_{n \rightarrow \infty} r_n h_n \overline{F}^x(u_n^x) = 0$ . Since  $m_n r_n / n = \mathcal{O}(1)$ , the leading term in the third term is asymptotically proportional to

$$\frac{1}{h_n \overline{F}^x(u_n^x)} \sum_{j=1}^{r_n} \mathbb{E} \left[ K\left(\frac{x - X_0}{h_n}\right) 1_{\{Y_0 > tu_n^x\}} K\left(\frac{x - X_j}{h_n}\right) 1_{\{Y_j > su_n^x\}} \right].$$

To examine the limit, fix  $m \in \mathbb{N}$ , and consider

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \frac{1}{h_n \bar{F}^x(u_n^x)} \sum_{j=1}^{m-1} \mathbb{E} \left[ K \left( \frac{x - X_0}{h_n} \right) 1_{\{Y_0 > t u_n^x\}} K \left( \frac{x - X_j}{h_n} \right) 1_{\{Y_j > s u_n^x\}} \right] \\ &= \limsup_{n \rightarrow \infty} \sum_{j=1}^{m-1} h_n \int_{[-1,1]^2} \left\{ \frac{\mathbb{E} \left[ 1_{\{Y_0 > t u_n^x\}} 1_{\{Y_j > s u_n^x\}} \mid X_0 = x + w_1 h_n, X_j = x + w_2 h_n \right]}{\bar{F}^x(u_n^x)} \right. \\ & \quad \left. \times g(x + w_1 h_n, x + w_2 h_n) K(w_1) K(w_2) \right\} d(w_1, w_2) = 0. \end{aligned}$$

Fix  $\varepsilon > 0$ . By condition 2.2, it is possible to pick  $m_0$  large enough so that for all  $m \geq m_0$

$$\limsup_{n \rightarrow \infty} \frac{1}{h_n \bar{F}^x(u_n^x)} \sum_{j=m}^{r_n} \mathbb{E} \left[ K \left( \frac{x - X_0}{h_n} \right) 1_{\{Y_0 > t u_n^x\}} K \left( \frac{x - X_j}{h_n} \right) 1_{\{Y_j > s u_n^x\}} \right] \leq \varepsilon.$$

This finishes the proof.  $\square$

**Theorem 10.** Assume conditions 1.1-1.4, 2.1-2.2. Then the finite dimensional distributions of  $\tilde{\mathbb{V}}_n^{\dagger, x}$  converge to those of a Gaussian process  $\mathbb{V}^x$  with covariance function defined in (13).

**Proof.** For  $i = 1, \dots, m_n$ , let

$$\begin{aligned} S_{n,i}^x(s) &= \sum_{j=1}^{r_n} K \left( \frac{x - X_{(i-1)r_n+j}}{h_n} \right) 1_{\{Y_{(i-1)r_n+j} > s u_n^x\}}, & \bar{S}_{n,i}^x(s) &= S_{n,i}^x(s) - \mathbb{E} \left[ S_{n,1}^x(s) \right], \\ S_{n,i}^{\dagger, x}(s) &= \sum_{j=1}^{r_n} K \left( \frac{x - X_{(i-1)r_n+j}^{\dagger}}{h_n} \right) 1_{\{Y_{(i-1)r_n+j}^{\dagger} > s u_n^x\}}, & \bar{S}_{n,i}^{\dagger, x}(s) &= S_{n,i}^{\dagger, x}(s) - \mathbb{E} \left[ S_{n,1}^{\dagger, x}(s) \right], \end{aligned}$$

for  $s > 0$ . Then  $\tilde{\mathbb{V}}_n^{\dagger, x}$  can be written as

$$\tilde{\mathbb{V}}_n^{\dagger, x}(s) = \frac{1}{\sqrt{k_n h_n}} \sum_{i=1}^{m_n} \bar{S}_{n,i}^{\dagger, x}(s),$$

where the summands are i.i.d. In the calculations below we show that the conditions of the Lindeberg-Lévy CLT are fulfilled. By Lemma 6 it only remains to show the negligibility condition. Since  $\bar{S}_{n,1}^x(s)$  and  $\bar{S}_{n,1}^{\dagger, x}(s)$  have the same distribution, we may establish it for the former. Note that  $m_n / (n h_n \bar{F}^x(u_n^x)) = O((r_n h_n \bar{F}^x(u_n^x))^{-1})$ . Thus the negligibility condition thus follows if the expectation below is  $o(r_n h_n \bar{F}^x(u_n^x))$ :

$$\begin{aligned} & \mathbb{E} \left[ \left( \bar{S}_{n,1}^x(s) \right)^2 1_{\{|\bar{S}_{n,1}^x(s)| > \varepsilon \sqrt{k_n h_n}\}} \right] = \\ & \mathbb{E} \left[ \left( \sum_{j=1}^{r_n} K \left( \frac{x - X_j}{h_n} \right) 1_{\{Y_j > s u_n^x\}} \right)^2 1_{\{|\bar{S}_{n,1}^x(s)| > \varepsilon \sqrt{k_n h_n}\}} \right] \end{aligned}$$

$$\begin{aligned}
& + \left( r_n \mathbb{E} \left[ K \left( \frac{x - X_0}{h_n} \right) 1_{\{Y_0 > su_n^x\}} \right] \right)^2 \Pr \left( \left| \bar{S}_{n,1}^x(s) \right| > \varepsilon \sqrt{k_n h_n} \right) \\
& - 2r_n \mathbb{E} \left[ K \left( \frac{x - X_0}{h_n} \right) 1_{\{Y_0 > su_n^x\}} \right] \sum_{j=1}^{r_n} \mathbb{E} \left[ K \left( \frac{x - X_j}{h_n} \right) 1_{\{Y_j > su_n^x\}} 1_{\left\{ \left| \bar{S}_{n,1}^x(s) \right| > \varepsilon \sqrt{k_n h_n} \right\}} \right].
\end{aligned}$$

Consider the second term. Lemma 5 yields that it is bounded by

$$r_n^2 \mathbb{E} \left[ K \left( \frac{x - X_0}{h_n} \right) 1_{\{Y_0 > su_n^x\}} \right]^2 = r_n^2 \mathcal{O} \left( \left( h_n \bar{F}^x(u_n^x) \right)^2 \right) = o \left( r_n h_n \bar{F}^x(u_n^x) \right).$$

The third term follows similarly. We turn the first term and decompose it according to an expansion of the square, yielding diagonal terms  $I_1$  and off-diagonal terms  $2I_2$ . We get

$$\begin{aligned}
I_1 & \leq \frac{1}{\varepsilon \sqrt{k_n h_n}} \sum_{j=1}^{r_n} \mathbb{E} \left[ K \left( \frac{x - X_j}{h_n} \right)^2 1_{\{Y_j > su_n^x\}} \left| \bar{S}_{n,1}^x(s) \right| \right] \\
& \leq \frac{1}{\varepsilon \sqrt{k_n h_n}} \sum_{j=1}^{r_n} \mathbb{E} \left[ K \left( \frac{x - X_j}{h_n} \right)^2 1_{\{Y_j > su_n^x\}} \left( S_{n,1}^x(s) + \mathbb{E} \left[ S_{n,1}^x(s) \right] \right) \right]
\end{aligned}$$

The latter of the two terms is of order

$$\frac{\mathcal{O} \left( r_n h_n \bar{F}^x(u_n^x) \right)}{\varepsilon \sqrt{k_n h_n}} \sum_{j=1}^{r_n} \mathbb{E} \left[ K \left( \frac{x - X_j}{h_n} \right)^2 1_{\{Y_j > su_n^x\}} \right] = \frac{\mathcal{O} \left( \left( r_n h_n \bar{F}^x(u_n^x) \right)^2 \right)}{\varepsilon \sqrt{k_n h_n}} = o \left( r_n h_n \bar{F}^x(u_n^x) \right),$$

using condition 2.1 and Lemma 5. The former of the two terms can be rewritten as

$$\begin{aligned}
& \frac{r_n}{\varepsilon \sqrt{k_n h_n}} \left( \mathbb{E} \left[ K \left( \frac{x - X_0}{h_n} \right)^3 1_{\{Y_0 > su_n^x\}} \right] \right. \\
& \quad + \sum_{j=1}^{r_n} \left( 1 - \frac{j}{r_n} \right) \left\{ \mathbb{E} \left[ K \left( \frac{x - X_j}{h_n} \right)^2 K \left( \frac{x - X_0}{h_n} \right) 1_{\{Y_j > su_n^x\}} 1_{\{Y_0 > su_n^x\}} \right] \right. \\
& \quad \left. \left. + \mathbb{E} \left[ K \left( \frac{x - X_0}{h_n} \right)^2 K \left( \frac{x - X_j}{h_n} \right) 1_{\{Y_0 > su_n^x\}} 1_{\{Y_j > su_n^x\}} \right] \right\} \right),
\end{aligned}$$

where the first term is order  $o(r_n h_n \bar{F}^x(u_n^x))$  as above using condition 2.1 and Lemma 5. The leading term in the sum is dealt with as in the proof of Lemma 6, from which we infer that the expression is  $o(r_n h_n \bar{F}^x(u_n^x))$  and consequently  $I_1 = o(r_n h_n \bar{F}^x(u_n^x))$ .

Consider  $I_2$ . Given  $\delta > 0$ , condition 2.2 yields that we can choose a positive integer  $m$  such that for sufficiently large  $n$  it holds that

$$\sum_{j=m}^{r_n} \frac{\mathbb{E} \left[ K \left( \frac{x - X_j}{h_n} \right) K \left( \frac{x - X_0}{h_n} \right) 1_{\{Y_j > su_n^x\}} 1_{\{Y_0 > su_n^x\}} \right]}{h_n \bar{F}^x(u_n^x)} \leq \delta.$$

Split the inner sum of  $I_2$  according to the chosen  $m$ .

This yields two terms. By stationarity the first term can be rewritten as

$$\begin{aligned}
& \sum_{j=1}^m (m-j) \mathbb{E} \left[ K \left( \frac{x-X_0}{h_n} \right) K \left( \frac{x-X_j}{h_n} \right) 1_{\{Y_0 > su_n^x\}} 1_{\{Y_j > su_n^x\}} 1_{\{|\bar{S}_{n,1}^x(s)| > \varepsilon \sqrt{k_n h_n}\}} \right] \\
& \leq m \sum_{j=1}^m \mathbb{E} \left[ K \left( \frac{x-X_0}{h_n} \right) K \left( \frac{x-X_j}{h_n} \right) 1_{\{Y_0 > su_n^x\}} 1_{\{Y_j > su_n^x\}} \right] \\
& \leq \frac{m}{r_n} r_n h_n \bar{F}^x(u_n^x) \sum_{j=1}^{r_n} \frac{\mathbb{E} \left[ K \left( \frac{x-X_0}{h_n} \right) K \left( \frac{x-X_j}{h_n} \right) 1_{\{Y_0 > su_n^x\}} 1_{\{Y_j > su_n^x\}} \right]}{h_n \bar{F}^x(u_n^x)} \\
& = \frac{O(r_n h_n \bar{F}^x(u_n^x))}{r_n} = o(r_n h_n \bar{F}^x(u_n^x)).
\end{aligned}$$

The second term of  $I_2$  satisfies

$$\sum_{i=1}^{r_n} \sum_{j=i+m}^{r_n} \mathbb{E} \left[ K \left( \frac{x-X_i}{h_n} \right) K \left( \frac{x-X_j}{h_n} \right) 1_{\{Y_i > su_n^x\}} 1_{\{Y_j > su_n^x\}} 1_{\{|\bar{S}_{n,1}^x(s)| > \varepsilon \sqrt{k_n h_n}\}} \right] \leq \delta r_n h_n \bar{F}^x(u_n^x).$$

Gathering the above yields that

$$\limsup_{n \rightarrow \infty} \frac{I_2}{r_n h_n \bar{F}^x(u_n^x)} \leq \delta,$$

and since  $\delta$  can be made arbitrarily small we conclude that  $I_2 = o(r_n h_n \bar{F}^x(u_n^x))$  which concludes the proof.  $\square$

Now we show that terms constructed from the incomplete block  $\{(X_{m_n r_n + 1}, Y_{m_n r_n + 1}), \dots, (X_n, Y_n)\}$  are asymptotically negligible.

**Lemma 7.** *Assume conditions 1.1-1.4 and 2.1. Then*

$$\sup_{s \in [a, b]} \left| \sum_{j=m_n r_n + 1}^n \frac{K \left( \frac{x-X_j}{h_n} \right) 1_{\{Y_j > su_n^x\}} - \mathbb{E} \left[ K \left( \frac{x-X_j}{h_n} \right) 1_{\{Y_j > su_n^x\}} \right]}{\sqrt{nh_n \bar{F}^x(u_n^x)}} \right| \xrightarrow{\text{Pr}} 0.$$

**Proof.** Since  $s \mapsto K((x-X_j)/h_n) 1_{\{Y_j > su_n^x\}}$  is monotonically decreasing, and  $(n - m_n r_n) \leq r_n$ , i.e. the last incomplete block consists of fewer terms than a complete block, we get

$$\begin{aligned}
& \sup_{s \in [a, b]} \left| \sum_{j=m_n r_n + 1}^n \frac{K \left( \frac{x-X_j}{h_n} \right) 1_{\{Y_j > su_n^x\}} - \mathbb{E} \left[ K \left( \frac{x-X_j}{h_n} \right) 1_{\{Y_j > su_n^x\}} \right]}{\sqrt{nh_n \bar{F}^x(u_n^x)}} \right| \\
& \leq \frac{1}{\sqrt{nh_n \bar{F}^x(u_n^x)}} \left\{ \sum_{j=m_n r_n + 1}^n K \left( \frac{x-X_j}{h_n} \right) 1_{\{Y_j > au_n^x\}} + r_n \mathbb{E} \left[ K \left( \frac{x-X_j}{h_n} \right) 1_{\{Y_j > au_n^x\}} \right] \right\}.
\end{aligned}$$

The sum has variance

$$\begin{aligned} & \frac{1}{nh_n\bar{F}^x(u_n^x)} \operatorname{var} \left( \sum_{j=1}^{(n-m_n r_n)} K \left( \frac{x-X_j}{h_n} \right) 1_{\{Y_j > au_n^x\}} \right) \\ & \leq \frac{1}{nh_n\bar{F}^x(u_n^x)} \operatorname{var} \left( \sum_{j=1}^{r_n} K \left( \frac{x-X_j}{h_n} \right) 1_{\{Y_j > au_n^x\}} \right) = \frac{1}{m_n} \operatorname{var} \left( \tilde{\mathbb{V}}_n^{\dagger, x}(a) \right) \rightarrow 0, \end{aligned}$$

where we used that the joint process  $\{(X_j, Y_j)\}$  is stationary, Lemma 5 and Lemma 6. Further, by condition 2.1,

$$\frac{r_n}{\sqrt{nh_n\bar{F}^x(u_n^x)}} \mathbb{E} \left[ K \left( \frac{x-X_j}{h_n} \right) 1_{\{Y_j > au_n^x\}} \right] = \frac{r_n}{\sqrt{nh_n\bar{F}^x(u_n^x)}} O \left( h_n \bar{F}^x(u_n^x) \right) = \frac{O \left( r_n h_n \bar{F}^x(u_n^x) \right)}{\sqrt{nh_n\bar{F}^x(u_n^x)}} \rightarrow 0.$$

These two limits now immediately imply the desired result.  $\square$

**Lemma 8.** *Assume conditions 1.1-1.4 and 2.1-2.3. Then the finite dimensional limit distributions of  $\tilde{\mathbb{V}}_n^{\dagger, x}$  and  $\tilde{\mathbb{V}}_n^x$  coincide.*

**Proof.** By Lemma 7 the incomplete block is asymptotically negligible. Thus we apply Lemma 18 with  $h_{n, r_n}((X_1, Y_1), \dots, (X_{r_n}, Y_{r_n}))$  and  $f_n((X_1, Y_1), \dots, (X_{r_n}, Y_{r_n}))$  both equal to

$$\sum_{j=1}^{r_n} \frac{K \left( \frac{x-X_j}{h_n} \right) 1_{\{Y_j > su_n^x\}}}{\sqrt{nh_n\bar{F}^x(u_n^x)}}.$$

The result follows if  $\lim_{n \rightarrow \infty} m_n \beta \ell_n = 0$  and  $\lim_{n \rightarrow \infty} m_n \mathbb{E} \left[ f_n^2((X_1, Y_1), \dots, (X_{\ell_n}, Y_{\ell_n})) \right] = 0$ . The former condition holds by condition 2.3, while the latter condition follows from Lemma 6 since

$$m_n \mathbb{E} \left[ \left( \sum_{j=1}^{\ell_n} \frac{K \left( \frac{x-X_j}{h_n} \right) 1_{\{Y_j > su_n^x\}}}{\sqrt{nh_n\bar{F}^x(u_n^x)}} \right)^2 \right] \leq \frac{l_n}{r_n} \operatorname{var} \left( \tilde{\mathbb{V}}_n^{\dagger, x}(s) \right) \rightarrow 0.$$

$\square$

This concludes the finite-dimensional convergence and we now proceed to proving a functional central limit theorem.

Define the modulus of continuity over an arbitrary compact interval  $C$  as

$$w(f, \delta, C) = \sup_{\substack{s, t \in C \\ |t-s| < \delta}} |f(t) - f(s)|,$$

for  $f : C \rightarrow \mathbb{R}$ .

**Theorem 11.** *Assume conditions 1.1-1.4, 2.1-2.3. Then  $\tilde{\mathbb{V}}_n^x$  converges weakly to the centred Gaussian process  $\mathbb{V}^x$  in  $\mathbb{D}((0, \infty))$  endowed with the  $J_1$ -topology, and the process  $\mathbb{V}^x$  admits an almost surely continuous version.*

**Proof.** Theorem 10 yields that the finite dimensional distributions of  $\tilde{\mathbb{V}}_n^x$  converges to that of the proposed Gaussian process. It thus remains to prove asymptotic equicontinuity. According to Theorem 16 the statement above follows if  $\tilde{\mathbb{V}}_n^x$  is asymptotically equicontinuous w.r.t. the uniform norm on each interval  $[a, b]$  with  $0 < a < b$ , i.e if for all  $\varepsilon > 0$

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \Pr \left( w \left( \tilde{\mathbb{V}}_n^x, \delta, [a, b] \right) > \varepsilon \right) = 0.$$

To show this limit, we relate the asymptotic equicontinuity of  $\tilde{\mathbb{V}}_n^x$  to the pseudo-sample version  $\tilde{\mathbb{V}}_n^{\dagger, x}$  as follows. Define the processes

$$\begin{aligned} \tilde{\mathbb{V}}_{n,1}^x &= \sum_{\substack{1 \leq i \leq m_n \\ i \text{ odd}}} \frac{\bar{S}_{n,i}^x}{\sqrt{nh_n \bar{F}^x(u_n^x)}}, & \tilde{\mathbb{V}}_{n,2}^x &= \sum_{\substack{1 \leq i \leq m_n \\ i \text{ even}}} \frac{\bar{S}_{n,i}^x}{\sqrt{nh_n \bar{F}^x(u_n^x)}} \\ \tilde{\mathbb{V}}_{n,1}^{\dagger, x} &= \sum_{\substack{1 \leq i \leq m_n \\ i \text{ odd}}} \frac{\bar{S}_{n,i}^{\dagger, x}}{\sqrt{nh_n \bar{F}^x(u_n^x)}}, & \tilde{\mathbb{V}}_{n,2}^{\dagger, x} &= \sum_{\substack{1 \leq i \leq m_n \\ i \text{ even}}} \frac{\bar{S}_{n,i}^{\dagger, x}}{\sqrt{nh_n \bar{F}^x(u_n^x)}} \end{aligned}$$

These processes are sums of blocks separated by a distance of the block length  $r_n$  and thus Lemma 17 yields that

$$d_{\text{TV}} \left( \mathcal{L} \left( \tilde{\mathbb{V}}_{n,i}^{\dagger, x} \right), \mathcal{L} \left( \tilde{\mathbb{V}}_{n,i}^x \right) \right) \leq m_n \beta r_n, \quad i = 1, 2,$$

where  $d_{\text{TV}}$  denotes the total variation distance. The processes defined above allows for the following decompositions

$$\tilde{\mathbb{V}}_n^x = \tilde{\mathbb{V}}_{n,1}^x + \tilde{\mathbb{V}}_{n,2}^x + \tilde{\mathbb{V}}_n^{x, \text{incomplete}}, \quad \tilde{\mathbb{V}}_n^{\dagger, x} = \tilde{\mathbb{V}}_{n,1}^{\dagger, x} + \tilde{\mathbb{V}}_{n,2}^{\dagger, x},$$

where

$$\tilde{\mathbb{V}}_n^{x, \text{incomplete}}(s) := \sum_{j=m_n r_n + 1}^n \frac{K \left( \frac{x - X_j}{h_n} \right) 1_{\{Y_j > su_n^x\}} - \mathbb{E} \left[ K \left( \frac{x - X_j}{h_n} \right) 1_{\{Y_j > su_n^x\}} \right]}{\sqrt{nh_n \bar{F}^x(u_n^x)}}.$$

Let  $\varepsilon > 0$  be given and notice that  $w(f + g, \delta, [a, b]) \leq w(f, \delta, [a, b]) + w(g, \delta, [a, b])$  for any  $f, g : [a, b] \rightarrow \mathbb{R}$ . The total variation distance yields that

$$\begin{aligned} & \Pr \left( w \left( \tilde{\mathbb{V}}_n^x, \delta, [a, b] \right) > \varepsilon \right) \\ & \leq \Pr \left( w \left( \tilde{\mathbb{V}}_{n,1}^x, \delta, [a, b] \right) > \frac{\varepsilon}{3} \right) + \Pr \left( w \left( \tilde{\mathbb{V}}_{n,2}^x, \delta, [a, b] \right) > \frac{\varepsilon}{3} \right) + \Pr \left( w \left( \tilde{\mathbb{V}}_n^{x, \text{incomplete}}, \delta, [a, b] \right) > \frac{\varepsilon}{3} \right) \\ & \leq \Pr \left( w \left( \tilde{\mathbb{V}}_{n,1}^{\dagger, x}, \delta, [a, b] \right) > \frac{\varepsilon}{3} \right) + \Pr \left( w \left( \tilde{\mathbb{V}}_{n,2}^{\dagger, x}, \delta, [a, b] \right) > \frac{\varepsilon}{3} \right) + \Pr \left( w \left( \tilde{\mathbb{V}}_n^{x, \text{incomplete}}, \delta, [a, b] \right) > \frac{\varepsilon}{3} \right) \\ & \quad + 2m_n \beta r_n. \end{aligned}$$

Now, Lemma 7 implies asymptotic equicontinuity of  $\tilde{\mathbb{V}}_n^{x, \text{incomplete}}$  w.r.t. the uniform norm on compact positive intervals and hence

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \Pr \left( w \left( \tilde{\mathbb{V}}_n^{x, \text{incomplete}}, \delta, [a, b] \right) > \frac{\varepsilon}{3} \right) = 0.$$

Condition 2.3 yields  $m_n \beta_{l_n} \rightarrow 0$  and  $l_n/r_n \rightarrow 0$  which implies  $m_n \beta_{r_n} \rightarrow 0$ . Thus asymptotic equicontinuity of  $\widetilde{\mathbb{V}}_n^x$  follows if we show asymptotic equicontinuity of  $\widetilde{\mathbb{V}}_{n,1}^{\dagger,x}$  and  $\widetilde{\mathbb{V}}_{n,1}^{\dagger,x}$  or equivalently  $\widetilde{\mathbb{V}}_n^{\dagger,x}$  since these processes are sums of the same i.i.d. summands.

The summands of  $\widetilde{\mathbb{V}}_n^{\dagger,x}$  are i.i.d processes indexed on  $[a, b]$  with càdlàg sample paths, the latter implying separability of each summand. Consequently we can apply Theorem 19 to show asymptotic equicontinuity. The Lindeberg condition was already shown in Theorem 10. With  $v_n = nh_n \overline{F}^x(u_n^x)$  define the random metric  $d_n$  on  $[a, b]$  by

$$d_n^2(s, t) = \frac{1}{v_n} \sum_{i=1}^{m_n} \left\{ \overline{S}_{n,i}^{\dagger,x}(s) - \overline{S}_{n,i}^{\dagger,x}(t) \right\}^2.$$

Monotonicity of  $s \mapsto \overline{S}_{n,i}^{\dagger,x}(s)$  yields

$$\begin{aligned} \mathbb{E} [d_n^2(s, t)] &= \frac{m_n}{nh_n \overline{F}^x(u_n^x)} \mathbb{E} \left[ \left\{ \overline{S}_{n,i}^{\dagger,x}(s) - \overline{S}_{n,i}^{\dagger,x}(t) \right\}^2 \right] \\ &\leq \frac{m_n}{nh_n \overline{F}^x(u_n^x)} \left( \mathbb{E} \left[ \overline{S}_{n,i}^{\dagger,x}(s)^2 \right] - \mathbb{E} \left[ \overline{S}_{n,i}^{\dagger,x}(t)^2 \right] \right) \rightarrow \left( s^{-1/\gamma(x)} - t^{-1/\gamma(x)} \right) g(x) \int K^2(u) du, \end{aligned}$$

where Lemma 6 was invoked. Since the limit function is continuous, Dini's Theorem implies that the convergence holds uniformly on compact intervals bounded away from zero. Thus for every sequence  $\{\delta_n\}$  decreasing to zero it holds that

$$\lim_{n \rightarrow \infty} \sup_{\substack{s, t \in [a, b] \\ |s-t| < \delta_n}} \mathbb{E} [d_n^2(s, t)] \leq \lim_{n \rightarrow \infty} \sup_{\substack{s, t \in [a, b] \\ |s-t| < \delta_n}} \left\{ \frac{m_n}{nh_n \overline{F}^x(u_n^x)} \left( \mathbb{E} \left[ \overline{S}_{n,i}^{\dagger,x}(s)^2 \right] - \mathbb{E} \left[ \overline{S}_{n,i}^{\dagger,x}(t)^2 \right] \right) \right\} = 0.$$

In order to prove the random entropy integral condition, define the map  $\mathcal{E}_s$  on  $\ell_0(\mathbb{R})$  by

$$\mathcal{E}_s(\mathbf{y}, \mathbf{x}) = \sum_{j \in \mathbb{Z}} 1_{\{y_j > s\}} K(x_j), \quad (\mathbf{y}, \mathbf{x}) = ((y_j, x_j), j \in \mathbb{Z}),$$

and let  $\mathcal{G}$  be the class of functions on  $\ell_0(\mathbb{R})$  defined by  $\mathcal{G} = \{\mathcal{E}_s, a \leq s \leq b\}$ . In addition define a sequence of processes  $\{\mathbb{Z}_n\}$  indexed by  $\mathcal{G}$  by

$$\mathbb{Z}_n(\mathcal{E}_s) = v_n^{-1} \sum_{i=1}^{m_n} \mathcal{E}_s \left( (u_n^x)^{-1} \mathbf{Y}_{(i-1)r_n+1, ir_n}^{\dagger}, h_n^{-1} (\mathbf{x} - \mathbf{X}_{(i-1)r_n+1, ir_n}^{\dagger}) \right) = \frac{1}{nh_n \overline{F}^x(u_n^x)} \sum_{i=1}^{m_n} S_{n,i}^{\dagger,x}(s),$$

using the notation  $\mathbf{X}_{(i-1)r_n+1, ir_n}^{\dagger} = (X_{(i-1)r_n+1}^{\dagger}, \dots, X_{ir_n}^{\dagger})$  for all  $i = 1, \dots, m_n$ , and similarly for  $\mathbf{Y}^{\dagger}$ .

If we define a random metric  $\widetilde{d}_n$  on  $\mathcal{G}$  by

$$\begin{aligned} \widetilde{d}_n^2(\mathcal{E}_s, \mathcal{E}_t) &= v_n^{-1} \sum_{i=1}^{m_n} \left\{ \mathcal{E}_s \left( (u_n^x)^{-1} \mathbf{Y}_{(i-1)r_n+1, ir_n}^{\dagger}, h_n^{-1} (\mathbf{x} - \mathbf{X}_{(i-1)r_n+1, ir_n}^{\dagger}) \right) \right. \\ &\quad \left. - \mathcal{E}_t \left( (u_n^x)^{-1} \mathbf{Y}_{(i-1)r_n+1, ir_n}^{\dagger}, h_n^{-1} (\mathbf{x} - \mathbf{X}_{(i-1)r_n+1, ir_n}^{\dagger}) \right) \right\}^2, \end{aligned}$$

then  $\tilde{d}_n(\mathcal{E}_s, \mathcal{E}_t) = d_n(s, t)$ , and thus the random entropy condition can be proved for the class  $\mathcal{G}$  with respect to the random metric  $\tilde{d}_n$ . This reformulation allows us to invoke Lemma 20. Consider the envelope

$$\mathbf{G}(\mathbf{y}, \mathbf{x}) = \sup_{s \in [a, b]} |\mathcal{E}_s(\mathbf{y}, \mathbf{x})| = \sup_{s \in [a, b]} \left| \sum_{j \in \mathbb{Z}} 1_{\{y_j > s\}} K(x_j) \right| = \sum_{j \in \mathbb{Z}} 1_{\{y_j > a\}} K(x_j).$$

Lemma 13 implies that

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \frac{m_n}{nh_n \bar{F}^x(u_n^x)} \mathbb{E} \left[ \mathbf{G}^2 \left( (u_n^x)^{-1} \mathbf{Y}_{1, r_n}^\dagger, h_n^{-1} (\mathbf{x} - \mathbf{X}_{1, r_n}^\dagger) \right) \right] \\ &= \limsup_{n \rightarrow \infty} \frac{m_n}{nh_n \bar{F}^x(u_n^x)} \mathbb{E} \left[ \left( \sum_{j=1}^{r_n} K \left( \frac{x - X_j}{h_n} \right) 1_{\{Y_j > au_n^x\}} \right)^2 \right] \\ &= \limsup_{n \rightarrow \infty} \text{var} \left( \tilde{\mathbb{V}}_n^{\dagger, x}(a) \right) = a^{-1/\gamma(x)} g(x) \int K^2(u) du < \infty. \end{aligned}$$

We note that  $\mathcal{G}$  is linearly ordered. This indeed holds due to monotonicity:

$$s < t \iff \mathcal{E}_s(\mathbf{y}, \mathbf{x}) \geq \mathcal{E}_t(\mathbf{y}, \mathbf{x}).$$

Consequently, Lemma 21 yields that the uniform entropy condition in Lemma 20 is fulfilled, and hence the random entropy condition in Theorem 19 holds, and thus the asymptotic equicontinuity of Theorem 16 holds. This concludes the proof.  $\square$

**Proof of Proposition 2.** Consider the following decomposition

$$\tilde{\mathbb{T}}_n^x(s) = \sqrt{k_n h_n} \{g(x) - g_n(x)\} \frac{\tilde{T}_n^x(s)}{g(x)} + \frac{1}{g(x)} \tilde{\mathbb{V}}_n^x(s) + \sqrt{k_n h_n} \frac{1}{g(x)} \left\{ \mathbb{E} [\tilde{\mathbb{V}}_n^x(s)] - g(x) s^{-1/\gamma(x)} \right\}.$$

The absolute value of the last term vanishes, since it is bounded by  $\sqrt{k_n h_n} B_n^x(s_0) \rightarrow 0$ , so Lemma 16 and Theorem 11 imply that

$$\tilde{\mathbb{T}}_n^x(s) = \sqrt{k_n h_n} \{g(x) - g_n(x)\} \frac{\tilde{T}_n^x(s)}{g(x)} + \frac{1}{g(x)} \tilde{\mathbb{V}}_n^x(s) + o(1) \xrightarrow{w} \frac{1}{g(x)} \mathbb{V}^x(s).$$

$\square$

Recall the definitions  $u_n^x = F^{\leftarrow, x}(1 - k_n/n)$  and  $q_{n, k_n}^x = F_n^{\leftarrow, x}(1 - k_n/n)$ . Then  $n\bar{F}^x(u_n^x) = k_n$ . Define the conditional tail empirical quantile function  $\tilde{Q}_n^x$  on  $(0, k_n/n)$  by

$$\tilde{Q}_n^x(t) = \tilde{T}_n^{\leftarrow, x} \left( \frac{\lfloor k_n t \rfloor}{k_n} \right) = \frac{q_{\lfloor k_n t \rfloor, n}}{u_n^x},$$

and let  $Q^x(t) = t^{-\gamma(x)}$ ,  $t > 0$ .

**Theorem 12.** Assume (1), conditions 1.1-1.5 and 2.1-2.5. Then

$$\sqrt{h_n k_n} \left\{ \tilde{T}_n^x - T^x, \tilde{Q}_n^x - Q^x \right\} \xrightarrow{w} (\mathbb{T}^x, - (Q^x)' \cdot \mathbb{T}^x \circ Q^x),$$

in  $\mathbb{D}([s_0, \infty)) \times \mathbb{D}((0, s_0^{-\gamma(x)}])$  endowed with the product  $J_1$  topology. Consequently

$$\sqrt{h_n k_n} \left\{ \frac{q_{n,k_n}^x}{u_n^x} - 1 \right\} \xrightarrow{d} \gamma(x) \mathbb{T}^x(1),$$

jointly with the convergence of  $\sqrt{h_n k_n} \{ \tilde{T}_n^x - T^x \}$  to  $\mathbb{T}^x$ .

**Proof.** Note that  $[s_0, \infty) \ni s \mapsto T^x(s)$  is surjective on  $(0, s_0^{-\gamma(x)}]$ , continuous and decreasing with negative derivative. As the processes  $\tilde{T}_n^x$  have non-increasing sample paths for each  $n \geq 1$ , Vervaat's Lemma 20 yields that

$$\sqrt{h_n k_n} \left\{ \tilde{T}_n^x - T^x, \tilde{T}_n^{\leftarrow, x} - T^{\leftarrow, x} \right\} \xrightarrow{w} (\mathbb{T}^x, - (T^{\leftarrow, x})' \cdot \mathbb{T}^x \circ T^{\leftarrow, x}),$$

in  $\mathbb{D}([s_0, \infty)) \times \mathbb{D}((0, s_0^{-\gamma(x)}])$  endowed with the product  $J_1$  topology. This concludes the proof.  $\square$

**Proof of Proposition 3.** The limit follows directly from Theorem 12. The asymptotic variance follows as

$$\text{var}(\gamma(x) \mathbb{T}^x(1)) = \gamma(x)^2 \text{cov}(\mathbb{T}^x(1), \mathbb{T}^x(1)) = \frac{\gamma(x)^2}{g(x)} \int K^2(u) du.$$

$\square$

**Proof of Theorem 4.** Decompose  $\hat{\mathbb{T}}_n^x$  into the following components

$$\hat{\mathbb{T}}_n^x = \tilde{\mathbb{T}}_n^x \left( s \frac{q_{n,k_n}^x}{u_n^x} \right) + \sqrt{k_n h_n} \left\{ T^x \left( s \frac{q_{n,k_n}^x}{u_n^x} \right) - T^x(s) \right\}.$$

The continuous mapping theorem and Theorem 12 yields that the first term above converges to  $\mathbb{T}^x$ . As for the second term, Theorem 12 and the delta-method yields that

$$\begin{aligned} \sqrt{k_n h_n} \left\{ T^x \left( s \frac{q_{n,k_n}^x}{u_n^x} \right) - T^x(s) \right\} &= s^{-1/\gamma(x)} \sqrt{k_n h_n} \left\{ T^x \left( \frac{q_{n,k_n}^x}{u_n^x} \right) - T^x(1) \right\} \\ &\xrightarrow{d} s^{-1/\gamma(x)} (T^x)'(1) \cdot \gamma(x) \mathbb{T}^x(1) = -T^x(s) \mathbb{T}^x(1), \end{aligned}$$

jointly with  $\tilde{\mathbb{T}}_n^x \left( s \frac{q_{n,k_n}^x}{u_n^x} \right)$ . The argument can be transferred unchanged to pathwise weak convergence.  $\square$

### A.5. Proof of Theorem 5 (Hill CLT)

**Proof of Theorem 5.** We consider the decomposition

$$\sqrt{k_n h_n} (\widehat{\gamma}_n(x) - \gamma(x)) = \sqrt{k_n h_n} \int_{q_{n,k_n}^x/u_n^x}^1 \frac{\widetilde{T}_n^x(s)}{s} ds + \int_1^\infty \frac{\widetilde{\mathbb{T}}_n^x(s)}{s} ds.$$

We examine the first term. Since Theorem 12 yields  $q_{n,k_n}^x/u_n^x \xrightarrow{\text{Pr}} 1$ , we may assume without loss of generality that for given  $\varepsilon \in (0, 1)$ , there exists  $N \in \mathbb{N}$  such that for  $n \geq N$  it holds that  $q_{n,k_n}^x/u_n^x \in (1 - \varepsilon, 1 + \varepsilon)$ .

Thus for  $n \geq N$  we have that

$$\sqrt{k_n h_n} \left| \int_{q_{n,k_n}^x/u_n^x}^1 \frac{\widetilde{T}_n^x(s) - T^x(s)}{s} ds \right| \leq \frac{\sqrt{k_n h_n}}{1 - \varepsilon} \left| \frac{q_{n,k_n}^x}{u_n^x} - 1 \right| \sup_{t \in (1 - \varepsilon, 1 + \varepsilon)} \left| \widetilde{T}_n^x(t) - T^x(t) \right|.$$

Moreover, Theorem 12 and the continuous mapping theorem yield that

$$\sqrt{k_n h_n} \left| \frac{q_{n,k_n}^x}{u_n^x} - 1 \right| \xrightarrow{d} \gamma(x) |\mathbb{T}^x(1)|,$$

while Theorem 1 and Dini's theorem yield  $\sup_{t \in (1 - \varepsilon, 1 + \varepsilon)} \left| \widetilde{T}_n^x(t) - T^x(t) \right| \xrightarrow{\text{Pr}} 0$  and thus

$$\sqrt{k_n h_n} \int_{q_{n,k_n}^x/u_n^x}^1 \frac{\widetilde{T}_n^x(s)}{s} ds = \sqrt{k_n h_n} \int_{q_{n,k_n}^x/u_n^x}^1 \frac{T^x(s)}{s} ds + o_{\text{Pr}}(1).$$

Applying the delta-method with  $x \mapsto \int_x^1 \frac{T^x(s)}{s} ds$ , on the sequence  $(q_{n,k_n}^x/u_n^x)$  yields that

$$\sqrt{k_n h_n} \int_{q_{n,k_n}^x/u_n^x}^1 \frac{T^x(s)}{s} ds \xrightarrow{d} -\gamma(x) \mathbb{T}^x(1).$$

This provides the asymptotics of the first term. We proceed to the second term

$$\begin{aligned} \int_1^\infty \frac{\widetilde{\mathbb{T}}_n^x(s)}{s} ds &= \sqrt{k_n h_n} \frac{1}{g(x)} \{g(x) - g_n(x)\} \int_1^\infty \frac{\widetilde{T}_n^x(s)}{s} ds + \frac{1}{g(x)} \int_1^\infty \frac{\widetilde{\mathbb{V}}_n^x(s)}{s} ds \\ &\quad + \sqrt{k_n h_n} \frac{1}{g(x)} \int_1^\infty \frac{\mathbb{E}[\widetilde{\mathbb{V}}_n^x(s)] - g(x)s^{-1/\gamma(x)}}{s} ds. \end{aligned}$$

The last term is negligible due to condition 2.7. Theorem 12 yields that equation (2) holds, and consequently  $\int_1^\infty \widetilde{T}_n^x(s)/s ds \xrightarrow{\text{Pr}} \gamma(x)$  as in the proof of Theorem 3. Combined with condition 2.4 we have that the first term also vanishes. Finally we consider  $g(x)^{-1} \int_1^\infty \widetilde{\mathbb{V}}_n^x(s)/s ds$ . It holds that

$$\int_1^A \frac{\widetilde{\mathbb{V}}_n^x(s)}{s} ds \xrightarrow[n \rightarrow \infty]{d} \int_1^A \frac{\widetilde{\mathbb{V}}^x(s)}{s} ds, \quad \int_A^\infty \frac{\widetilde{\mathbb{V}}^x(s)}{s} ds \xrightarrow[A \rightarrow \infty]{d} 0.$$

Introduce

$$\mathbb{J}_{n,A} = \int_A^\infty \frac{\widetilde{\mathbb{V}}_n^x(s)}{s} ds, \quad \mathbb{J}_{n,A}^\dagger = \int_A^\infty \frac{\widetilde{\mathbb{V}}_n^{\dagger,x}(s)}{s} ds.$$

Normality is established if we can show that, for every  $\varepsilon > 0$ ,

$$\lim_{A \rightarrow \infty} \limsup_{n \rightarrow \infty} \Pr(|\mathbb{J}_{n,A}| > \varepsilon) = 0$$

By repeating the arguments in the proof of Theorem 11, it holds that

$$\Pr(|\mathbb{J}_{n,A}| > \varepsilon) \leq \Pr(|\mathbb{J}_{n,A}^\dagger| > \varepsilon) + m_n \beta r_n,$$

and thus it suffices to show that

$$\lim_{A \rightarrow \infty} \limsup_{n \rightarrow \infty} \Pr(|\mathbb{J}_{n,A}^\dagger| > \varepsilon) = 0.$$

As  $\mathbb{J}_{n,A}^\dagger$  is centred, we show negligibility of the variance. By independence of blocks we have that

$$\begin{aligned} \text{var}(\mathbb{J}_{n,A}^\dagger) &= \frac{m_n}{nh_n \overline{F}^x(u_n^x)} \text{var}\left(\sum_{j=1}^{r_n} K\left(\frac{x-X_j}{h_n}\right) \log_+\left(\frac{Y_j}{u_n^x A}\right)\right) \\ &\leq \frac{2m_n r_n}{nh_n \overline{F}^x(u_n^x)} \sum_{j=0}^{r_n} \mathbb{E}\left[K\left(\frac{x-X_0}{h_n}\right) K\left(\frac{x-X_j}{h_n}\right) \log_+\left(\frac{Y_0}{u_n^x A}\right) \log_+\left(\frac{Y_j}{u_n^x A}\right)\right] \end{aligned}$$

Now, for fixed  $m \in \mathbb{N}$ , stationarity yields

$$\begin{aligned} &\frac{m_n r_n}{nh_n \overline{F}^x(u_n^x)} \sum_{j=0}^{m-1} \mathbb{E}\left[K\left(\frac{x-X_0}{h_n}\right) K\left(\frac{x-X_j}{h_n}\right) \log_+\left(\frac{Y_0}{u_n^x A}\right) \log_+\left(\frac{Y_j}{u_n^x A}\right)\right] \\ &\leq \frac{m_n r_n}{nh_n \overline{F}^x(u_n^x)} m \mathbb{E}\left[K^2\left(\frac{x-X_0}{h_n}\right) \log_+\left(\frac{Y_0}{u_n^x A}\right)^2\right]. \end{aligned}$$

Then, recalling  $m_n r_n = O(n)$ , and that the kernel  $K$  has support  $[-1, 1]$ , we have that  $1_{\{(z-x)/h_n \leq 1\}} = 1_{\{z \leq x+h_n\}}$  and hence we have the further bound

$$\begin{aligned} &\frac{m}{nh_n \overline{F}^x(u_n^x)} \int K^2\left(\frac{x-z}{h_n}\right) \mathbb{E}\left[\log_+\left(\frac{Y_0}{u_n^x A}\right)^2 \middle| X_0 = z\right] g(z) dz \\ &\leq m \sup_{u \in [-h_n+x, x+h_n]} \left(\frac{1}{\overline{F}^u(u_n^x)} \mathbb{E}\left[\log_+\left(\frac{Y_0}{u_n^x A}\right)^2 \middle| X_0 = u\right]\right) \times \frac{1}{h_n} \int K^2\left(\frac{x-z}{h_n}\right) g(z) \frac{\overline{F}^z(u_n^x)}{\overline{F}^x(u_n^x)} dz, \end{aligned}$$

where it holds that

$$\begin{aligned} &\frac{1}{h_n} \int K^2\left(\frac{x-z}{h_n}\right) g(z) \frac{\overline{F}^z(u_n^x)}{\overline{F}^x(u_n^x)} dz \leq \frac{1}{h_n} \int K^2\left(\frac{x-z}{h_n}\right) g(z) (u_n^x)^{-(c_\gamma+c_L)|x-z|} dz \\ &\leq (u_n^x)^{(c_\gamma+c_L)h_n} \int K^2(w) g(x-wh_n) dw = O(1), \end{aligned}$$

and

$$\begin{aligned}
 & \lim_{n \rightarrow \infty} \sup_{u \in [-h_n + x, x + h_n]} \left( \frac{1}{\overline{F}^u(u_n^x)} \mathbb{E} \left[ \log_+ \left( \frac{Y_0}{u_n^x A} \right)^2 \middle| X_0 = u \right] \right) \\
 & \leq \lim_{n \rightarrow \infty} \left( \frac{1}{\overline{F}^x(u_n^x)} \mathbb{E} \left[ \log_+ \left( \frac{Y_0}{u_n^x A} \right)^2 \middle| X_0 = x \right] \right) \\
 & \quad + \lim_{n \rightarrow \infty} \sup_{u \in [-h_n + x, x + h_n]} \left| \frac{1}{\overline{F}^u(u_n^x)} \mathbb{E} \left[ \log_+ \left( \frac{Y_0}{u_n^x A} \right)^2 \middle| X_0 = u \right] - \frac{1}{\overline{F}^x(u_n^x)} \mathbb{E} \left[ \log_+ \left( \frac{Y_0}{u_n^x A} \right)^2 \middle| X_0 = x \right] \right| \\
 & = \int_A^\infty \log^2(t) 1/\gamma(x) t^{-1/\gamma(x)-1} dt,
 \end{aligned}$$

where we have used Potter's bound to invoke the dominated convergence theorem for the first limit, and the term in the difference can be bounded using Potter's bound and applying twice the Lipschitz conditions on  $L^x$  and  $1/\gamma(x)$  as was done in Theorem 3. The remaining upper sum is asymptotically negligible by condition 2.6, i.e.

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \sum_{j=m}^{r_n} \frac{\mathbb{E} \left[ K \left( \frac{x-X_0}{h_n} \right) K \left( \frac{x-X_j}{h_n} \right) \log_+ \left( \frac{Y_0}{u_n^x A} \right) \log_+ \left( \frac{Y_j}{u_n^x A} \right) \right]}{h_n \overline{F}^x(u_n^x)} = 0.$$

Thus, taking limit as  $A \rightarrow \infty$  establishes the required asymptotic normality. By construction,  $\mathbb{T}^x(\cdot) = \mathbb{V}^x(\cdot)/g(x)$ , and thus we have the representation

$$\sqrt{k_n h_n} (\widehat{\gamma}_n(x) - \gamma(x)) \xrightarrow{d} \frac{1}{g(x)} \int_1^\infty \frac{\mathbb{V}^x(s)}{s} ds - \frac{\gamma(x)}{g(x)} \mathbb{V}^x(1).$$

It remains to establish the variance of the limit distribution. We provide the straightforward calculations. Change of variable yields that the limit can be rewritten as

$$\gamma(x) \int_0^1 \frac{\mathbb{T}^x(t^{-\gamma(x)}) - t \mathbb{T}^x(1)}{t} dt.$$

Let  $\mathbb{W}^x(t) = \mathbb{T}^x(t^{-\gamma(x)})$  and note that  $\mathbb{W}^x$  has the following covariance structure

$$\text{cov}(\mathbb{W}^x(s), \mathbb{W}^x(t)) = (s \wedge t) \frac{1}{g(x)} \int K^2(u) du =: c^*(s, t).$$

The limit process is a linear functional of centred Gaussian processes, and hence centred itself. The variance of the limit variable is thus given by

$$\begin{aligned}
 & \gamma(x)^2 \mathbb{E} \left[ \int_0^1 \frac{\mathbb{T}^x(s^{-\gamma(x)}) - s \mathbb{T}^x(1)}{s} ds \int_0^1 \frac{\mathbb{T}^x(t^{-\gamma(x)}) - t \mathbb{T}^x(1)}{t} dt \right] \\
 & = \gamma(x)^2 \int_0^1 \int_0^1 \frac{c^*(s, t) - s c^*(t, 1) - t c^*(s, 1) + s t c^*(1, 1)}{s t} ds dt
 \end{aligned}$$

$$= \gamma(x)^2 \left\{ c^*(1, 1) + \int_0^1 \int_0^1 \frac{c^*(s, t)}{st} ds dt - 2 \int_0^1 \frac{c^*(s, 1)}{s} ds \right\}.$$

We have that  $c^*(s, t) = c^*(t, s) = tc^*(s/t, 1)$ , so that the latter two terms cancel and we get variance of the limiting distribution given by

$$\gamma(x)^2 c^*(1, 1) = \gamma(x)^2 \frac{1}{g(x)} \int K^2(u) du.$$

□

## Appendix B: Proofs of results from Section 4

**Proof of Theorem 6.** Due to the fact that  $\delta \in ((\eta + 1)^{-1}, 1)$ , we can let

$$e \in \left( 0, \frac{\delta}{1-\delta} \wedge \frac{\delta(\eta+1)-1}{2(1-\delta)} \right),$$

since the interval is non-empty.

Notice that by regular variation it holds that

$$\log(u_n^x) = \gamma(x) \log(\ell^x(u_n^x)) - \gamma(x) \log(k_n/n).$$

Consequently, for any strictly positive  $e$  it holds that

$$\begin{aligned} h_n \log(u_n^x) &= h_n (\gamma(x) \log(\ell^x(u_n^x)) - \gamma(x) \log(k_n/n)) \\ &= (\bar{F}^x(u_n^x))^e \left( \gamma(x) \log(\ell^x(u_n^x)) - \gamma(x) \log(n^{-(1-\delta)}) \right) \\ &= \gamma(x) (u_n^x)^{-e/\gamma(x)} \ell^x(u_n^x)^e \log(\ell^x(u_n^x)) + \mathcal{O}\left(n^{-(1-\delta)e} \log(n)\right) \\ &\sim \gamma(x) (u_n^x)^{-e/\gamma(x)} \ell^x(u_n^x)^e \log(\ell^x(u_n^x)) \\ &\leq \gamma(x) (u_n^x)^{-e/\gamma(x)} \ell^x(u_n^x)^{1+e} = \gamma(x) \left( (u_n^x)^{-(e/1+e)/\gamma(x)} \ell^x(u_n^x) \right)^{1+e} \rightarrow 0. \end{aligned}$$

The upper bound on  $e$  ensures

$$nh_n \bar{F}^x(u_n^x) \asymp n \left( \bar{F}^x(u_n^x) \right)^{1+e} \asymp n \left( \frac{k_n}{n} \right)^{1+e} = n^{-(1-\delta)(1+e)+1} = n^{-(1-\delta)e+\delta} \rightarrow \infty.$$

As for the third condition we have

$$\frac{\bar{F}^x(u_n^x) + h_n^{c_x \eta - 2}}{nh_n^{c_x} (\bar{F}^x(u_n^x))^2} = \frac{1}{nh_n^{c_x} \bar{F}^x(u_n^x)} + \frac{h_n^{c_x(\eta-1)-2}}{n(\bar{F}^x(u_n^x))^2}.$$

Let  $c_x$  be a positive constant such that

$$c_x \in \left( \frac{2e(1-\delta) + 1 - 2\delta}{e(1-\delta)(\eta-1)} \vee 0, \frac{\delta}{e(1-\delta)} \right),$$

and note that the restrictions on  $\delta$  and  $e$  ensures that this interval is non-empty. The upper bound on  $c_x$  ensures that the first of the above two terms vanishes asymptotically since

$$\left(nh_n^{c_x} \bar{F}^x(u_n^x)\right)^{-1} \asymp \left(n(\bar{F}^x(u_n^x))^{c_x e+1}\right)^{-1} = \left(n\left(\frac{k_n}{n}\right)^{c_x e+1}\right)^{-1} \asymp \left(n^{1-(1-\delta)(c_x e+1)}\right)^{-1} \rightarrow 0,$$

while the lower bound ensures

$$\begin{aligned} \frac{h_n^{c_x(\eta-1)-2}}{n(\bar{F}^x(u_n^x))^2} &= \frac{1}{nh_n \bar{F}^x(u_n^x)} \frac{h_n^{c_x(\eta-1)-1}}{\bar{F}^x(u_n^x)} \\ &\asymp n^{(1-\delta)e-\delta} \bar{F}^x(u_n^x)^{e(c_x(\eta-1)-1)-1} \asymp n^{(1-\delta)e-\delta-(1-\delta)(e(c_x(\eta-1)-1)-1)} \rightarrow 0, \end{aligned}$$

as  $n \rightarrow \infty$ . □

**Proof of Theorem 7.** Note that  $nh_n \bar{F}^x(u_n^x) = k_n h_n \asymp \log(n)^{2-e} \rightarrow \infty$ . Now, the relation

$$\log(n/k_n) = -\log\left(\bar{F}^x(u_n^x)\right) = (1/\gamma(x)) \log(u_n^x) - \log(\ell^x(u_n^x)),$$

implies that

$$\frac{\gamma(x) \log(n/k_n)}{\log(u_n^x)} = 1 - \frac{\gamma(x) \log(\ell^x(u_n^x))}{\log(u_n^x)} \rightarrow 1,$$

using [Bingham, Goldie and Teugels \(1987\)](#) Proposition 1.3.6 (i). Hence  $\log(n/k_n) \sim (1/\gamma(x)) \log(u_n^x)$  and consequently

$$h_n \log(u_n^x) \sim \gamma(x) h_n \log(n/k_n) \leq \gamma(x) h_n \log(n) \asymp \gamma(x) \log(n)^{1-e} \rightarrow 0.$$

□

**Proof of Lemma 1.** By definition of  $u_n^x$  and properties of generalised inverses, it holds that  $n\bar{F}^x(u_n^x) = k_n$ . The second order assumption of equation (4) implies that  $u_n^x = O((n/k_n)^{\gamma(x)})$ .

Then recall that

$$B_n^x(s_0) = \sup_{t \geq s_0} \left| \mathbb{E} [\tilde{V}_n^x(t)] - t^{-1/\gamma(x)} g(x) \right|.$$

The above difference can be split as follows

$$\begin{aligned} \mathbb{E} [\tilde{V}_n^x(t)] - t^{-1/\gamma(x)} g(x) &= \int K(u) \left( \frac{\bar{F}^{x-uh_n}(tu_n^x)}{\bar{F}^x(u_n^x)} g(x-h_n u) - t^{-1/\gamma(x)} g(x) \right) du \\ &= \int K(u) t^{-1/\gamma(x)} (g(x-h_n u) - g(x)) du + \int K(u) \left( \frac{\bar{F}^x(tu_n^x)}{\bar{F}^x(u_n^x)} - t^{-1/\gamma(x)} \right) g(x-h_n u) du \\ &\quad + \int K(u) \left( \frac{\bar{F}^{x-uh_n}(tu_n^x)}{\bar{F}^x(u_n^x)} - \frac{\bar{F}^x(tu_n^x)}{\bar{F}^x(u_n^x)} \right) g(x-h_n u) du. \end{aligned}$$

The above expression is analysed term by term. The density  $g$  is Lipschitz by assumption and hence the first of the above three terms is of order

$$\int K(u) t^{-1/\gamma(x)} (g(x - h_n u) - g(x)) du \leq t^{-1/\gamma(x)} h_n c_g \int K(u) |u| du = O(t^{-1/\gamma(x)} h_n).$$

When examining the third term, we rely on calculations from the proof of Lemma 4 and utilize the inequality  $|e^z - 1| \leq |z| + (1 + e^{|z|})z^2/2$ . This yields

$$\begin{aligned} & \left| \frac{\bar{F}^{x-uh_n}(tu_n^x)}{\bar{F}^x(u_n^x)} - \frac{\bar{F}^x(tu_n^x)}{\bar{F}^x(u_n^x)} \right| \leq \frac{\bar{F}^x(tu_n^x)}{\bar{F}^x(u_n^x)} \left| \frac{\bar{F}^{x-uh_n}(tu_n^x)}{\bar{F}^x(tu_n^x)} - 1 \right| \\ & \leq \frac{\bar{F}^x(tu_n^x)}{\bar{F}^x(u_n^x)} \left( \left| \log \left( \frac{\bar{F}^{x-uh_n}(tu_n^x)}{\bar{F}^x(tu_n^x)} \right) \right| + \frac{1}{2} \log \left( \frac{\bar{F}^{x-uh_n}(tu_n^x)}{\bar{F}^x(tu_n^x)} \right)^2 \left( 1 + \frac{\bar{F}^{x-uh_n}(tu_n^x)}{\bar{F}^x(tu_n^x)} \right) \right) \\ & \leq \text{cst} \frac{\bar{F}^x(tu_n^x)}{\bar{F}^x(u_n^x)} \log(tu_n^x) |u| h_n \left( 1 + \frac{1}{2} \log(tu_n^x) |u| h_n \left( 1 + (tu_n^x)^{(c_\gamma+c_L)|u|h_n} \right) \right) \\ & \leq \text{cst} t^{-1/\gamma(x)+\varepsilon} \log(tu_n^x) |u| h_n \left( 1 + \frac{1}{2} \log(tu_n^x) |u| h_n \left( 1 + (tu_n^x)^{(c_\gamma+c_L)|u|h_n} \right) \right), \end{aligned}$$

for  $\varepsilon \in (0, 1/\gamma(x))$ . The first term is the leading term and of order

$$O\left(t^{-1/\gamma(x)+\varepsilon} \log(tu_n^x) h_n\right) = O\left(t^{-1/\gamma(x)+\varepsilon} \log\left(t \frac{n}{k_n}\right) h_n\right).$$

To determine the order at which the second and final term converge to zero the second order condition is needed. Indeed

$$\begin{aligned} & \left| \frac{\bar{F}^x(tu_n^x)}{\bar{F}^x(u_n^x)} - t^{-1/\gamma(x)} \right| = \left| \frac{n\bar{F}^x(tu_n^x)}{k_n} - t^{-1/\gamma(x)} \right| \\ & \leq \frac{n}{k_n} \left( \left| \bar{F}^x(tu_n^x) - c_x(u_n^x t)^{-1/\gamma(x)} \right| + t^{-1/\gamma(x)} \left| c_x(u_n^x)^{-1/\gamma(x)} - \bar{F}^x(u_n^x) \right| \right) \\ & \leq \text{cst} \frac{n}{(u_n^x)^{\zeta_x} k_n} \left( t^{-\zeta_x} + t^{-1/\gamma(x)} \right) \leq \text{cst} \frac{n}{(u_n^x)^{\zeta_x} k_n} t^{-1/\gamma(x)} = O\left(k_n^{\zeta_x \gamma(x)-1} n^{1-\zeta_x \gamma(x)} t^{-1/\gamma(x)}\right). \end{aligned}$$

The above three O-terms are uniformly bounded in  $t$  for  $t \geq s_0 \in (0, 1)$ . Thus, we get

$$\begin{aligned} \sqrt{k_n h_n} B_n^x(s_0) & \leq O(\sqrt{k_n h_n} h_n) + O\left(\sqrt{k_n h_n} \log(n/k_n) h_n\right) + O\left(\sqrt{k_n h_n} k_n^{\zeta_x \gamma(x)-1} n^{1-\zeta_x \gamma(x)}\right) \\ & = O\left(\sqrt{k_n h_n^3} \log(n/k_n)\right) + O\left(\sqrt{h_n} k_n^{\zeta_x \gamma(x)-\frac{1}{2}} n^{1-\zeta_x \gamma(x)}\right) \rightarrow 0. \end{aligned}$$

This concludes condition 2.5. As for condition 2.7 we have

$$\begin{aligned} & \int_1^\infty \frac{\mathbb{E}[\tilde{V}_n^x(t)] - t^{-1/\gamma(x)} g(x)}{t} dt \\ & \leq O(h_n) \int_1^\infty t^{-1/\gamma(x)-1} dt + O\left(k_n^{\zeta_x \gamma(x)-1} n^{1-\zeta_x \gamma(x)}\right) \int_1^\infty t^{-1/\gamma(x)-1} dt \end{aligned}$$

$$+ O(h_n) \int_1^\infty t^{-1/\gamma(x)+\varepsilon-1} \log(tu_n^x) \left( 1 + \frac{1}{2} \log(tu_n^x) h_n \left( 1 + (tu_n^x)^{(c_\gamma+c_L)|u|/h_n} \right) \right) dt,$$

for  $\varepsilon \in (0, 1/\gamma(x))$ . The first two integrals are obviously finite, and thus we examine the last integral, which itself consists of three sub-integrals. The first of these subintegrals is finite, and of order  $O(h_n \log(u_n^x))$ , likewise for the second which is of order  $O(h_n^2 \log(u_n^x)^2)$ . Finally, let  $\varepsilon' \in (0, 1/\gamma(x) - \varepsilon)$  be given. It is possible to find  $n'$  such that  $(c_\gamma + c_L) h_n \leq \varepsilon'$  for all  $n \geq n'$ . For such  $n$ 's the third integral is also finite and of order  $O(h_n \log(u_n^x))$ , since  $(u_n^x)^{h_n} = O(1)$ .

Consequently

$$\sqrt{k_n h_n} \int_1^\infty \frac{\mathbb{E} [\tilde{V}_n^x(t)] - t^{-1/\gamma(x)} g(x)}{t} dt \leq O\left(\sqrt{k_n h_n^3} \log(n/k_n)\right) + O\left(\sqrt{h_n} k_n^{\zeta_x \gamma(x) - \frac{1}{2}} n^{1 - \zeta_x \gamma(x)}\right),$$

which vanishes as  $n \rightarrow \infty$ .  $\square$

**Proof of Lemma 2.** As above, it holds that  $\bar{F}^x(tu_n^x)/\bar{F}^x(u_n^x) = \bar{F}^x(tu_n^x n/k_n)$ . Consequently

$$\begin{aligned} & \left| \frac{\bar{F}^x(tu_n^x)}{\bar{F}^x(u_n^x)} - t^{-1/\gamma(x)} \right| \\ &= \frac{n}{k_n} \left| \bar{F}^x(tu_n^x) - c_x \log(u_n^x t) (u_n^x t)^{-1/\gamma(x)} + c_x \log(u_n^x t) (u_n^x t)^{-1/\gamma(x)} - \bar{F}^x(u_n^x) t^{-1/\gamma(x)} \right| \\ &\leq \frac{n}{k_n} \left( \left| \bar{F}^x(tu_n^x) - c_x \log(u_n^x t) (u_n^x t)^{-1/\gamma(x)} \right| \right. \\ &\quad \left. + t^{-1/\gamma(x)} \left| c_x \log(u_n^x) (u_n^x)^{-1/\gamma(x)} - \bar{F}^x(u_n^x) \right| + c_x \log(t) (u_n^x t)^{-1/\gamma(x)} \right) \\ &\leq \text{cst} \frac{n}{k_n} \left( (tu_n^x)^{-\zeta_x} + t^{-1/\gamma(x)} (u_n^x)^{-\zeta_x} + c_x \log(t) (u_n^x t)^{-1/\gamma(x)} \right) \\ &\leq \text{cst} \frac{n}{k_n} t^{-1/\gamma(x)} (u_n^x)^{-1/\gamma(x)} (1 + \log(t)). \end{aligned}$$

Equation (6) yields that

$$(u_n^x)^{-1/\gamma(x)} \log(u_n^x) = \frac{k_n}{n} + O\left((u_n^x)^{-\zeta_x}\right),$$

and consequently

$$\begin{aligned} & \sqrt{k_n h_n} \left| \frac{\bar{F}^x(tu_n^x)}{\bar{F}^x(u_n^x)} - t^{-1/\gamma(x)} \right| \leq \text{cst} t^{-1/\gamma(x)} (1 + \log(t)) \sqrt{k_n h_n} \frac{n}{k_n} \frac{(u_n^x)^{-1/\gamma(x)} \log(u_n^x)}{\log(u_n^x)} \\ &= \text{cst} t^{-1/\gamma(x)} (1 + \log(t)) \sqrt{k_n h_n} \frac{n}{k_n} \frac{\frac{k_n}{n} + O\left((u_n^x)^{-\zeta_x}\right)}{\log(u_n^x)} \\ &= \text{cst} t^{-1/\gamma(x)} (1 + \log(t)) \left( \frac{\sqrt{k_n h_n}}{\log(u_n^x)} + \frac{O\left(n \sqrt{h_n} (u_n^x)^{-\zeta_x}\right)}{\log(u_n^x) \sqrt{k_n}} \right). \end{aligned}$$

The mapping  $t \mapsto t^{-1/\gamma(x)-1} (1 + \log(t))$  is integrable on  $(1, \infty)$  and hence it remains to examine the deterministic sequence

$$\left( \frac{\sqrt{k_n h_n}}{\log(u_n^x)} + \frac{O(n\sqrt{h_n}(u_n^x)^{-\zeta_x})}{\log(u_n^x)\sqrt{k_n}} \right) = \frac{\sqrt{k_n h_n}}{\log(u_n^x)} \left( 1 + \frac{n}{k_n(u_n^x)^{\zeta_x}} \right).$$

Notice first that, since  $\zeta_x \gamma(x) > 1$ ,

$$\frac{n}{k_n(u_n^x)^{\zeta_x}} = \frac{1}{\bar{F}^x(u_n^x)(u_n^x)^{\zeta_x}} \rightarrow 0,$$

As argued in the proof of Theorem 7 we have  $\log(n/k_n) \sim (1/\gamma(x)) \log(u_n^x)$  and thus

$$\frac{\sqrt{k_n h_n}}{\log(u_n^x)} \left( 1 + \frac{n}{k_n(u_n^x)^{\zeta_x}} \right) \sim \frac{\sqrt{k_n h_n}}{\gamma(x) \log(n/k_n)} \left( 1 + \frac{n}{k_n(u_n^x)^{\zeta_x}} \right) \rightarrow 0.$$

To verify that condition 2.5 holds, we can repeat some of the arguments from Lemma 1:

$$\begin{aligned} \sqrt{k_n h_n} B_n^x(s_0) &\leq O(\sqrt{k_n h_n} h_n) + O\left(\sqrt{k_n h_n} \log(u_n^x) h_n\right) + O\left(\frac{\sqrt{k_n h_n}}{\log(n/k_n)}\right) \\ &= O\left(\sqrt{k_n h_n^3} \log(n/k_n)\right) + O\left(\frac{\sqrt{k_n h_n}}{\log(n/k_n)}\right) \rightarrow 0. \end{aligned}$$

In a similar way, arguments akin to those of Lemma 1 also yield that condition 2.7 holds.  $\square$

**Proof of Lemma 3.** Consider the expression of condition 2.2:

$$\begin{aligned} &\sum_{j=m}^{r_n} \frac{\mathbb{E} \left[ K\left(\frac{x-X_0}{h_n}\right) 1_{\{Y_0 > su_n^x\}} K\left(\frac{x-X_j}{h_n}\right) 1_{\{Y_j > tu_n^x\}} \right]}{h_n \bar{F}^x(u_n^x)} \\ &= \sum_{j=m}^{r_n} \frac{\text{cov} \left( K\left(\frac{x-X_0}{h_n}\right) 1_{\{Y_0 > su_n^x\}} K\left(\frac{x-X_j}{h_n}\right) 1_{\{Y_j > tu_n^x\}} \right)}{h_n \bar{F}^x(u_n^x)} \\ &\quad + (r_n - m) h_n \bar{F}^x(u_n^x) \frac{\mathbb{E} \left[ K\left(\frac{x-X_0}{h_n}\right) 1_{\{Y_0 > su_n^x\}} \right]}{h_n \bar{F}^x(u_n^x)} \frac{\mathbb{E} \left[ K\left(\frac{x-X_0}{h_n}\right) 1_{\{Y_0 > tu_n^x\}} \right]}{h_n \bar{F}^x(u_n^x)}, \end{aligned}$$

where the latter term is negligible due to condition 2.1. Calculations similar to those of the proof of Theorem 2 yield

$$\begin{aligned} \sum_{j=m}^{r_n} \frac{\text{cov} \left( K\left(\frac{x-X_0}{h_n}\right) 1_{\{Y_0 > su_n^x\}} K\left(\frac{x-X_j}{h_n}\right) 1_{\{Y_j > tu_n^x\}} \right)}{h_n \bar{F}^x(u_n^x)} &\leq \frac{\text{cst}}{h_n \bar{F}^x(u_n^x)} \left( h_n^{2-c_x} \bar{F}^x(u_n^x) + h_n^{c_x(\eta-1)} \right) \\ &= \text{cst} (h_n^{1-c_x} + h_n^{c_x(\eta-1)-1} \bar{F}^x(u_n^x)^{-1}) \rightarrow 0, \end{aligned}$$

as  $n \rightarrow \infty$ , for some constant  $c_x > 0$ , by assumption. This yields condition 2.2.

We proceed to establish the logarithmic anti-clustering condition 2.6, focusing, as above, on the covariance rather than the cross-term. Let  $q > 2$  and  $p > 1$  such that  $2/q = 1 - 1/p$ . Rio's covariance

inequality and calculations similar to those in the proof of Theorem 5 yield

$$\begin{aligned} & \left| \text{cov} \left( K \left( \frac{x - X_0}{h_n} \right) \log_+ \left( \frac{Y_0}{u_n^x A} \right), K \left( \frac{x - X_j}{h_n} \right) \log_+ \left( \frac{Y_j}{u_n^x A} \right) \right) \right| \\ & \leq 2p(2\alpha(j))^{1/p} \mathbb{E} \left[ K^q \left( \frac{x - X_0}{h_n} \right) \log_+^q \left( \frac{Y_0}{u_n^x A} \right) \right]^{2/q} = O(\alpha(j)^{1/p} h_n^{2/q} \bar{F}^x(u_n^x)^{2/q}). \end{aligned}$$

In addition it holds that

$$\begin{aligned} & \mathbb{E} \left[ K \left( \frac{x - X_0}{h_n} \right) K \left( \frac{x - X_j}{h_n} \right) \log_+ \left( \frac{Y_0}{u_n^x A} \right) \log_+ \left( \frac{Y_j}{u_n^x A} \right) \right] \\ & = h_n^2 \bar{F}^x(u_n^x) \int_{[-1,1]^2} \left\{ \frac{\mathbb{E} \left[ \log_+ \left( \frac{Y_0}{u_n^x A} \right) \log_+ \left( \frac{Y_j}{u_n^x A} \right) \mid X_0 = x + h_n w_1, X_j = x + h_n w_2 \right]}{\bar{F}^x(u_n^x)} \right\} \\ & \quad \times g(x + w_1 h_n, x + w_2 h_n) K(w_1) K(w_2) \Big\} d(w_1, w_2) = O \left( h_n^2 \bar{F}^x(u_n^x) \right), \end{aligned}$$

due to (8). The above two inequalities give

$$\begin{aligned} & \frac{1}{h_n \bar{F}^x(u_n^x)} \left| \sum_{j=m}^{r_n} \text{cov} \left( K \left( \frac{x - X_0}{h_n} \right) \log_+ \left( \frac{Y_0}{u_n^x A} \right), K \left( \frac{x - X_j}{h_n} \right) \log_+ \left( \frac{Y_j}{u_n^x A} \right) \right) \right| \\ & \leq \frac{\text{cst}}{h_n \bar{F}^x(u_n^x)} \left( \sum_{j=1}^{d_n-1} h_n^2 \bar{F}^x(u_n^x) + \sum_{j=d_n}^{\infty} h_n^{2/q} \bar{F}^x(u_n^x)^{2/q} j^{-\eta/p} \right) \\ & \leq \text{cst} \left( d_n h_n + d_n^{-\eta/p+1} h_n^{2/q-1} \bar{F}^x(u_n^x)^{2/q-1} \right) \sim \text{cst} \left( h_n^{1-c'_x} + h_n^{-1/p+c'_x(\eta/p-1)} \bar{F}^x(u_n^x)^{-1/p} \right) \rightarrow 0, \end{aligned}$$

where as above, we have  $d_n = \lceil h_n^{-c'_x} \rceil$  for some  $c'_x > 0$ .  $\square$

**Proof of Theorem 8.** Theorem 6 ensures that the first two limit conditions for consistency are fulfilled for  $k_n \asymp n^\delta$  for  $\delta \in (0, 3\nu/(2(1+\nu)))$  and  $h_n \asymp \bar{F}^x(u_n^x)^e$  for  $e \in (\delta/(3(1-\delta)), \delta/(1-\delta))$ . Let  $r_n \asymp n^a$ ,  $a = 1 - 2\delta/3$ . Then  $r_n/n \rightarrow 0$  and

$$r_n h_n \bar{F}^x(u_n^x) \asymp n^a \left( \frac{k_n}{n} \right)^{1+e} \asymp n^{a-(1-\delta)(1+e)} \rightarrow 0,$$

since  $(1-\delta)(1+e) > 1 - 2\delta/3$ . Let  $\ell_n \asymp n^\psi$ ,  $\psi \in (0, a)$ . Then  $1/\ell_n \rightarrow 0$ ,  $\ell_n/r_n \rightarrow 0$  and

$$\frac{n}{r_n} \beta_{\ell_n} \leq \text{cst} n^{1-a-\nu\psi} \rightarrow 0,$$

since upper bound on  $\delta$  ensures that for every  $\nu > 0$  there exists a small  $\varepsilon > 0$  such that for  $\psi = a - \varepsilon$ , it holds that  $1 - a - \nu\psi < 0$ . This concludes the second limit condition. Lemma 16 yields that the third condition holds if  $\lim_{n \rightarrow \infty} \sqrt{\bar{F}^x(u_n^x) n h_n h_n^2} = 0$ . This is indeed true since

$$\sqrt{\bar{F}^x(u_n^x) n h_n h_n^2} \asymp \sqrt{n \left( \frac{k_n}{n} \right)^{1+5e}} \asymp \sqrt{n^{1-(1+5e)(1-\delta)}},$$

which vanishes asymptotically for  $\delta/(5(1-\delta)) < \delta/(3(1-\delta)) < e$ . According to Lemma 1 it suffices to check the limits of equation (5) in order ensure the last two of the above five conditions. The lower bound on  $e$  ensures

$$\sqrt{k_n h_n^3} \log(n/k_n) \asymp \sqrt{n^{\delta-3e(1-\delta)}} \log(n^{1-\delta}) \rightarrow 0.$$

Further restricting, let

$$\delta \in \left(0, \frac{3}{2} \frac{\nu}{1+\nu} \wedge \frac{2(\zeta_x - 1/\gamma(x))}{2\zeta_x - 1/\gamma(x)}\right),$$

where the latter fraction of the upper bound of  $\delta$  ensures that  $e$  is in a non-empty interval, since  $\zeta_x \gamma(x) > 1$ , by assumption. It then holds that

$$\sqrt{h_n k_n^{\zeta_x \gamma(x) - \frac{1}{2}}} n^{1-\zeta_x \gamma(x)} \asymp \sqrt{h_n n^{\delta(\zeta_x \gamma(x) - \frac{1}{2}) + 1 - \zeta_x \gamma(x)}} \rightarrow 0,$$

which concludes the last two of the above five conditions.

According to Lemma 3, the anti-clustering conditions follow if equation (9) holds. The first limit expression of equation (9) is given by

$$h_n^{1-c_x} + h_n^{c_x(\eta-1)-1} \bar{F}^x(u_n^x)^{-1} \asymp \bar{F}^x(u_n^x)^{e(1-c_x)} + \bar{F}^x(u_n^x)^{ec_x(\eta-1)+(e+1)}.$$

Since  $\eta > 2$ , there exists  $c_x \in (0, 1)$ , such that the exponents above are positive, and consequently that the expression converges to zero as  $n \rightarrow \infty$ . For the second limit expression of equation (9), note that

$$h_n^{1-c'_x} + h_n^{-1/p+c'_x(\eta/p-1)} \bar{F}^x(u_n^x)^{-1/p} \asymp \bar{F}^x(u_n^x)^{e(1-c'_x)} + \bar{F}^x(u_n^x)^{ec'_x(\eta/p-1)+(e+1)/p}.$$

For  $c'_x \in (0, 1)$  and  $p \in (1, (1+e)/(ec'_x) + \eta)$  the exponents above are positive, and consequently the expression converges to zero as  $n \rightarrow \infty$ .  $\square$

**Proof of Theorem 9.** Let  $h_n \asymp \log(n)^{-e}$  for  $e \in (4/3, 2)$  and let  $k_n \asymp \log(n)^2$ . Then, according to Theorem 7, the first two conditions for consistency is satisfied. The proof is similar to that of Theorem 8, thus we only show that the statements hold, where adjustment is needed. It holds that  $k_n/n \asymp \log(n)^2/n \rightarrow 0$  and hence  $\lim_{n \rightarrow \infty} u_n^x = \infty$ . Secondly

$$nh_n \bar{F}^x(u_n^x) = k_n h_n \asymp \log(n)^{2-e} \rightarrow \infty.$$

As argued in the proof of Theorem 7, we have  $\log(n/k_n) \sim (1/\gamma(x)) \log(u_n^x)$  and thus

$$h_n \log(u_n^x) \sim \gamma(x) h_n \log(n/k_n) \leq \text{cst} \log(n)^{1-e} \rightarrow 0.$$

This concludes the three consistency conditions. Let  $r_n \asymp n^a$ ,  $a \in (0, 1)$ . Then  $r_n/n \rightarrow 0$  and

$$r_n h_n \bar{F}^x(u_n^x) \asymp n^{a-1} \log(n)^{2-e} \rightarrow 0.$$

Let  $\ell_n \asymp n^\psi$ ,  $\psi \in (0, a)$ . Then  $1/\ell_n \rightarrow 0$ ,  $\ell_n/r_n \rightarrow 0$  and

$$\frac{n}{r_n} \beta \ell_n \leq \text{cst} n^{1-a-\nu\psi} \rightarrow 0,$$

since for every  $\nu > 0$  there exists  $a \in (0, 1)$  and  $\psi \in (0, a)$  such that  $1 - a - \nu\psi < 0$ . This concludes the second normality limit condition above. Note also that

$$\sqrt{\bar{F}^x(u_n^x)nh_n h_n^2} \asymp \sqrt{k_n h_n h_n^2} \asymp \log(n)^{1-5e/2} \rightarrow 0,$$

which, according to Lemma 16, implies the third condition. The last two conditions for asymptotic normality are fulfilled by Lemma 2, since

$$\frac{\sqrt{k_n h_n}}{\log(n/k_n)} \sim \frac{\sqrt{k_n h_n}}{\log(n)} \asymp \frac{\sqrt{\log(n)^{2-e}}}{\log(n)} = \log(n)^{-e/2} \rightarrow 0,$$

and

$$\sqrt{k_n h_n h_n} \log(u_n^x) \sim \gamma(x) \sqrt{k_n h_n h_n} \log(n/k_n) \leq \text{cst} \sqrt{k_n h_n h_n} \log(n) \asymp \log(n)^{2-3e/2} \rightarrow 0.$$

which concludes the last two of the above five normality conditions.  $\square$

## Appendix C: Anti-clustering conditions for simulations

When checking that the time series in the simulation study satisfy the required conditions, Lemma 3 can be used to verify anti-clustering conditions 2.2 and 2.6. In this appendix we prove that equation (8) holds for the proposed conditional Fréchet and Pareto models, and also provide a result that can be used to check the anti-clustering conditions for the proposed conditional max-stable process.

### C.1. Conditional Fréchet and Pareto

Consider covariate process  $\{X_j\}$ , belonging to some stationary (possibly long-memory) time series model and  $\{U_j\}$  stationary (possibly long-memory) with marginals in  $[0, 1]$ . The constructions

$$Y_j = (-\log(U_j))^{-\gamma(X_j)}, \quad j \in \mathbb{Z}, \quad (14)$$

$$Y_j = U_j^{-\gamma(X_j)}, \quad j \in \mathbb{Z}, \quad (15)$$

gives rise to a time series models with marginal Fréchet  $(1/\gamma(X_j), 1, 0)$  and Pareto  $(1, 1/\gamma(X_j))$  distributions respectively for every  $j$ .

**Lemma 9.** *Assume conditions 1.1-1.4, with either (14) or (15). Then condition 1.6 and (8) holds.*

**Proof.** Let  $s, t > 0$  be given. In both equation (14) and (15), the Cauchy-Schwarz inequality yields that

$$\begin{aligned} & \mathbb{E} \left[ \mathbb{1}_{\{Y_0 > su_n^x\}} \mathbb{1}_{\{Y_j > tu_n^x\}} \mid X_0 = x + h_n w_1, X_j = x + h_n w_2 \right] / \bar{F}^x(u_n^x) \\ & \leq \frac{\mathbb{E} \left[ \mathbb{1}_{\{Y_0 > su_n^x\}} \mid X_0 = x + h_n w_1, X_j = x + h_n w_2 \right]^{1/2} \mathbb{E} \left[ \mathbb{1}_{\{Y_j > tu_n^x\}} \mid X_0 = x + h_n w_1, X_j = x + h_n w_2 \right]^{1/2}}{\bar{F}^x(u_n^x)} \\ & = \left( \frac{\mathbb{E} \left[ \mathbb{1}_{\{Y_0 > su_n^x\}} \mid X_0 = x + h_n w_1 \right]}{\bar{F}^x(u_n^x)} \right)^{1/2} \left( \frac{\mathbb{E} \left[ \mathbb{1}_{\{Y_j > tu_n^x\}} \mid X_j = x + h_n w_2 \right]}{\bar{F}^x(u_n^x)} \right)^{1/2} = O(1), \end{aligned}$$

uniformly in  $w_1, w_2 \in [-1, 1]$ , where we have used Potter's bound and the Lipschitz conditions on  $L^x$  and  $1/\gamma(x)$  as was done in Lemma 4. This shows condition 1.6.

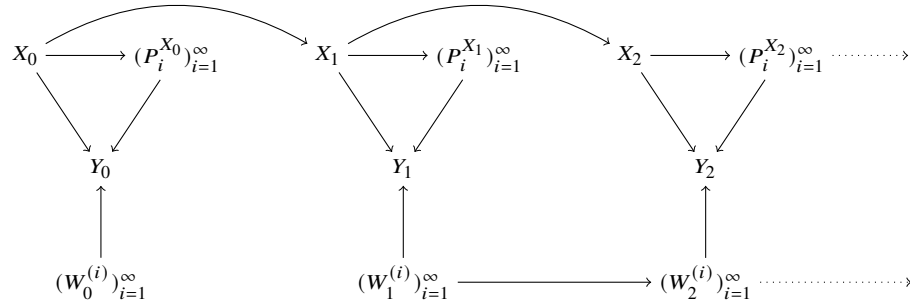
Showing equation (8) is done analogously

$$\begin{aligned} & \frac{\mathbb{E} \left[ \log_+ \left( \frac{Y_0}{u_n^x A} \right) \log_+ \left( \frac{Y_j}{u_n^x A} \right) \mid X_0 = x + h_n w_1, X_j = x + h_n w_2 \right]}{\bar{F}^x(u_n^x)} \\ & \leq \left( \frac{\mathbb{E} \left[ \log_+^2 \left( \frac{Y_0}{u_n^x A} \right) \mid X_0 = x + h_n w_1 \right]}{\bar{F}^x(u_n^x)} \right)^{1/2} \left( \frac{\mathbb{E} \left[ \log_+^2 \left( \frac{Y_0}{u_n^x A} \right) \mid X_0 = x + h_n w_2 \right]}{\bar{F}^x(u_n^x)} \right)^{1/2} \\ & \rightarrow \int_A^\infty \log^2(t) (1/\gamma(x)) t^{-1/\gamma(x)-1} dt, \end{aligned}$$

uniformly in  $w_1, w_2 \in [-1, 1]$  as shown in the proof of Theorem 5. We conclude that also equation (8) holds.  $\square$

## C.2. Conditional Max-Stable process

The construction of equation (12), can be summarised in the following diagram, when the centred Gaussian processes are started at zero and  $(P^{X_i} \perp P^{X_j})|(X_i, X_j)$ . Since  $(W_0^{(i)})_{i=1}^\infty$  is an infinite dimen-



**Figure 7.** Diagram of the conditional Max-stable (CSGMS) construction. Since the Gaussian processes  $\mathbf{W}^{(i)}, i \geq 1$  are all started in zero, the distribution of the processes at time 1, namely  $(W_1^{(i)})_{i=1}^\infty$ , doesn't depend on the starting values  $(W_0^{(i)})_{i=1}^\infty$  as they are deterministic. Hence no arrow between  $(W_0^{(i)})_{i=1}^\infty$  and  $(W_1^{(i)})_{i=1}^\infty$ .

sional vector of zeroes, it holds that  $(Y_0 \perp Y_j)|(X_0, X_j)$  for every  $j$  and  $Y_j|X_j, X_i \sim Y_j|X_j$  for  $i \neq j$ . For time series in our setup, satisfying these distributional properties, it turns out that showing the anti-clustering conditions can be done without the use of Lemma 3. This is content of the Lemma below,

**Lemma 10.** *Assume (1), conditions 1.1 - 1.4 and 2.1. If  $(Y_0 \perp Y_j)|(X_0, X_j)$  for every  $j$ , and  $Y_j|X_j, X_i \sim Y_j|X_j$  for  $i \neq j$ , then conditions 2.2 and 2.6 hold.*

**Proof.** Consider the numerator in condition 2.2

$$\begin{aligned}
& \mathbb{E} \left[ K \left( \frac{x - X_0}{h_n} \right) 1_{\{Y_0 > su_n^x\}} K \left( \frac{x - X_j}{h_n} \right) 1_{\{Y_j > tu_n^x\}} \right] \\
&= \mathbb{E} \left[ K \left( \frac{x - X_0}{h_n} \right) K \left( \frac{x - X_j}{h_n} \right) \mathbb{E} [1_{\{Y_0 > su_n^x\}} \mid X_0, X_j] \mathbb{E} [1_{\{Y_j > tu_n^x\}} \mid X_0, X_j] \right] \\
&= \mathbb{E} \left[ K \left( \frac{x - X_0}{h_n} \right) K \left( \frac{x - X_j}{h_n} \right) \mathbb{E} [1_{\{Y_0 > su_n^x\}} \mid X_0] \mathbb{E} [1_{\{Y_j > tu_n^x\}} \mid X_j] \right] \\
&= (h_n \bar{F}^x(u_n^x))^2 \int_{[-1,1]^2} \left\{ \frac{\mathbb{E} [1_{\{Y_0 > su_n^x\}} \mid X_0 = x + w_1 h_n]}{\bar{F}^x(u_n^x)} \frac{\mathbb{E} [1_{\{Y_0 > tu_n^x\}} \mid X_0 = x + w_2 h_n]}{\bar{F}^x(u_n^x)} \right. \\
&\quad \left. \times g(x + w_1 h_n, x + w_2 h_n) K(w_1) K(w_2) \right\} d(w_1, w_2) \\
&= O((h_n \bar{F}^x(u_n^x))^2),
\end{aligned}$$

where we have used Potter's bound and the Lipschitz conditions on  $L^x$  and  $1/\gamma(x)$  as was done in Lemma 4. Consequently

$$\begin{aligned}
& \lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \sum_{j=m}^{r_n} \frac{\mathbb{E} \left[ K \left( \frac{x - X_0}{h_n} \right) 1_{\{Y_0 > su_n^x\}} K \left( \frac{x - X_j}{h_n} \right) 1_{\{Y_j > tu_n^x\}} \right]}{h_n \bar{F}^x(u_n^x)} \\
&\leq \limsup_{n \rightarrow \infty} \sum_{j=1}^{r_n} \frac{\mathbb{E} \left[ K \left( \frac{x - X_0}{h_n} \right) 1_{\{Y_0 > su_n^x\}} K \left( \frac{x - X_j}{h_n} \right) 1_{\{Y_j > tu_n^x\}} \right]}{h_n \bar{F}^x(u_n^x)} \leq \limsup_{n \rightarrow \infty} O(r_n h_n \bar{F}^x(u_n^x)) = 0,
\end{aligned}$$

which concludes condition 2.2. Showing condition 2.6 is done analogously, using that

$$\begin{aligned}
& \mathbb{E} \left[ K \left( \frac{x - X_0}{h_n} \right) K \left( \frac{x - X_j}{h_n} \right) \log_+ \left( \frac{Y_0}{u_n^x A} \right) \log_+ \left( \frac{Y_j}{u_n^x A} \right) \right] \\
&= (h_n \bar{F}^x(u_n^x))^2 \int_{[-1,1]^2} \left\{ \frac{\mathbb{E} \left[ \log_+ \left( \frac{Y_0}{u_n^x A} \right) \mid X_0 = x + h_n w_1 \right]}{\bar{F}^x(u_n^x)} \frac{\mathbb{E} \left[ \log_+ \left( \frac{Y_0}{u_n^x A} \right) \mid X_0 = x + h_n w_2 \right]}{\bar{F}^x(u_n^x)} \right. \\
&\quad \left. \times g(x + w_1 h_n, x + w_2 h_n) K(w_1) K(w_2) \right\} d(w_1, w_2) \\
&= O((h_n \bar{F}^x(u_n^x))^2),
\end{aligned}$$

using calculations similar to those in the proof of Theorem 5. □

## Appendix D: Kernel density estimation for mixing sequences

The main task of this section is to prove consistency of the Parzen–Rosenblatt estimator for a stationary time series, under suitable assumptions involving either some form of anti-clustering or mixing condition. We also analyse the speed of convergence.

We begin with an auxiliary result which is used in all subsequent proofs.

**Lemma 11.** Consider a stationary time series  $\{(X_j), j \in \mathbb{Z}\}$  where each  $X_j$  takes values in  $[-1, 1]$  and admits twice continuously differentiable density  $g$ . Let the kernel  $K$  be a bounded, symmetric and defined on  $[-1, 1]$ , and let the bandwidth sequence  $\{h_n\}$  satisfy

$$\lim_{n \rightarrow \infty} h_n = 0.$$

For any  $x \in \mathbb{R}$ , it holds that

$$\frac{\mathbb{E} \left[ K^q \left( \frac{x - X_0}{h_n} \right) \right]}{h_n} \rightarrow g(x) \int K^q(u) du.$$

**Proof.** The left hand side of the limit can be rewritten as follows

$$\frac{\mathbb{E} \left[ K^q \left( \frac{x - X_0}{h_n} \right) \right]}{h_n} = \frac{1}{h_n} \int g(z) K^q \left( \frac{x - z}{h_n} \right) dz = \int K^q(u) g(x + h_n u) du.$$

The Taylor expansion  $g(x + h_n u) = g(x) + h_n u g^{(1)}(x) + h_n^2 u^2 g^{(2)}(x)/2 + O(|h_n u|^3)$  yields

$$\begin{aligned} \frac{\mathbb{E} \left[ K^q \left( \frac{x - X_0}{h_n} \right) \right]}{h_n} &= g(x) \int K^q(u) du + h_n g^{(1)}(x) \int u K^q(u) du \\ &\quad + \frac{h_n^2 g^{(2)}(x)}{2} \int u^2 K(u) du + o(h_n^2) \rightarrow g(x) \int K^q(u) du, \end{aligned}$$

as  $n \rightarrow \infty$ . □

The following Lemma provides consistency when assuming anti-clustering.

**Lemma 12.** Assume that  $h_n \rightarrow 0$ ,  $nh_n \rightarrow \infty$ , that the time series  $\{(X_j), j \in \mathbb{Z}\}$  is stationary with

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{nh_n} \sum_{j=m}^n \frac{\mathbb{E} \left[ K \left( \frac{x - X_0}{h_n} \right) K \left( \frac{x - X_j}{h_n} \right) \right]}{h_n} = 0,$$

where each  $X_j$  admits twice continuously differentiable density  $g$ , and the joint density  $g_{X_i, X_j}$  of  $(X_i, X_j)$  exists, and is bounded for all  $i, j \in \mathbb{Z}$ .

Then for any  $x \in \mathbb{R}$ , it holds that  $g_n(x) \xrightarrow{\text{Pr}} g(x)$ .

**Proof.** Fix  $x \in \mathbb{R}$ . The bias is negligible since

$$\begin{aligned} \mathbb{E} [g_n(x)] - g(x) &= \int \frac{1}{h_n} K \left( \frac{x - y}{h_n} \right) \{g(y) - g(x)\} dy \\ &= \int K(u) \{g'(x)(-h_n u) + \frac{1}{2} g''(x)(h_n u)^2 + o(h_n^2)\} du = \frac{h_n^2}{2} g''(x) \int u^2 K(u) du + o(h_n^2) \rightarrow 0, \end{aligned}$$

as  $n \rightarrow \infty$ . We consider the variance

$$\begin{aligned} \text{var}(g_n(x)) &= \frac{1}{(nh_n)^2} \text{var} \left( \sum_{i=1}^n K \left( \frac{x - X_i}{h_n} \right) \right) \\ &= \frac{1}{(nh_n)^2} \left( n \mathbb{E} \left[ K^2 \left( \frac{x - X_0}{h_n} \right) \right] - n \mathbb{E} \left[ K \left( \frac{x - X_0}{h_n} \right) \right]^2 + 2n \sum_{i=1}^n \left( 1 - \frac{i}{n} \right) \text{cov} \left( K \left( \frac{x - X_0}{h_n} \right), K \left( \frac{x - X_i}{h_n} \right) \right) \right) \\ &\leq \frac{1}{nh_n^2} \text{var} \left( K \left( \frac{x - X_0}{h_n} \right) \right) + \frac{2}{nh_n^2} \sum_{i=1}^n \left| \text{cov} \left( K \left( \frac{x - X_0}{h_n} \right), K \left( \frac{x - X_i}{h_n} \right) \right) \right|, \end{aligned}$$

where we used stationarity. The first term is negligible by Lemma 11, since  $nh_n \rightarrow \infty$ . Since the joint density  $g_{X_0, X_i}$  is bounded for all  $i \in \mathbb{N}$  it holds that

$$\begin{aligned} &\mathbb{E} \left[ K \left( \frac{x - X_0}{h_n} \right) K \left( \frac{x - X_i}{h_n} \right) \right] \\ &= h_n^2 \int \int K(u) K(y) g_{X_0, X_i}(x - uh_n, x - yh_n) du dy \leq h_n^2 \sup_{x_1, x_2 \in \mathbb{R}} (g_{X_0, X_i}(x_1, x_2)) = O(h_n^2). \end{aligned}$$

Similarly  $\mathbb{E} [K^2((x - X_0)/h_n)] = O(h_n^2)$ , and consequently, for fixed  $m \in \mathbb{N}$  we have

$$\limsup_{n \rightarrow \infty} \frac{2}{nh_n^2} \sum_{i=1}^m \left| \text{cov} \left( K \left( \frac{x - X_0}{h_n} \right), K \left( \frac{x - X_i}{h_n} \right) \right) \right| \leq \text{cst} \limsup_{n \rightarrow \infty} \frac{m}{n} = 0,$$

and, by assumption, for given  $\varepsilon > 0$ , it is possible to pick  $m_0$  so large that for all  $m \geq m_0$  it holds that

$$\limsup_{n \rightarrow \infty} \frac{2}{nh_n^2} \sum_{i=m}^n \text{cov} \left( K \left( \frac{x - X_0}{h_n} \right), K \left( \frac{x - X_i}{h_n} \right) \right) \leq \varepsilon,$$

where  $\varepsilon$  can be made arbitrarily small. This implies that  $\text{var}(g_n(x)) \rightarrow 0$ , and hence  $g_n(x) \xrightarrow{\text{Pr}} g(x)$ .  $\square$

The assumptions of the following Lemma implies those of Lemma 12, and is thus a special case thereof.

**Lemma 13.** *Assume that  $h_n \rightarrow 0$ ,  $nh_n \rightarrow \infty$ , that the time series  $\{(X_j), j \in \mathbb{Z}\}$  is stationary,  $\alpha$ -mixing with  $\alpha$ -mixing coefficient bounded by  $O(j^{-\eta})$ ,  $\eta > 2$  for every  $j \in \mathbb{Z}$ , and each  $X_j$  admits twice continuously differentiable density  $g$ , and the joint density  $g_{X_i, X_j}$  of  $(X_i, X_j)$  exists, and is bounded for all  $i, j \in \mathbb{Z}$ . Then for any  $x \in \mathbb{R}$ , it holds that  $g_n(x) \xrightarrow{\text{Pr}} g(x)$ .*

**Proof.** The proof is similar to that of Lemma 12, different only in showing that the sum of covariances disappears as  $n \rightarrow \infty$ .

Let  $\{d_n, n \in \mathbb{N}\}$  be an integer sequence given by  $d_n = \lceil h_n^{-2/\eta} \rceil$ . Using Billingsley's inequality with the  $\alpha$ -mixing assumption yields

$$\frac{2}{nh_n^2} \sum_{i=1}^n \text{cov} \left( K \left( \frac{x - X_0}{h_n} \right), K \left( \frac{x - X_i}{h_n} \right) \right) \leq \text{cst} \frac{1}{nh_n^2} \left( \sum_{i=1}^{d_n-1} h_n^2 + \sum_{i=d_n}^{\infty} \alpha(i) \right)$$

$$\leq \text{cst} \left( \frac{d_n}{n} + \frac{d_n^{-\eta+1}}{nh_n^2} \right) \sim \text{cst} (nh_n^{2/\eta})^{-1} \rightarrow 0,$$

since  $2/\eta \leq 1$ . □

**Lemma 14.** *Assume that  $h_n \rightarrow 0$ ,  $nh_n \rightarrow \infty$ , that the time series  $\{(X_j), j \in \mathbb{Z}\}$  is stationary,  $m$ -dependent and each  $X_j$  admits twice continuously differentiable density  $g$ . Then for any  $x \in \mathbb{R}$ , it holds that  $g_n(x) \xrightarrow{\text{Pr}} g(x)$ .*

**Proof.** By the same argument from Lemma 12, the bias is asymptotically negligible. Let  $m \in \mathbb{N}$  be given. By  $m$ -dependence we split  $g_n(x)$  into a main component, which is a sum of independent terms, and a negligible remainder term as follows

$$g_n(x) = \frac{1}{m} \sum_{i=1}^m g_n^{(i)}(x) + R_n(x),$$

where

$$g_n^{(i)}(x) = \frac{m}{nh_n} \sum_{j=1}^{\lfloor \frac{n}{m} \rfloor} K\left(\frac{x - X_{(j-1)m+i}}{h_n}\right), \quad R_n(s) = \frac{1}{nh_n} \sum_{j=\lfloor \frac{n}{m} \rfloor m}^n K\left(\frac{x - X_j}{h_n}\right).$$

Stationarity and  $m$ -dependence implies that each summand in  $g_n^{(i)}(x)$  are i.i.d. which allow for a direct variance calculation

$$\begin{aligned} \text{var} \left( \frac{1}{m} \sum_{i=1}^m g_n^{(i)}(x) \right) &= \frac{1}{m} \text{var} \left( g_n^{(i)}(x) \right) \frac{m}{(nh_n)^2} \left\lfloor \frac{n}{m} \right\rfloor \text{var} \left( K\left(\frac{x - X_0}{h_n}\right) \right) \\ &= \frac{m}{nh_n} \left\lfloor \frac{n}{m} \right\rfloor \left( \frac{\mathbb{E} \left[ K^2\left(\frac{x - X_0}{h_n}\right) \right]}{h_n} - \frac{\mathbb{E} \left[ K\left(\frac{x - X_0}{h_n}\right) \right]^2}{h_n} \right) \rightarrow 0, \end{aligned}$$

using Lemma 11.

Turning our attention to the remainder term, notice that

$$\begin{aligned} \mathbb{E} [R_n(x)^2] &= \frac{1}{n^2 h_n^2} \mathbb{E} \left[ \left( \sum_{j=\lfloor \frac{n}{m} \rfloor m}^n K\left(\frac{x - X_j}{h_n}\right) \right)^2 \right] \\ &= \frac{1}{n^2 h_n^2} \left\{ \sum_{j=\lfloor \frac{n}{m} \rfloor m}^n \mathbb{E} \left[ K^2\left(\frac{x - X_j}{h_n}\right) \right] + 2 \sum_{j=\lfloor \frac{n}{m} \rfloor m}^n \sum_{i=j}^n \mathbb{E} \left[ K\left(\frac{x - X_i}{h_n}\right) K\left(\frac{x - X_j}{h_n}\right) \right] \right\} \\ &\leq \frac{(n - \lfloor \frac{n}{m} \rfloor m)}{n^2 h_n^2} \left\{ \mathbb{E} \left[ K^2\left(\frac{x - X_0}{h_n}\right) \right] + 2 \left( n - \lfloor \frac{n}{m} \rfloor m \right) \sqrt{\mathbb{E} \left[ K^2\left(\frac{x - X_0}{h_n}\right) \right]^2} \right\} \\ &= \frac{(n - \lfloor \frac{n}{m} \rfloor m)}{n^2 h_n} \frac{\mathbb{E} \left[ K^2\left(\frac{x - X_0}{h_n}\right) \right]}{h_n} \left( 1 + 2 \left( n - \lfloor \frac{n}{m} \rfloor m \right) \right) \rightarrow 0, \end{aligned}$$

where we used the Cauchy-Schwarz inequality. The above calculations imply that the remainder is negligible which finishes the proof.  $\square$

**Lemma 15.** Assume that  $h_n \rightarrow 0$ ,  $u_n^x \rightarrow \infty$ ,  $nh_n \rightarrow \infty$ , that the time series  $\{(X_j), j \in \mathbb{Z}\}$  is stationary, each  $X_j$  admits twice continuously differentiable density  $g$ ,

$$\sqrt{k_n h_n} h_n^2 \rightarrow 0, \quad (16)$$

and

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \bar{F}^x(u_n^x) \sum_{j=m}^n \frac{\mathbb{E} \left[ K \left( \frac{x-X_0}{h_n} \right) K \left( \frac{x-X_j}{h_n} \right) \right]}{h_n} = 0.$$

Then

$$\sqrt{k_n h_n} \{g(x) - g_n(x)\} \xrightarrow{\text{Pr}} 0.$$

**Proof.** Consider the following decomposition

$$\sqrt{k_n h_n} \{g_n(x) - g(x)\} = \sqrt{k_n h_n} \{g_n(x) - \mathbb{E}[g_n(x)]\} + \sqrt{k_n h_n} \{\mathbb{E}[g_n(x)] - g(x)\} =: A_n + B_n.$$

In Lemma 12, we showed that the bias is of order  $h_n^2$  and hence  $B_n$  vanishes asymptotically as

$$B_n = \sqrt{k_n h_n} \{\mathbb{E}[g_n(x)] - g(x)\} = O\left(\sqrt{k_n h_n} h_n^2\right) \rightarrow 0,$$

by assumption. The term  $A_n$  has mean zero, and hence it is asymptotically negligible if its variance goes to zero. For fixed  $m \in \mathbb{N}$  we have

$$\limsup_{n \rightarrow \infty} \frac{2\bar{F}^x(u_n^x)}{h_n} \sum_{i=1}^m \text{cov} \left( K \left( \frac{x-X_0}{h_n} \right), K \left( \frac{x-X_i}{h_n} \right) \right) \leq \text{cst} \limsup_{n \rightarrow \infty} m \bar{F}^x(u_n^x) h_n = 0,$$

and for given  $\varepsilon > 0$ , it is possible to pick  $m_0$  so large that for all  $m \geq m_0$  it holds that

$$\limsup_{n \rightarrow \infty} \frac{2\bar{F}^x(u_n^x)}{h_n} \sum_{i=m}^n \text{cov} \left( K \left( \frac{x-X_0}{h_n} \right), K \left( \frac{x-X_i}{h_n} \right) \right) \leq \varepsilon,$$

where  $\varepsilon$  can be made arbitrarily small.  $\square$

**Lemma 16.** Assume that (16) holds, that  $h_n \rightarrow 0$ ,  $u_n^x \rightarrow \infty$ ,  $nh_n \rightarrow \infty$ , that the time series  $\{(X_j), j \in \mathbb{Z}\}$  is stationary,  $\alpha$ -mixing with  $\alpha$ -mixing coefficient bounded by  $O(j^{-\eta})$ ,  $\eta > 2$  for every  $j \in \mathbb{Z}$ , and each  $X_j$  admits twice continuously differentiable density  $g$ .

Then

$$\sqrt{k_n h_n} \{g(x) - g_n(x)\} \xrightarrow{\text{Pr}} 0.$$

**Proof.** The proof is similar to that of Lemma 15, different only in showing that the sum of covariances disappears as  $n \rightarrow \infty$ . By similar calculations as in the proof of Lemma 12 it holds that

$$\text{var}(A_n) \leq \frac{\bar{F}^x(u_n^x)}{h_n} \text{var} \left( K \left( \frac{x-X_0}{h_n} \right) \right) + \frac{2\bar{F}^x(u_n^x)}{h_n} \sum_{i=1}^n \left| \text{cov} \left( K \left( \frac{x-X_0}{h_n} \right), K \left( \frac{x-X_i}{h_n} \right) \right) \right|.$$

The first term has leading term of order  $\bar{F}^x(u_n^x)$ , which vanishes asymptotically. For the second term, we again consider the integer sequence  $\{d_n, n \in \mathbb{N}\}$  given by  $d_n = \lceil h_n^{-2/\eta} \rceil$ . Using Davydov's and Billingsley's inequality with the  $\alpha$ -mixing assumption yields

$$\begin{aligned} & \frac{2\bar{F}^x(u_n^x)}{h_n} \sum_{i=1}^n \left| \text{cov} \left( K \left( \frac{x - X_0}{h_n} \right), K \left( \frac{x - X_i}{h_n} \right) \right) \right| \leq \text{cst} \frac{\bar{F}^x(u_n^x)}{h_n} \left( \sum_{i=1}^{d_n-1} h_n^2 + \sum_{i=d_n}^{\infty} \alpha(i) \right) \\ & \leq \text{cst} \left( \bar{F}^x(u_n^x) h_n d_n + \frac{\bar{F}^x(u_n^x)}{h_n} d_n^{-\eta+1} \right) \sim \text{cst} \bar{F}^x(u_n^x) h_n^{1-\frac{2}{\eta}} \rightarrow 0. \end{aligned}$$

This yields that  $\text{var}(A_n) \rightarrow 0$ , and hence by Chebyshev's inequality we conclude the desired result.  $\square$

## Appendix E: Background on fundamental limit theorems

### Mixing

We introduce  $\alpha$ - and  $\beta$ -mixing using [Rosenblatt \(1956\)](#), [Volkonskii and Rozanov \(1959\)](#) as primary references (see also [Bradley \(2005\)](#), [Kulik and Soulier \(2020\)](#)) and state useful results on  $\beta$ -mixing sequences.

Let  $(\Omega, \mathcal{F}, \text{Pr})$  be a probability space and let  $\mathcal{A}$  and  $\mathcal{B}$  be sub-sigma-algebras of  $\mathcal{F}$ .

**Definition 13.** Define

$$\begin{aligned} \alpha(\mathcal{A}, \mathcal{B}) &= \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |\text{Pr}(A \cap B) - \text{Pr}(A) \text{Pr}(B)|, \\ \beta(\mathcal{A}, \mathcal{B}) &= \sup \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^J |\text{Pr}(A_i \cap B_j) - \text{Pr}(A_i) \text{Pr}(B_j)|, \end{aligned}$$

where the latter supremum is over all pairs of finite partitions  $(A_1, A_2, \dots, A_I)$  and  $(B_1, B_2, \dots, B_J)$  of  $\Omega$  such that  $A_i \in \mathcal{A}$  for each  $i$  and  $B_j \in \mathcal{B}$  for each  $j$ .

The functions  $\alpha$  and  $\beta$  are measures of dependence.

**Definition 14.** Let  $\{X_j, j \in \mathbb{Z}\}$  be a stationary sequence and let  $\mathcal{F}_a^b$  be the  $\sigma$ -field generated by  $(X_a, X_{a+1}, \dots, X_b)$ ,  $-\infty \leq a < b \leq \infty$ . We introduce the sequence of  $\alpha$ -mixing coefficients and  $\beta$ -mixing coefficients respectively as

$$\alpha(n) = \alpha(\mathcal{F}_{-\infty}^0, \mathcal{F}_n^\infty), \quad \beta(n) = \beta(\mathcal{F}_{-\infty}^0, \mathcal{F}_n^\infty).$$

The sequence  $\{X_j, j \in \mathbb{Z}\}$  is said to be

1.  $\alpha$ -mixing (strongly mixing) if  $\lim_{n \rightarrow \infty} \alpha(n) = 0$ ;
2.  $\beta$ -mixing (absolutely regular) if  $\lim_{n \rightarrow \infty} \beta(n) = 0$ .

It holds that  $2\alpha(\mathcal{A}, \mathcal{B}) \leq \beta(\mathcal{A}, \mathcal{B})$  and thus  $\beta$ -mixing implies  $\alpha$ -mixing. Absolute regularity allows for comparison between the laws of a times series and another sequence based on independent blocks. The following Lemma validates the blocking method.

**Lemma 17.** Let  $l \leq r$  and  $m$  be integers. Let  $(\Omega, \mathcal{F}, \Pr)$  be a probability space on which the following sequences are defined, (i) a stationary  $\mathbb{R}^d$ -valued sequence  $\{X_j, j \in \mathbb{Z}\}$  with  $\beta$ -mixing coefficients  $\beta_j$ , (ii) a sequence  $\{X_j, j \in \mathbb{Z}\}$  such that the vectors  $(X_{(i-1)r+1}^\dagger, \dots, X_{ir}^\dagger), i \in \mathbb{Z}$  are mutually independent and each have the same distribution as  $(X_1, \dots, X_r)$ , (iii) a sequence  $\{\xi_j, j \in \mathbb{Z}\}$  of i.i.d. random elements taking value in a measurable set  $(E, \mathcal{E})$ , independent of the previous sequences. Let  $Q_m$  and  $Q_m^\dagger$  be the respective laws of  $\{(\xi_i, X_{(i-1)r+1}, \dots, X_{ir-l}), 1 \leq i \leq m\}$  and  $\{(\xi_i, X_{(i-1)r+1}^\dagger, \dots, X_{ir-l}^\dagger), 1 \leq i \leq m\}$ . Then

$$d_{TV}(Q_m, Q_m^\dagger) \leq m\beta_l,$$

where  $d_{TV}(\cdot, \cdot)$  denotes the total variation distance between measures.

For  $n, k \geq 1$ , let  $h_{n,k} : (\mathbb{R}^d)^k \rightarrow \mathbb{R}$  be measurable function and define

$$\mathbb{H}_n = \sum_{i=1}^{m_n} h_{n,r_n}(X_{(i-1)r_n+1}, \dots, X_{ir_n}), \quad \mathbb{H}_n^\dagger = \sum_{i=1}^{m_n} h_{n,r_n}(X_{(i-1)r_n+1}^\dagger, \dots, X_{ir_n}^\dagger),$$

$$\mathbb{H}_n^0 = \mathbb{H}_n - \mathbb{E}[\mathbb{H}_n], \quad \mathbb{H}_n^{\dagger,0} = \mathbb{H}_n^\dagger - \mathbb{E}[\mathbb{H}_n^\dagger].$$

**Lemma 18.** Let  $\{X_j, j \in \mathbb{Z}\}$  be a stationary sequence in  $\mathbb{R}^d$  with  $\beta$ -mixing coefficients  $\beta_j$ . Let  $\ell_n, m_n, r_n$  be non-decreasing sequences of integers such that  $\ell_n/r_n \rightarrow 0$ . Assume that for each  $n$  there exists a measurable function  $f_n$  such that

$$|h_{n,r_n}(X_1, \dots, X_{r_n}) - h_{n,r_n-\ell_n}(X_1, \dots, X_{r_n-\ell_n})| \leq f_n(X_{r_n-\ell_n+1}, \dots, X_{r_n}),$$

where  $\lim_{n \rightarrow \infty} m_n \mathbb{E}[f_n^2(X_1, \dots, X_{\ell_n})] = 0$ .

If  $\lim_{n \rightarrow \infty} m_n \beta_{l_n} = 0$ , and  $\mathbb{H}_n^{\dagger,0}$  converges weakly, then  $\mathbb{H}_n^0$  converges weakly to the same distribution.

## Central Limit Theorems

In this section we present existing results on regarding Central Limit Theorems using [Billingsley \(2009\)](#) and [van der Vaart and Wellner \(2023\)](#) as primary reference, see also [Kulik and Soulier \(2020\)](#).

For  $a < b$ , define by  $\mathbb{D}[a, b]$  the space of cadlag functions indexed on the compact interval  $[a, b]$ . We are interested in establishing weak convergence in  $\mathbb{D}[a, b]$  endowed with the  $J_1$ -topology (also called the Skorohod topology)

**Definition 15.** For  $x$  and  $y$  in  $\mathbb{D}$ , define

$$d_{J_1}(x, y) = \inf_{\lambda \in \Lambda} \{\|\lambda - I\| \vee \|x - y(\lambda)\|\},$$

where  $\Lambda$  is the set of strictly increasing and continuous mappings defined on the index set of  $\mathbb{D}$  onto itself and  $I$  is the identity map on the index set. Then  $d_{J_1}$  is a metric on  $\mathbb{D}$  and we denote by  $J_1$  the topology metrizable by this metric.

**Theorem 16 (Asymptotic equicontinuity with uniform modulus of continuity).** *Assume that a sequence of  $\mathbb{D}[a, b]$ -valued stochastic processes  $(\mathbb{X}_n)$  has converging finite-dimensional distributions. If moreover, for any  $\varepsilon > 0$*

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \Pr \left( \sup_{\substack{a \leq s \leq t \leq b \\ |s-t| \leq \delta}} |\mathbb{X}_n(t) - \mathbb{X}_n(s)| > \varepsilon \right) = 0, \quad (17)$$

*then there exists a stochastic process  $\mathbb{X}$ , with  $\Pr(\mathbb{X} \in \mathbb{C}[a, b]) = 1$ , such that  $\mathbb{X}_n$  converges weakly to  $\mathbb{X}$  in  $\mathbb{D}[a, b]$  endowed with the  $J_1$  topology.*

As the condition of equation (17) is difficult to show directly, one often invokes the covering Central Limit Theorem when dealing with arrays of stochastic processes. In the theorem we below we present an adaption to the covering CLT in the case of separable processes.

**Definition 17 (Separability).** Let  $(\mathcal{G}, \rho)$  be a semi-metric space. A process  $\{\mathbb{X}(f), f \in \mathcal{G}\}$  is said to be separable if there exists a countable sub-class  $\mathcal{G}_0 \subset \mathcal{G}$  an  $\Omega_0 \subseteq \Omega$  with  $\Pr(\Omega_0) = 1$  such the for all  $\omega \in \Omega_0$ ,  $f \in \mathcal{G}$  and  $\varepsilon > 0$ ,

$$\mathbb{X}(f, \omega) \in \overline{\{\mathbb{X}(g, \omega) : g \in \mathcal{G}_0 \cap B(f, \varepsilon)\}},$$

where  $B(f, \varepsilon)$  is the open ball around  $f$  with radius  $\varepsilon$  wrt  $\rho$ .

**Definition 18 (Covering number).** The covering number  $N(\mathcal{G}, d, \varepsilon)$  is the minimal number of balls  $\{g : d(g, f) < \varepsilon\}$  of radius  $\varepsilon$  needed to cover the set  $\mathcal{G}$ . The centres of the balls need not belong to  $\mathcal{G}$ , but the should have finite norms.

**Lemma 19.** *Assume  $\mathbb{X}$  is a separable stochastic process indexed by the set  $\mathcal{F}$ . Then  $\|\mathbb{X}\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} \{\mathbb{X}(f)\}$  is measurable.*

Let  $\mathbb{Z}_n$  be the centred empirical process indexed by a class of functions  $\mathcal{G}$  defined by

$$\mathbb{Z}_n(f) = \sum_{i=1}^{m_n} \{Z_{n,i}(f) - \mathbb{E}[Z_{n,i}(f)]\}, \quad f \in \mathcal{G},$$

where  $m_n \rightarrow \infty$  as  $n \rightarrow \infty$  and  $\{Z_{n,i}, n \in \mathbb{N}\}, i = 1, \dots, m_n$  are i.i.d. separable stochastic processes. We define the random semi-metric  $d_n$  on  $\mathcal{G}$  by

$$d_n^2(f, g) = \sum_{i=1}^{m_n} \{Z_{n,i}(f) - Z_{n,i}(g)\}^2, \quad f, g \in \mathcal{G}.$$

**Theorem 19.** *Consider the sequence of processes*

$$\mathbb{Z}_n(f) = \sum_{i=1}^{m_n} \{Z_{n,i}(f) - \mathbb{E}[Z_{n,i}(f)]\},$$

*where for each  $n \in \mathbb{N}$ ,  $\{Z_{n,i}, i = 1, \dots, m_n\}$ , are i.i.d. separable stochastic processes and assume that the space  $(\mathcal{G}, \rho)$  is totally bounded. Suppose that*

1. For all  $\eta > 0$ ,

$$\lim_{n \rightarrow \infty} m_n \mathbb{E} \left[ \left\| Z_{n,1} \right\|_{\mathcal{G}}^2 \mathbf{1} \left\{ \left\| Z_{n,1} \right\|_{\mathcal{G}} > \eta \right\} \right] = 0.$$

2. For every sequence  $\{\delta_n\}$  decreasing to zero,

$$\lim_{n \rightarrow \infty} \sup_{\substack{f, g \in \mathcal{G} \\ \rho(f, g) \leq \delta_n}} \mathbb{E} \left[ d_n^2(f, g) \right] = 0.$$

3. There exists a measurable majorant  $N^*(\mathcal{G}, d_n, \varepsilon)$  of the covering number  $N(\mathcal{G}, d_n, \varepsilon)$  such that for every sequence  $\{\delta_n\}$  which decreases to zero,

$$\int_{[0, \delta_n]} \sqrt{\log(N^*(\mathcal{G}, d_n, \varepsilon))} d\varepsilon \xrightarrow{\text{Pr}} 0. \quad (18)$$

Then  $\{Z_n, n \in \mathbb{N}\}$  is asymptotically  $\rho$ -equicontinuous.

Consider a sequence of stochastic processes

$$Z_{n,i}(f) = \frac{1}{\sqrt{v_n}} f(Y_{n,i}), \quad i = 1, \dots, m_n,$$

where  $Y_{n,i}$  are i.i.d. random elements in a measurable space  $E$  and  $v_n$  is non-decreasing tending to infinity. For a class of real-valued functions  $\mathcal{G}$  defined on  $E$ , define the envelope function  $\mathbf{G} : E \rightarrow \mathbb{R}$  by

$$\mathbf{G}(y) = \sup_{f \in \mathcal{G}} |f(y)|.$$

Denote by  $\mathcal{P}$  the set of finite discrete probability measures on  $E$ . For  $P \in \mathcal{P}$  define  $L^2(P) = \{f : E \rightarrow \mathbb{R}, P(f^2) < \infty\}$ . For each  $f, g \in L^2(P)$  write

$$\|f\|_{L^2(P)} = \sqrt{P(f^2)}, \quad d_{L^2(P)}(f, g) = \sqrt{P(\{f - g\}^2)}.$$

The entropy condition of equation (18) in Theorem 19 can be replaced with a uniform metric entropy condition with respect to the deterministic semi-metric  $d_{L^2(P)}$ .

**Lemma 20.** Assume that

$$\limsup_{n \rightarrow \infty} \frac{m_n \mathbb{E} \left[ \mathbf{G}^2(Y_{n,1}) \right]}{v_n} < \infty, \quad (19)$$

and

$$\lim_{\delta \rightarrow 0} \int_{[0, \delta]} \sup_{Q \in \mathcal{P}} \sqrt{\log \left( N \left( \mathcal{G}, d_{L^2(P)}, \|\mathbf{G}\|_{L^2(Q)} \varepsilon \right) \right)} d\varepsilon = 0. \quad (20)$$

Then equation (18) holds.

**Lemma 21.** *If  $\mathcal{G}$  is a VC-subgraph class with measurable envelope function  $\mathbf{G}$ , then (20) holds. In particular, (20) holds for linearly ordered classes.*

**Theorem 20 (Vervaat's Lemma).** *Consider non-empty intervals  $I, J$  and let  $\{\mathbb{X}, \mathbb{X}_n, n \geq 1\}$  be  $\mathbb{D}(I)$ -valued stochastic processes and assume that  $\mathbb{X}$  has almost surely continuous sample paths. Let  $h : I \rightarrow J$  be a surjective, continuous and increasing function with positive derivative  $h'$ . If there exists a deterministic sequence  $c_n \rightarrow \infty$  such that  $c_n (\mathbb{X}_n - h) \xrightarrow{d} \mathbb{X}$  in  $\mathbb{D}(I)$  endowed with the  $J_1$ -topology and  $\mathbb{X}_n, n \geq 1$  have non-decreasing sample paths, then*

$$c_n (\mathbb{X}_n - h, \mathbb{X}_n^\leftarrow - h^\leftarrow) \xrightarrow{d} (\mathbb{X}, -(h^\leftarrow)' \times \mathbb{X} \circ h^\leftarrow),$$

*in  $\mathbb{D}(I) \times \mathbb{D}(J)$  endowed with the product  $J_1$  topology.*

A simple application of the Continuous Mapping Theorem yields that Vervaat's Lemma holds in the case where  $h$  is decreasing and  $\mathbb{X}_n, n \geq 1$  have non-increasing sample paths.