Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Estimators for temporal dependence of extremes

Dissertation

zur Erlangung des Doktorgrades
an der Fakultät für Mathematik, Informatik und Naturwissenschaft
Fachbereich Mathematik
der Universität Hamburg

vorgelegt von

Sebastian A. Neblung

Hamburg, 2021

Folgende Gutachter empfehlen die Annahme der Dissertation:
Prof. Dr. Holger Drees (Universität Hamburg)
Prof. Dr. Anja Janßen (Otto-von-Guericke Universität Magdeburg)
Dr. Stanislav Volgushev, Assistant Professor (University of Toronto)

Vorsitzender der Prüfungskommission:
Prof. Dr. Thomas Schmidt (Universität Hamburg)

Tag der Disputation: 05. August 2021

Acknowledgment

During the preparation of this dissertation, some people have supported me in particular, whom I would like to express my gratitude.

First, I would like to thank my supervisor Prof. Dr. Holger Drees and co-supervisor Prof. Dr. Anja Janßen. Holger contributed to the thematic orientation of the dissertation, always found (even spontaneously) time for discussions with me and gave me many valuable tips. Anja also took her time and gave many helpful hints which contributed to the quality of this thesis. In particular, I would like to thank Anja for the opportunity of a research visit in Stockholm. Anja and Holger opened doors for me to the international scientific community, for which I am grateful to both of them.

As a reviewer, Dr. Stanislav Volgushev took the time to work through the entire dissertation, for which I am really grateful.

I wish to acknowledge the anonymous referees of my papers (see A.3), whose comments led to an improved presentation in this thesis.

All my colleagues in the working group Mathematical Statistic and Stochastic Processes at the University of Hamburg deserve my gratitude for providing a pleasant working atmosphere (even digitally during the pandemic) and for fruitful discussions. Moreover, there was always room for private conversations and many nice moments in- and outside of the Geomatikum. In particular, I sincerely wish to thank my fellow doctoral students who took it upon themselves to proofread this dissertation: Dr. Florian Hildebrandt, Dr. Miran Knežević, Suada Mulgeci, Dr. Berenice Neumann, Jan Rabe and Maximilian Steffen.

Finally, I like to express my gratitude to my family, who encouraged me in my decision to do my PhD and always supported me in continuing on my way.

Thank you all very much!

Contents

List of Figures

List of Tables

Symbols, Notation and Conventions

This list contains a selection of symbols, abbreviations and conventions used in this thesis, including an indication of where they first appear, if they are explicitly defined.

$\asymp$	asymptotically of the same order
$\sim$	asymptotic equivalent or distributed as
$\lfloor\cdot\rfloor$, $\lceil\cdot\rceil$	floor function, ceiling function
$\overset{d}{=}$	equal in distribution
$\to$	converges to, usually as $n \to \infty$ if not indicated otherwise
$\overset{w}{\to}$, $\overset{P}{\to}$, $\overset{P^*}{\longrightarrow}$	weak convergence, convergence in probability, convergence in outer probability, usually as $n \to \infty$
$\lVert\cdot\rVert$	some norm on a metric space E or $\mathbb{R}^d$
$\lVert\cdot\rVert_{P,2}$	L_2-norm w.r.t. some probability measure P
$\lVert\cdot\rVert_{TV}$	norm of total variation for measures
$\lVert(z_t)_{t\in\mathbb{Z}}\rVert_\alpha$	$(\sum_{t\in\mathbb{Z}} \lVert z_t\rVert^\alpha)^{1/\alpha}$, p.14
$\vee$, $\wedge$	maximum, minimum
$\leq_L$	Loewner order
f^+, f^-	$\max(f,0)$, $-\min(f,0)$, absolute positive and negative part of a function f
$\lim_{n\to\infty}$	limit for $n \to \infty$
o, O	small and big Landau symbol
o_P, O_P	small and big stochastic Landau symbol
∂A	topological boundary of the set A
A^C	complement of a set A
$\frac{\partial}{\partial x_j} f(x_1,...,x_d)$	j-th partial derivative of the function f defined on $\mathbb{R}^d$
R^{RS}, Q^{RS}	RS-transformed of a process R or a measure Q, p.15
$U^*_{s,t}$	$\sup_{s\leq\lvert j\rvert\leq t} \lVert U_j\rVert$ for $s < t$ and some process $(U_t)_{t\in\mathbb{Z}}$, p.70
$x \in \mathbb{R}^d$	$x = (x_1,...,x_d)$ with $x_j \in \mathbb{R}$, $1 \leq j \leq d$
$x \leq y \in \mathbb{R}^d$	componentwise $x_j \leq y_j$ for $1 \leq j \leq d$
$x < y \in \mathbb{R}^d$, $x \lneq y$	$x \leq y$ and $x \neq y$
$x \vee y \in \mathbb{R}^d$	componentwise $(x_1 \vee y_1, ..., x_d \vee y_d)$ ($\wedge$ analog)
δ_x	Dirac measure with pointmass 1 in x

F_X	distribution function (cdf) of random variable X
$F_X^{\leftarrow}$	quantile function /inverse cdf of random variable X
$\mathbb{1}_A$	indicator function of the set A
$\mathcal{L}(X)$, P^X	both notations for the distribution of X
$\mathcal{L}(X \mid Y)$	conditional distribution of X given Y
$\mathcal{N}_k(\mu, \Sigma)$	k-dimensional normal distribution with mean μ and covariance Σ ($k = 1$ is usually omitted)
P, $E[\cdot]$	probability measure, expectation
P^*, $E^*[\cdot]$	outer probability, outer expectation
$\mathrm{Par}(\alpha)$	Pareto distribution with parameter α
$Var(X)$, $Cov(X,Y)$	variance of a random variable X, covariance of X and Y
$\mathcal{A}$	some family of Borel sets
$\mathcal{B}(M)$	Borel-σ-algebra for the set M
$\mathcal{G}$	some family of real valued functions on l_α
l_0	$\{(x_t)_{t\in\mathbb{Z}} \in (\mathbb{R}^d)^{\mathbb{Z}} \mid \lim_{\lvert t\rvert\to\infty} \lvert x_t\rvert = 0\}$, p.70
l_α	$\left\{z \in (\mathbb{R}^d)^{\mathbb{Z}} \mid 0 \leq \lVert z\rVert_\alpha^\alpha < \infty\right\}$, p.12
$l^\infty(M)$	space of real-valued uniform bounded functions on the set M, p.23
$\mathbb{N}$	$\{1, 2, 3, ...\}$ set of natural numbers
$\mathbb{N}_0$	$\{0, 1, 2, 3, ...\}$ set of natural numbers including 0
$\mathbb{R}$, $\mathbb{R}^+$, $\bar{\mathbb{R}}$	real numbers, $[0, \infty)$ and $\mathbb{R} \cup \{-\infty, \infty\}$
$\mathbb{Z}$	set of integers
$\beta_{n,k}^\Gamma$	β-mixing coefficient of $(\Gamma_{n,i})_{1\leq i\leq m_n}$, p.21
$N(\varepsilon, \mathcal{G}, d_n)$	ε-covering number of $\mathcal{G}$ w.r.t. the semi-metric d_n, p.28
$N_{[\cdot]}(\varepsilon, \mathcal{G}, \lVert\cdot\rVert)$	ε-bracketing number of $\mathcal{G}$ w.r.t. the norm $\lVert\cdot\rVert$, p.27
α	tail index of regular variation
$\hat{\alpha}_n$	Hill-type estimator for α, p.142
$M_{j,k}$	$\max(X_j, ..., X_k)$ for $j \leq k$ and $X_j, ..., X_k$ real valued, p.70
$\nu^*(H)$	cluster index for the cluster functional H, p.69
p_A	$P(\Theta_i \in A)$
$\hat{p}_{n,A}$	projection based estimator for p_A with known α, p.142
$\hat{\hat{p}}_{n,A}$	projection based estimator for p_A with estimated α, p.142
$\hat{p}_{n,A}^f$, $\hat{p}_{n,A}^b$	forward and backward estimator for p_A, p.139
s_n, r_n, l_n	sliding block, big block, small block lengths
$T_{n,A}$	$\sum_{t=1}^n g_A(X_{n,t-s_n}, ..., X_{n,t+s_n})$, statistic in Chapter 5, p.144
ϑ	candidate extremal index, p.70
θ	extremal index, p.74

$\hat{\theta}_n^s$, $\hat{\theta}_n^d$, $\hat{\theta}_n^r$	sliding blocks, disjoint blocks and runs estimator for θ, pp.75 sq.
$\theta_{sl}(S)$	stop loss index, p.89
$\hat{\theta}_{sl,n}^s(S)$, $\hat{\theta}_{sl,n}^d(S)$	sliding and disjoint blocks estimator for $\theta_{sl}(S)$, p.92
$\hat{\theta}_{sl,n}^r(S)$	runs estimator for $\theta_{sl}(S)$, p.90
$\Theta = (\Theta_t)_{t\in\mathbb{Z}}$	spectral tail process, p.10
u_n	thresholds for extreme values
v_n, p_n	probability of one extreme observation and of at least one extreme observation in a big block
$W_{n,t}$	$(X_{t+h}/u_n)_{\lvert h\rvert\le s_n}$ block of observations in Chapter 5 (defined slightly differently as $(X_{n,t+h})_{0\le h\le s_n-1}$ in Chapter 3 and parts of Chapter 4)
$X = (X_t)_{t\in\mathbb{Z}}$	(strict) stationary $\mathbb{R}^d$-valued time series
$Y = (Y_t)_{t\in\mathbb{Z}}$	tail process, p.10

AR-model	autoregressive model
a.s.	almost surely
fidi	finite dimensional marginal distribution
GARCH-model	generalized autoregressive conditional heteroscedasticity model
iid	independent and identically distributed
POT	peak-over-threshold
RMSE	root mean squared error
SRE	stochastic recurrence equation
SR-model	stochastic recurrence equation model
SV-model	stochastic volatility model
TCF	time change formula
w.l.o.g.	without loss of generality
w.r.t.	with respect to

We typically embed $(\mathbb{R}^d)^{t-s+1}$ in $(\mathbb{R}^d)^{\mathbb{Z}}$ (or l_0) by the mapping $(x_s, ..., x_t) \mapsto x = (x_h)_{h\in\mathbb{Z}} \in (\mathbb{R}^d)^{\mathbb{Z}}$ with $x_h := 0$ for $h \notin \{s, ..., t\}$.
An additional overview of sequences like r_n, l_n, v_n, p_n occurring in Chapter 3 is given in Table 3.1.

Chapter 1

Introduction

Extreme value theory, as a discipline of mathematical statistics, deals with the modeling and statistical treatment of rare and extreme risks. Unlike for frequently occurring events, there are not enough empirical observations for rare and extreme events such that standard statistical methods often underestimate the probabilities of such events. Nevertheless, extreme events in particular have serious consequences for humans, the environment and the economy. For this reason, modeling these probabilities of extreme events as accurately as possible is extremely important for risk management and this is where extreme value theory comes into play. Examples of such extreme events are natural disasters such as floods, heavy rainfall, heat waves, air concentrations of pollutants and other natural catastrophes. In flood protection, dikes should be built high enough to hold back almost all floods. However, there are usually only few water levels of floods and many data of normal water levels. These non-extreme observations have to be extrapolated in a suitable way to estimate the probabilities of flooding as accurately as possible. Extreme value theory provides tools for this, e.g. one can use extreme value distributions in order to estimate a value-at-risk for future floods. Extreme events are also important in the economy, such as crashes on the stock market or in reinsurance. On the financial market there is much data available for normal price fluctuations, but an investor should also consider extreme losses for an adequate risk coverage. Probabilities for this can also be determined from data with tools of the extreme value theory. Further and more detailed descriptions of the illustrating examples for fields of application of extreme value theory can be found e.g. in Beirlant et al. (2004), Coles (2001), De Haan and Ferreira (2006) or Embrechts et al. (2013). Further examples from engineering can be found in Castillo (2012) and from finance in Finkenstädt and Rootzén (2003). More examples for the concrete statistical application of extreme value theory in the context of insurance, finance and hydrology can be found in Reiss and Thomas (1997).

For one-dimensional and independent data, extreme value statistics is already well developed, for an overview see De Haan and Ferreira (2006). For dependent data the further development of extreme value statistics is part of the ongoing research. The dependencies of extreme events are of great importance for the overall risk assessment. For example, a heavy rainfall on one day can cause short-term flooding, but if it rains heavily for several days in a row, it can drench the ground and, in the case of mountainous unstable ground, lead to mudslides. Another example is found on the financial market, where the extreme loss of one value in the portfolio can possibly be compensated by other items, but a longer period of extreme losses or a simultaneous extreme loss in several positions can quickly lead to bankruptcy. Therefore, to understand the overall extreme behavior of a time series, it is important to statistically investigate the dependencies of extreme events. It is often observed in empirical data that extreme events occur over time in clusters and not alone. This dissertation deals with temporal dependencies in extreme data and their statistical treatment.

The modern foundation of extreme value theory for time series was introduced by Basrak and Segers (2009) and has been widely used in the literature since then, see the overview in Kulik and Soulier (2020). Usually a (strict) stationary $\mathbb{R}^d$-valued time series $(X_t)_{t\in\mathbb{Z}}$ is considered, which is regularly varying with an index α (see Definition 2.1.5). An observation of this time series is considered as extreme if the norm of this observation exceeds a certain threshold u_n, i.e. $\|X_t\| > u_n$ for some norm $\|\cdot\|$ on $\mathbb{R}^d$ (this is the peak-over-threshold (POT) setting). The spectral tail process $(\Theta_t)_{t\in\mathbb{Z}}$ is then defined by the weak convergence

$$\mathcal{L}\left((X_s, ..., X_t)/\|X_0\| \mid \|X_0\| > u_n\right) \to \mathcal{L}\left(\Theta_s, ..., \Theta_t\right)$$

as $n \to \infty$ for $s \leq t \in \mathbb{Z}$ and $u_n \to \infty$. Thus, it describes asymptotically the extreme behavior of the time series $(X_t)_{t\in\mathbb{Z}}$, given that at the fixed time 0 an extreme event occurred and $(X_t)_{t\in\mathbb{Z}}$ is standardized such that $\|\Theta_0\| = 1$ a.s. Hence, by this definition, the spectral tail process can be used to describe the extreme dependency of the time series $(X_t)_{t\in\mathbb{Z}}$, independently of the heaviness of the tail of the distribution of $\|X_0\|$. In other words, the spectral tail process contains all information about the extreme dependence structure of the underlying time series $(X_t)_{t\in\mathbb{Z}}$. The process $(\Theta_t)_{t\in\mathbb{Z}}$ is not stationary, nevertheless it fulfills a certain structural property called the *time change formula* (Definition 2.2.2), which follows from the stationarity of $(X_t)_{t\in\mathbb{Z}}$.

If one wants to pursue statistical inference for the extreme dependency of $(X_t)_{t\in\mathbb{Z}}$, it is useful to estimate the distribution of $(\Theta_t)_{t\in\mathbb{Z}}$ or the probability of some events depending on the spectral tail process, respectively. From this, one could approximate the probabilities of extreme events of $(X_t)_{t\in\mathbb{Z}}$ by using the definition of $(\Theta_t)_{t\in\mathbb{Z}}$.

The estimation of the distribution of $(\Theta_t)_{t\in\mathbb{Z}}$ with Markovian structure was already considered in Drees et al. (2015). There, a naive empirical estimator, named the forward estimator, and the so-called backward estimator are considered. This backward estimator is derived from the empirical estimator by using a partial aspect of the time change formula. In simulation studies it has been shown that for certain sets A this estimator performs better than the empirical estimator when estimating $P(\Theta_i \in A)$. This estimation approach was generalized by Davis et al. (2018) for general real-valued time series $(\Theta_t)_{t\in\mathbb{Z}}$ with the same result.

New estimators for the distribution of the spectral tail process

To provide new and improved estimators for the distribution of $(\Theta_t)_{t\in\mathbb{Z}}$ is one goal of this thesis. The backward estimator uses only a fraction of the structure described by the time change formula. Janßen (2019) introduced with the RS-transformation (Definition 2.2.4) an equivalent formulation of this structural property. This RS-transformation makes it possible to derive a new estimator for the distribution of $(\Theta_t)_{t\in\mathbb{Z}}$ by using the entire structural property of the time change formula. In this thesis, we introduce this estimator as so-called projection based estimator and for the estimation of $P(\Theta_i \in A)$ for a fixed $i > 0$ we define it by

$$\hat{p}_{n,A} := \frac{1}{\sum_{t=1}^{n} \mathbb{1}_{\{\|X_t\|>u_n\}}} \sum_{t=1}^{n} \mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h=-s_n}^{s_n} \frac{\|X_{t+h}\|^\alpha}{\sum_{k=-s_n}^{s_n} \|X_{t+k}\|^\alpha} \times \left(\mathbb{1}_{\{h\le s_n-i\}} \mathbb{1}_A\left(\frac{X_{t+h+i}}{\|X_{t+h}\|}\right) + \mathbb{1}_{\{h>s_n-i\}} \mathbb{1}_A(0) \right),$$

for observations $X_{1-s_n}, ..., X_{n+s_n}$, some $s_n \in \mathbb{N}$ and the index α of the regular variation of the time series $(X_t)_{t\in\mathbb{Z}}$. A theoretical advantage of this estimator is, that, if we estimate for a family of sets $\mathcal{A}$, the estimated distribution of $(\Theta_t)_{t\in\mathbb{Z}}$ fulfills the structural property of the time change formula and, thus, surely itself is the distribution of a spectral tail process. Neither the forward nor the backward estimator known from the literature possess this property. In some sense, the application of the RS-transformation is a projection on the set of admissible distributions for the spectral tail process. Projection methods for the construction of estimators are also used in the literature, see e.g. Fils-Villetard et al. (2008), however they use a different kind of projection. The method to construct this estimator $\hat{p}_{n,A}$ can easily be generalized to define other estimators for probabilities of events depending on the spectral tail process. This method introduces a new approach to construct general estimators for the extremal dependence of stationary time series.

Simulation results presented below also show that our new estimator often has a smaller RMSE and, thus, can perform better on a finite sample than the estimators from the literature. In particular, the variance of this new estimator is smaller than the variance of the aforementioned two competing estimators, whereas the bias can be slightly larger when estimating $P(\Theta_i \le x)$ for a small $|x|$. Overall, the projection-based estimator is a new and useful alternative to estimating the distribution of $(\Theta_t)_{t\in\mathbb{Z}}$. The use of the whole structure of the time change formula improves the estimation results.

In this thesis, the motivation of this estimator $\hat{p}_{n,A}$ is discussed and the asymptotic behavior of the appropriately normalized estimator is analyzed. For this, two main problems have to be considered. First, the index of the regular variation α is generally unknown and has to be estimated itself. To this end, α has to be replaced by a suitable estimator. This makes the proof for asymptotic normality technically more complex and requires some additional conditions. The second more fundamental problem in the asymptotic analysis of the projection based estimator is that this estimator is a so-called sliding blocks estimator, for whose asymptotic treatment there exist no general suitable limit theorem in the literature. More concretely, this means that the numerator of $\hat{p}_{n,A}$ has the form

$$\sum_{t=1}^{n} g_A(X_{n,t-s_n}, ..., X_{n,t+s_n})$$

for the blocks of observations $(X_{n,t-s_n}, ..., X_{n,t+s_n})$ and a suitable function g_A, where $X_{n,t} = X_t/u_n$. In particular, two successive blocks have a considerable overlap of $2s_n$ X-observations, which causes dependencies between the individual summands in the estimator and complicates the asymptotic treatment.

Sliding blocks estimators and their asymptotic behavior

The consideration and discussion of this second problem in a much more general framework will be a substantial contribution of this thesis: In extreme value statistics, one often considers estimators, which are defined as the average or sums of block statistics $g(W_{n,t})$ for suitable functions g. Here

$$W_{n,t} = (X_{n,t-s_n}, ..., X_{n,t+s_n})$$

is a block of observation for a growing sequence s_n and in the peak-over-threshold (POT) setting $X_{n,t}$ is a X_t-measurable random variable, e.g. $X_{n,t} = X_t/u_n \mathbb{1}_{\{\|X_t\|>u_n\}}$ for a stationary time series $(X_t)_{t\in\mathbb{Z}}$. Typical examples are estimators of the extremal index (cf. Section 4.2), the empirical extremogram (Davis and Mikosch, 2009*b*) or the cluster size distribution (Hsing, 1991). A more recent example are estimators for the ordinal pattern in extremes (Oesting and Schnurr, 2020). Such block statistics can either be defined as averages over $g(W_{n,t})$ with $1 \le t \le n$, in which case we have so-called sliding blocks (overlapping blocks), as in our projection based estimator $\hat{p}_{n,A}$. Or it could be an average over $g(W_{n,t(2s_n+1)+1})$ with $0 \le t \le \lfloor n/(2s_n + 1)\rfloor - 1$. In the latter case, the individual blocks have no overlap and they are so-called disjoint blocks. The sliding blocks method uses more data, but this data has also stronger dependencies than disjoint blocks.

For the asymptotic analysis of general disjoint blocks statistics in the POT setting the powerful results of Drees and Rootzén (2010) can be used, as it is done for example in Drees et al. (2015), Davis et al. (2018) or Drees and Knezevic (2020). However, the setting of Drees and Rootzén (2010) is too restrictive for the treatment of sliding blocks estimators, and there is no other directly suitable and known result in the literature. Some specific sliding blocks statistics are analyzed in the literature e.g. in Bücher and

Segers (2018*a*), Bücher and Jennessen (2020*b*), Zou et al. (2021) or recently Cissokho and Kulik (2021) and Oesting and Schnurr (2020), but their results are really specific for their problems and there is no general result for sliding blocks statistics. In general the asymptotic analysis of sliding blocks statistics is much more complex than the analysis of disjoint blocks. For example Northrop (2015) proposed a sliding blocks estimator for the extremal index but did not consider any asymptotic results due to the complex methods needed.

In this thesis we introduce a first general setting which allows for a systematic asymptotic analysis of blocks statistics in the POT setting. Based on the setting in Drees and Rootzén (2010), an even more abstract setting for the derivation of a uniform central limit theorem for general suitably standardized blocks statistics is derived. In this framework the setting of Drees and Rootzén (2010) for the treatment of disjoint blocks statistics can be embedded as well as a more special setting which can be used to deal with sliding blocks statistics. The result is a uniform central limit theorem for standardized blocks statistics. More precisely, this is a uniform central limit theorem for the empirical process

$$Z_n(g) = \frac{1}{\sqrt{p_n} b_n(g)} \sum_{t=1}^{n} (g(W_{n,t}) - E[g(W_{n,t})]), \quad g \in \mathcal{G},$$

for some suitable function class $\mathcal{G}$ and normalization $b_n(g)$ and $p_n = P(\exists g \in \mathcal{G} : g(W_{n,1}) \neq 0) \to 0$. In particular, the abstract setting developed in this thesis provides a common basis for deriving asymptotic statements for disjoint blocks statistics and sliding blocks statistics under unified conditions. This limit theorem is the main tool to solve the second problem of the asymptotic treatment of the projection based estimator $\hat{p}_{n,A}$ mentioned above and to achieve an asymptotic normality result for this new type of estimator.

In the literature it has been suggested that sliding blocks are often more efficient, see Beirlant et al. (2004), p. 390, for a statement on the extremal index. In fact this has only been proven in a few concrete examples and a general result in the POT setting is not known. Robert et al. (2009) have proven for a particular estimator for the extremal index that the sliding blocks version of their estimator always has a strictly smaller variance than the disjoint blocks version. In the so-called block maxima setting only the value of the maximum observation per block is included in the statistic, this is different from the POT setting, where all observations whose norm exceeds a certain limit u_n are used for the estimator. Because of the alternative definition of observations which are considered to be extreme, there is a different asymptotic behavior for estimators in the block maxima setting, especially more observations are included in the estimation. A general comparison of the performance in the POT and block maxima setting can be found e.g. in the overview article of Bücher and Zhou (2018). A comparison of different assumptions in both settings can be found e.g. in Bücher et al. (2019) for second order conditions. In the block maxima setting Zou et al. (2021) has proven under quite general conditions that sliding blocks statistics are at least as efficient as disjoint blocks statistics. Zou et al. (2021) has shown

this more concretely for estimators for copulas. Bücher and Segers (2018*a*) also observed for the maximum likelihood estimator of the parameters of a Fréchet distribution in the block maxima setting that the sliding blocks estimator is more efficient.
By applying the general abstract limit theorem developed in Chapter 3 below, comparable conditions and weak convergences can be derived for disjoint blocks statistics and related sliding blocks statistics. For these statistics, variances can be compared, with the result that also in the POT setting the asymptotic variance of a quite general sliding blocks statistic is never greater than the asymptotic variance of the disjoint blocks counterpart. This implies that, in contrast to common practice, sliding blocks statistics should be considered primarily. The projection estimator $\hat{p}_{n,A}$ results directly from the motivation as a sliding blocks statistic, but due to the general results it makes no sense to analyze a disjoint blocks counterpart.

Estimators for cluster indexes

Besides the spectral tail process $(\Theta_t)_{t\in\mathbb{Z}}$, there are also a number of other quantities and indexes which can be used to describe certain properties of the extremal dependency of a stationary time series $(X_t)_{t\in\mathbb{Z}}$. A whole family of such indexes are the so-called cluster indexes which are systematically defined in Definition 4.1.1, see also Kulik and Soulier (2020), Chapters 6 and 10. Each cluster index describes a certain property of the extreme dependence structure. For cluster indexes, disjoint and sliding blocks estimators can be motivated, which can be analyzed asymptotically in the above mentioned general setting. The family of cluster indexes also includes well-known indexes defined independently of this generalizing concept, such as the extremal index or the deviation index (Mikosch and Wintenberger, 2013). The extremal index θ was introduced by Leadbetter (1983) and is the inverse of the mean cluster length, where a cluster is defined asymptotically as consecutive extreme observations. Thus, the extremal index is a measure for how many extreme observations occur together on average. A disjoint blocks estimator for θ was introduced by Hsing (1991). For blocks estimators all extreme observations in a block are interpreted as a cluster. Another interpretation of a cluster is that all extreme observations that are not separated by a certain number of non-extreme observations form a cluster. This interpretation leads to so-called runs estimators as motivated by Hsing (1993). For both estimators, the asymptotic normality under different conditions has been proven by Weissman and Novak (1998). In this thesis we prove the asymptotic normality again under slightly different and especially comparable conditions. For the application of the abstract setting for the runs estimator note that each runs estimator can be interpreted as a special sliding blocks estimator. As a result we get that both estimators have the same asymptotic variance, a fact that was unknown so far. This shows that the uniform abstract setting developed in this thesis allows a gain in information. Furthermore, in this work the asymptotic normality of the sliding blocks estimator for the extremal index

is proven for the first time, whereby the asymptotic variance is again the same as for the other two estimators.
More generally, sliding and disjoint blocks estimators as well as runs estimators for any cluster index can be analyzed with the limit theorems of the abstract setting or with the limit theorem for sliding blocks statistics. In this thesis, this is also done for the example of the family of stop loss indexes. The stop loss index describes the distribution of the overall extreme losses given one loss at time point 0. It is shown that for these cluster indexes the sliding and disjoint blocks estimators have the same asymptotic distribution. This is in agreement with the result of Cissokho and Kulik (2021), who recently showed this for a large class of cluster indexes, but using slightly different conditions. Among others, they use a so-called ANSJB condition which controls the occurrence of small jumps in the time series. In our analysis we also consider the runs estimator for the stop loss index, which is not considered in Cissokho and Kulik (2021). The resulting asymptotic variance of the runs estimator cannot be compared directly with that of the blocks estimators.

Structure of the thesis

This dissertation is structured as follows: In Chapter 2 the fundamentals of extreme value theory for time series are introduced, among others regular variation, the spectral tail process and the time change formula. Subsequently, in Chapter 3 the abstract setting for the derivation of uniform central limit theorems for disjoint and sliding blocks statistics is introduced. This chapter is also an essential preparation for the asymptotic analysis of $\hat{p}_{n,A}$ later on. The limit theorem for sliding blocks statistics is proved in this chapter and a general comparison of disjoint vs. sliding blocks statistics in the abstract POT setting is presented. In order to provide application examples for the abstract setting and the sliding blocks limit theorem, the asymptotic normality of estimators for cluster indexes is discussed in Chapter 4. In particular, the extremal index and the stop-loss index are analyzed, with new insights about the estimators for the extremal index. Here, the advantage of a unified framework for the asymptotic analysis is presented. The new concept of the projection based estimator $\hat{p}_{n,A}$ for the distribution of the spectral tail process is motivated in the Chapter 5. In this chapter, the asymptotic normality of this estimator is established, first with known α and then with a more sophisticated technical proof for estimated α. The chapter also contains the example of stochastic recurrence equations, for which all assumed conditions are fulfilled. Finally, the finite sample performance of the new estimator for the distribution of the spectral tail process is considered in a simulation study. Chapter 6 finalizes this thesis with a brief outlook on open research questions. All proofs are deferred to the end of each chapter.

Chapter 2

Extreme value theory for time series

In the examples given in the beginning of the introduction, there are not only independent observations. Rather, the observed data may show dependencies over time. In order to describe extremes of such dependent data in a mathematically precise way, extreme value theory for time series was developed. The modern approach to the extreme value theoretical consideration of $\mathbb{R}^d$-valued stationary time series has been introduced by Basrak and Segers (2009). As for independent data, the standard assumption for time series is the regular variation of all finite dimensional distributions. Under this condition Basrak and Segers (2009) showed the existence of a so-called spectral tail process $(\Theta_t)_{t\in\mathbb{Z}}$. This process describes the extreme behavior of the time series $(X_t)_{t\in\mathbb{Z}}$, assuming that an extreme event occurred at the initial time 0. Therefore, $(\Theta_t)_{t\in\mathbb{Z}}$ can also be used for modeling the extreme ranges of the underlying time series. In preparation for the rest of this work, we will briefly introduce the basic concepts of regular varying random variables, regular varying time series and recall some basic properties.

2.1 Regular varying time series

We start this chapter with the concept of regular variation in the univariate case.

Definition 2.1.1 (Regular variation)**.** *(i) A measurable function $f : (x_0, \infty) \to \mathbb{R}$, $x_0 \in \mathbb{R}$, is called **(univariate) regularly varying** if there exists some $\rho \in \mathbb{R}$ such that*

$$\lim_{t\to\infty} \frac{f(xt)}{f(t)} = x^\rho \qquad \forall x > 0.$$

The function f is called slowly varying if $\rho = 0$.
(ii) A measurable function $f : (x_0, x_1) \to \mathbb{R}$, $x_0, x_1 \in \mathbb{R}$, is called regularly varying in x_0 (x_1 resp.) if there exists some $\rho \in \mathbb{R}$ such that

$$\lim_{t\to 0} \frac{f(x_0 + xt)}{f(x_0 + t)} = x^\rho \qquad \left(\lim_{t\to 0} \frac{f(x_1 - xt)}{f(x_1 - t)} = x^\rho \ \textit{resp.}\right) \qquad \forall x > 0,$$

(iii) A real valued random variable X with distribution function F is called (univariate) ***regularly varying with index*** *ρ, if the survival function $1-F$ is regularly varying, i.e.*

$$\lim_{t\to\infty}\frac{P(X>xt)}{P(X>t)}=\lim_{t\to\infty}\frac{1-F(xt)}{1-F(t)}=x^{\rho}\qquad\forall x>0.$$

A regularly varying function f has the tail behavior of a power function, i.e. there exist $\rho\in\mathbb{R}$ and a slowly varying function l such that $f(x)=x^{\rho}l(x)$ for all $x\in(x_0,\infty)$ (Bingham et al. (1989), Theorem 1.4.1). In particular, $f(x)\to 0$ for $x\to\infty$ if $\rho<0$ and $f(x)\to\infty$ for $x\to\infty$ if $\rho>0$.

There is a broad theory about properties of regularly varying functions or random variables, some of it will be used in this thesis. For an overview, we refer to e.g. Bingham et al. (1989) and De Haan and Ferreira (2006). Here, we just mention the Potter bounds (Bingham et al. (1989), Theorem 1.5.6):

Theorem 2.1.2. *If $f:(x_0,\infty)\to\mathbb{R}$ is regularly varying with index ρ, then for all $\varepsilon>0$ there exists some $x_\varepsilon>x_0$ such that for all $t>x_\varepsilon$ and x with $xt>x_\varepsilon$*

$$(1-\varepsilon)x^{\rho}\min(x^{\varepsilon},x^{-\varepsilon})\le\frac{f(xt)}{f(t)}\le(1+\varepsilon)x^{\rho}\max(x^{\varepsilon},x^{-\varepsilon}).$$

More central for this thesis is the concept of multivariate regular variation for random vectors $X=(X_1,...,X_d)\in\mathbb{R}^d$.

Definition 2.1.3 (Multivariate regular variation)**.** *An $\mathbb{R}^d$-valued random vector X is called* ***(multivariate) regularly varying*** *if there exists a non-degenerate measure μ on $(\mathbb{R}^d\backslash\{0\},\mathcal{B}(\mathbb{R}^d\backslash\{0\}))$ with $\mu(A)<\infty$ for all $A\in\mathcal{B}(\mathbb{R}^d\backslash\{0\})$ bounded away from 0 and*

$$\lim_{t\to\infty}\frac{P(X\in tA)}{P(\|X\|>t)}=\mu(A)$$

for all $A\in\mathcal{B}(\mathbb{R}^d\backslash\{0\})$ bounded away from 0 with $\mu(\partial A)=0$.

Here, ∂B denotes the topological boundary of the set B. The measure μ is non degenerate if $\mu(\{x\}^C)>0$ for all $x\in\mathbb{R}^d\backslash\{0\}$, i.e. if it is not concentrated in a single point. The norm $\|\cdot\|$ is arbitrary: if such a limit measure μ exists for one norm, then there exists a limit measure for each norm and the measures are equal up to a multiplicative constant. The limit measure μ is homogeneous with index $-\alpha$ for some $\alpha>0$, i.e. $\mu(tA)=t^{-\alpha}\mu(A)$ for all $A\in\mathcal{B}(\mathbb{R}^d\backslash\{0\})$ bounded away from 0 and $t>0$ (Resnick, 1987).

Definition 2.1.4 (Index of regular variation)**.** *If X is regularly varying with the $-\alpha$-homogeneous limit measure μ, then the parameter $\alpha>0$ is called the* ***index of regular variation*** *of X.*

By choosing $A = \{x \in \mathbb{R}^d \mid \|x\| > r\}$ for $r > 0$, the multivariate regular variation of X directly implies the univariate regular variation of $\|X\|$ with index $-\alpha$.
With the help of a polar transformation, a spectral decomposition can be specified for the limit measure of the regular variation. Define $\mathbb{S}^{d-1} := \{x \in \mathbb{R}^d : \|x\| = 1\}$ as the unit sphere with respect to the norm $\|\cdot\|$. Then, one can show that X is multivariate regularly varying if and only if there exists a probability measure Φ on $(\mathbb{S}^{d-1}, \mathcal{B}(\mathbb{S}^{d-1}))$ such that

$$\lim_{t\to\infty} \frac{P\left(\|X\| > rt, X/\|X\| \in A\right)}{P(\|X\| > t)} = r^{-\alpha}\Phi(A)$$

for all $r > 0$ and $A \in \mathcal{B}(\mathbb{S}^{d-1})$ with $\Phi(\partial A) = 0$. In particular, $\Phi(A) = \mu(\{x \in \mathbb{R}^d \mid \|x\| \geq 1, x/\|x\| \in A\})$ and Φ is called spectral measure of X. Moreover, $\nu_\alpha((r, \infty)) := r^{-\alpha}$ defines a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Thus, this spectral representation implies that the stochastic behavior of the heaviness of the tail and the dependency in the extremes can be considered independently. Indeed, the heaviness of the tail is described by α and the dependence is modeled by Φ.
An overview about further properties of regularly varying functions and random variables can be found e.g. in the basic references Bingham et al. (1989) and De Haan and Ferreira (2006). Further properties and methods to model heavy tails can also be found in Resnick (2007). Here, we will continue with the extension of the concept of regular variation for strictly stationary time series.

Definition 2.1.5 (Regular varying time series)**.** *A $\mathbb{R}^d$-valued time series $(X_t)_{t\in\mathbb{Z}}$ is **regularly varying with index** α if all finite dimensional marginal distributions (fidis) $(X_s, ..., X_t)$, $s, t \in \mathbb{Z}$, $s \leq t$ are multivariate regularly varying with index α. The value α is called the index of regular variation of $(X_t)_{t\in\mathbb{Z}}$.*

In particular, regular variation of the time series $(X_t)_{t\in\mathbb{Z}}$ implies regular variation of X_0 and thereby univariate regular variation of $\|X_0\|$ with index $-\alpha$.
In this work we will consider only (strictly) stationary time series $(X_t)_{t\in\mathbb{Z}}$. As mentioned above, the extreme behavior of a time series can be described in some sense by the spectral tail process. This process is defined in the following definition.

Definition 2.1.6 (Tail process and spectral tail process)**.** *Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary $\mathbb{R}^d$-valued time series. If there exists a non degenerate time series $(Y_t)_{t\in\mathbb{Z}}$ with*

$$\mathcal{L}\left(\left(\frac{X_s}{u_n}, ..., \frac{X_t}{u_n}\right) \Big| \|X_0\| > u_n\right) \xrightarrow[u_n\to\infty]{w} \mathcal{L}(Y_s, ..., Y_t), \tag{2.1.1}$$

*for all $s, t \in \mathbb{Z}$, $s \leq t$, then $(Y_t)_{t\in\mathbb{Z}}$ is called the **tail process** of $(X_t)_{t\in\mathbb{Z}}$. If there exists a time series $(\Theta_t)_{t\in\mathbb{Z}}$ with*

$$\mathcal{L}\left(\left(\frac{X_s}{\|X_0\|}, ..., \frac{X_t}{\|X_0\|}\right) \Big| \|X_0\| > u_n\right) \xrightarrow[u_n\to\infty]{w} \mathcal{L}(\Theta_s, ..., \Theta_t),$$

for all $s, t \in \mathbb{Z}$, $s \leq t$, then $(\Theta_t)_{t\in\mathbb{Z}}$ is called the ***spectral tail process*** *(or in short* ***spectral process****) of $(X_t)_{t\in\mathbb{Z}}$. The process $(\Theta_t)_{t\in\mathbb{N}_0}$ is called forward spectral tail process and $(\Theta_{-t})_{t\in\mathbb{N}_0}$ is called backward spectral tail process.*

By this definition it is obvious that the tail process and the spectral tail process describe asymptotically the extreme behavior of the underlying time series $(X_t)_{t\in\mathbb{Z}}$. Therefore, these processes are important tools for the extreme value analysis for time series. The spectral tail process contains the information about the dependence of the extremes while $(Y_t)_{t\in\mathbb{Z}}$ also includes information about the heaviness of the tails. There is a close connection between $(Y_t)_{t\in\mathbb{Z}}$ and $(\Theta_t)_{t\in\mathbb{Z}}$ which will be investigated after the next theorem. Before that, we state a criterion for the existence of the tail process. Basrak and Segers (2009) have proved that $(Y_t)_{t\in\mathbb{Z}}$ exists if and only if the time series $(X_t)_{t\in\mathbb{Z}}$ is regularly varying.

Theorem 2.1.7. *Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary $\mathbb{R}^d$-valued time series and let $\alpha \in (0,\infty)$. The following statements are equivalent:*

(i) $(X_t)_{t\in\mathbb{Z}}$ is regularly varying with index α.

(ii) There exists an $\mathbb{R}^d$-valued process $(Y_t)_{t\in\mathbb{Z}}$ with $P(\|Y_0\| > y) = y^{-\alpha}$ for $y \geq 1$ (i.e. $\|Y_0\|$ is Par(α) distributed) such that for all $t \in \mathbb{N}$

$$\mathcal{L}\left(\left(\frac{X_0}{u_n}, ..., \frac{X_t}{u_n}\right) \middle| \|X_0\| > u_n\right) \xrightarrow[u_n\to\infty]{w} \mathcal{L}(Y_0, ..., Y_t).$$

(iii) There exists an $\mathbb{R}^d$-valued process $(Y_t)_{t\in\mathbb{Z}}$ with $P(\|Y_0\| > y) = y^{-\alpha}$ for $y \geq 1$ such that for all $s, t \in \mathbb{Z}$ with $s \leq t$

$$\mathcal{L}\left(\left(\frac{X_s}{u_n}, ..., \frac{X_t}{u_n}\right) \middle| \|X_0\| > u_n\right) \xrightarrow[u_n\to\infty]{w} \mathcal{L}(Y_s, ..., Y_t).$$

Furthermore, Basrak and Segers (2009) have shown that the regular variation of $(X_t)_{t\in\mathbb{Z}}$ is also equivalent to the existence of the spectral tail process. Moreover, tail process and spectral tail process are closely related via $(Y_t)_{t\in\mathbb{Z}} \stackrel{d}{=} Y(\Theta_t)_{t\in\mathbb{Z}}$ for an independent Par(α)-distributed random variable Y. Here, Par(α) stands for the Pareto distribution with parameter α, i.e. $P(Y > y) = y^{-\alpha}$ for $y \in [1,\infty)$.

Theorem 2.1.8. *Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary $\mathbb{R}^d$-valued time series and let the function $x \mapsto P(\|X_0\| > x)$ be regularly varying with index $-\alpha$ for some $\alpha \in (0,\infty)$. The following statements are equivalent*

(i) $(X_t)_{t\in\mathbb{Z}}$ is regularly varying with index α.

(ii) There exists an $\mathbb{R}^d$-valued process $(\Theta_t)_{t\in\mathbb{Z}}$ such that for all $t \in \mathbb{N}$

$$\mathcal{L}\left(\left(\frac{X_0}{\|X_0\|}, ..., \frac{X_t}{\|X_0\|}\right) \middle| \|X_0\| > u_n\right) \xrightarrow[u_n\to\infty]{w} \mathcal{L}(\Theta_0, ..., \Theta_t).$$

(iii) There exists an $\mathbb{R}^d$-valued process $(\Theta_t)_{t\in\mathbb{Z}}$ such that for all $s,t \in \mathbb{Z}$ with $s \le t$

$$\mathcal{L}\left(\left(\frac{X_s}{\|X_0\|}, ..., \frac{X_t}{\|X_0\|}\right) \Big| \|X_0\| > u_n\right) \xrightarrow[u_n\to\infty]{w} \mathcal{L}(\Theta_s, ..., \Theta_t).$$

If (ii) or (iii) is satisfied, then the tail process $(Y_t)_{t\in\mathbb{Z}}$ of $(X_t)_{t\in\mathbb{Z}}$ is given by $(Y_t)_{t\in\mathbb{Z}} \stackrel{d}{=} (Y\Theta_t)_{t\in\mathbb{Z}}$ for all $t \in \mathbb{Z}$, where the random variable Y is Par(α) distributed and independent of $(\Theta_t)_{t\in\mathbb{Z}}$.

By this theorem, it holds that

$$(Y_t)_{t\in\mathbb{Z}} \stackrel{d}{=} (\|Y_0\|\Theta_t)_{t\in\mathbb{Z}} \qquad \text{and} \qquad (\Theta_t)_{t\in\mathbb{Z}} \stackrel{d}{=} (Y_t/\|Y_0\|)_{t\in\mathbb{Z}}\,.$$

Due to this decomposition the random variable $\|Y_0\|$ and the parameter α describe the heaviness of the tail while the spectral tail process $(\Theta_t)_{t\in\mathbb{Z}}$ describes the serial dependence of the extremes of the time series $(X_t)_{t\in\mathbb{Z}}$. Note that $\|\Theta_0\| = 1$ a.s., i.e. Θ_0 models only the extremal dependence but $\|\Theta_t\| \neq 1$ for $t \neq 1$ is possible, i.e. Θ_t might includes some information about the heaviness of the tail of X_t given that X_0 is extreme. In some sense, the decomposition of Y_t in $\|Y_0\|$ and Θ_t is comparable to the spectral decomposition for the multivariate regular variation above, but for $t \neq 0$ it is not the same.
The tail process and spectral tail process are defined by the weak limit of all standardized finite stretches $(X_s, ..., X_t)$, for $s \le t \in \mathbb{Z}$. Under an additional assumption, one can show that the tail process is also the weak limit of a standardized growing segment $(X_{-r_n}, ..., X_{r_n})$ for some suitable increasing sequence $r_n \to \infty$. The condition needed for this is the well known anti-clustering condition (AC) introduced by Davis and Hsing (1995), but we state this condition with a sequence of thresholds u_n satisfying $nP(\|X_0\| > u_n) \to \infty$, while originally $nP(\|X_0\| > u_n) \to \tau > 0$ was used, i.e. we use a smaller sequence u_n. The anti-clustering condition (or, more precisely, finite mean cluster size condition, since it allows clusters of extremes, but the mean of the size of these clusters may only be finite) is given by

(AC)

$$\lim_{m\to\infty} \limsup_{n\to\infty} P\Big(\max_{m\le|t|\le r_n} \|X_t\| > u_n c \,\Big|\, \|X_0\| > u_n c\Big) = 0 \tag{2.1.2}$$

for all $c \in (0,\infty)$, for a fixed sequence u_n with $nv_n := nP(\|X_0\| > u_n) \to \infty$ and some sequence $r_n \to \infty$ with $r_n v_n \to 0$.

The convergence considered in the following lemma is understood as weak convergence on the sequence space $l_\alpha \times l_\alpha$ equipped with the supremum norm, where

$$l_\alpha := \Big\{(x_t)_{t\in\mathbb{Z}} \in (\mathbb{R}^d)^{\mathbb{Z}} \,\Big|\, \sum_{t\in\mathbb{Z}} \|x_t\|^\alpha < \infty\Big\}.$$

For arbitrary $n \in \mathbb{N}$ the spaces $(\mathbb{R}^d)^{2n+1}$ is embedded in l_α by the mapping $(\mathbb{R}^d)^{2n+1} \ni$

$(z_t)_{|t|\leq n} \mapsto (z_t)_{t\in\mathbb{Z}} \in l_\alpha$ with $z_t = 0$ for $|t| \geq n$. Note that (AC) ensures that the realizations of the spectral tail process a.s. belongs to l_α (see next section or Remark 2.3 of Janßen (2019)).

Lemma 2.1.9. *Suppose $(X_t)_{t\in\mathbb{Z}}$ is a regularly varying time series which satisfies (AC), $(r_n)_{n\in\mathbb{N}} \subset \mathbb{N}$ with $r_n \to \infty$, $r_n v_n \to 0$ and fix $j \in \mathbb{Z}$. If either (AC) holds for r_n replaced by $r'_n = r_n + j$ or if $v_n^{-1}\beta^X_{n,r_n} \to 0$, then*

$$\mathcal{L}\left((u_n c)^{-1}(X_t)_{|t|\leq r_n}, (u_n c)^{-1}(X_{t+j})_{|t|\leq r_n} \,\middle|\, \|X_0\| > u_n c\right) \xrightarrow{w} \mathcal{L}\left((Y_t)_{t\in\mathbb{Z}}, (Y_{t+j})_{t\in\mathbb{Z}}\right).$$

Here β^X_{n,r_n} denotes the β-mixing coefficient of $(X_t)_{t\in\mathbb{Z}}$. For $j = 0$ the second assumption of (AC) for r'_n is trivially fulfilled. The proof of this lemma is given below in Section 2.3.

2.2 Time change formula and RS-transformation

So far, we introduced the basic concept of regular variation and defined the spectral tail process as a weak limit. The distribution of the spectral tail process shall be estimated in Chapter 5.

Next, we consider some properties of the spectral tail process. In general the stationarity of $(X_t)_{t\in\mathbb{Z}}$ does not imply the stationarity of $(\Theta_t)_{t\in\mathbb{Z}}$. This is due to the special role of the time point 0. Observe that $\left\|X_0/\|X_0\|\right\| = 1$ holds and, therefore, $\|\Theta_0\| = 1$ a.s. but, in general, $\|\Theta_1\|$ is not constant 1, i.e. $(\Theta_t)_{t\in\mathbb{Z}}$ cannot be stationary. However, the stationarity of $(X_t)_{t\in\mathbb{Z}}$ implies a different structural property of $(\Theta_t)_{t\in\mathbb{Z}}$, which is formalized by the so called *time change formula* (TCF). This property was discovered by Basrak and Segers (2009).

Theorem 2.2.1. *Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary, regularly varying time series and let $(Y_t)_{t\in\mathbb{Z}}$ and $(\Theta_t)_{t\in\mathbb{Z}}$ be the corresponding tail process and spectral tail process, respectively. Then,*

(i) $\|Y_0\|$ is independent of $(\Theta_t)_{t\in\mathbb{Z}}$.

(ii) For all $i, s, t \in \mathbb{Z}$ with $s \leq 0 \leq t$ and for all continuous and bounded functions $g : (\mathbb{R}^d)^{t-s+1} \to \mathbb{R}$ with $g(y_s, ..., y_t) = 0$ if $y_0 = 0$ it holds that

$$E[g(Y_{s-i}, ..., Y_{t-i})] = \int_0^\infty E\left[g(r\Theta_s, ..., r\Theta_t)\mathbb{1}_{\{r\|\Theta_i\|>1\}}\right]\alpha r^{-\alpha-1}dr.$$

(iii) For all $i, s, t \in \mathbb{Z}$ with $s \leq 0 \leq t$ and for all continuous and bounded functions $f : (\mathbb{R}^d)^{t-s+1} \to \mathbb{R}$ with $f(y_s, ..., y_t) = 0$ if $y_0 = 0$ it holds that

$$E[f(\Theta_{s-i}, ..., \Theta_{t-i})] = E\left[f\left(\frac{\Theta_s}{\|\Theta_i\|}, ..., \frac{\Theta_t}{\|\Theta_i\|}\right)\|\Theta_i\|^\alpha\right]. \tag{2.2.1}$$

If (iii) in the theorem is satisfied, then (2.2.1) holds for all measurable and bounded functions $f : (\mathbb{R}^d)^{t-s+1} \to \mathbb{R}$ with $f(y_s, ..., y_t) = 0$ if $y_0 = 0$, since each measurable function can be approximated by continuous functions. The property in part (iii) is very central and is called the time change formula.

Definition 2.2.2 (TCF)**.** *An $\mathbb{R}^d$-valued time series $(\Theta_t)_{t\in\mathbb{Z}}$ possesses the property **TCF** (Time Change Formula), if $P(\|\Theta_0\| = 1) = 1$ and the **time change formula** (2.2.1) is fulfilled, i.e. there exists some $\alpha > 0$ such that*

$$E[f(\Theta_{s-i}, ..., \Theta_{t-i})] = E\Big[f\Big(\frac{\Theta_s}{\|\Theta_i\|}, ..., \frac{\Theta_t}{\|\Theta_i\|}\Big)\mathbb{1}_{\{\|\Theta_i\|>0\}}\|\Theta_i\|^\alpha\Big] \tag{2.2.2}$$

holds for all $s \leq 0 \leq t$, $i \in \mathbb{Z}$ and for all bounded and measurable functions $f : (\mathbb{R}^d)^{t-s+1} \to \mathbb{R}$ with $f(\theta_s, ..., \theta_t) = 0$ if $\theta_0 = 0$.

By the theorems above, every spectral tail process possesses the property (TCF). Since the finite dimensional marginal distributions define the distribution of a stochastic process uniquely, (2.2.2) can be generalized to

$$E[f((\Theta_{t-i})_{t\in\mathbb{Z}})] = E\Big[f\Big(\Big(\frac{\Theta_t}{\|\Theta_i\|}\Big)_{t\in\mathbb{Z}}\Big)\|\Theta_i\|^\alpha\Big].$$

for all $i \in \mathbb{Z}$ and for all bounded and $\mathcal{B}((\mathbb{R}^d)^{\mathbb{Z}})$-$\mathcal{B}(\mathbb{R})$-measurable functions $f : (\mathbb{R}^d)^{\mathbb{Z}} \to \mathbb{R}$ with $f((\theta_t)_{t\in\mathbb{Z}}) = 0$ if $\theta_0 = 0$.

The time change formula is not only a property of a spectral tail process. Rather, Janßen (2019), Theorem 4.2, and Planinić and Soulier (2018), Theorem 5.1, have shown, that each process satisfying the property (TCF) is already the spectral tail process of some max-stable time series. Thus, the property (TCF) characterizes the class of all admissible spectral tail processes.

Until recently, no general interpretation of the time change formula from Definition 2.2.2 was known in the literature. A first possible interpretation was given by Janßen (2019) for one (quite general) case. She introduced an equivalent representation of the structural properties implied by the time change formula under the following summability condition (SC) that is often satisfied as we will shortly see.

(SC) The $\mathbb{R}^d$-valued time series $(\Theta_t)_{t\in\mathbb{Z}}$ satisfies for some $\alpha > 0$ that

$$0 < \sum_{t\in\mathbb{Z}} \|\Theta_t\|^\alpha < \infty \qquad \text{a.s.} \tag{2.2.3}$$

This condition depends on α, which is the index of regular variation whenever a spectral tail process is considered. The first inequality in (2.2.3) is always satisfied for a spectral tail process, since $\|\Theta_0\| = 1$ a.s. If the summability condition is fulfilled, we will use the following notation:

$$\|z\|_\alpha := \|(z_t)_{t\in\mathbb{Z}}\|_\alpha := \Big(\sum_{t\in\mathbb{Z}} \|z_t\|^\alpha\Big)^{1/\alpha} \tag{2.2.4}$$

for all $(z_t)_{t\in\mathbb{Z}} \in (\mathbb{R}^d)^{\mathbb{Z}}$ and $\alpha > 0$. Note that $\|\cdot\|_\alpha$ is a norm on $(\mathbb{R}^d)^{\mathbb{Z}}$ for $\alpha \geq 1$ but only a quasinorm for $\alpha < 1$.
Furthermore, denote $\|\Theta^*\| := \sup_{t\in\mathbb{Z}} \|\Theta_t\|$ and $T^* := T^*((\Theta_t)_{t\in\mathbb{Z}}) := \inf\{t \in \mathbb{Z} : \|\Theta_t\| = \|\Theta^*\|\}$. Janßen (2019) proved the following equivalent conditions for (SC).

Lemma 2.2.3. *Assume that the $\mathbb{R}^d$-valued time series $(\Theta_t)_{t\in\mathbb{Z}}$ satisfies the property (TCF) with some $\alpha > 0$. Then the following statements are equivalent*

(i) $\sum_{t\in\mathbb{Z}} \|\Theta_t\|^\alpha < \infty$ *a.s. (i.e. condition (SC) is fulfilled.)*

(ii) $\|\Theta_t\| \to 0$ *a.s. for* $|t| \to \infty$

(iii) $P(T^* \in \mathbb{Z}) = 1$

Condition (SC) is not very restrictive. Basrak and Segers (2009), Proposition 4.2, have shown that this condition is satisfied for a spectral tail process if the underlying time series satisfies Condition (AC). This condition in turn is fulfilled e.g. for stochastic volatility models, ARMA models, max-moving average processes (Mikosch and Zhao, 2014) and GARCH models (Basrak et al., 2002). Recently, Kulik et al. (2019) showed that the anti-clustering condition is also met for the broad class of stationary geometrically ergodic Markov chains.
Under Condition (SC), Janßen (2019) proved the equivalence of the property (TCF) and the invariance under the so-called RS-transformation of the time series.

Definition 2.2.4 (RS-transformation)**.** *Consider a time series $(\Theta_t)_{t\in\mathbb{Z}}$ which satisfies the condition (SC). The* ***RS-transformation*** *$(\Theta_t^{RS})_{t\in\mathbb{Z}}$ of $(\Theta_t)_{t\in\mathbb{Z}}$ is defined by*

$$(\Theta_t^{RS})_{t\in\mathbb{Z}} \stackrel{d}{=} \left(\frac{\Theta_{t+K(\Theta)}}{\|\Theta_{K(\Theta)}\|}\right)_{t\in\mathbb{Z}}, \tag{2.2.5}$$

where $K(\Theta) = K((\Theta_t)_{t\in\mathbb{Z}})$ is a $\mathbb{Z}$-valued random variable with conditional density

$$P(K(\Theta) = k \mid (\Theta_t)_{t\in\mathbb{Z}}) = \frac{\|\Theta_k\|^\alpha}{\sum_{t\in\mathbb{Z}} \|\Theta_t\|^\alpha} = \frac{\|\Theta_k\|^\alpha}{\|\Theta\|_\alpha^\alpha},$$

for all $k \in \mathbb{Z}$.

This definition and taking iterated expectations directly leads to

$$P\left((\Theta_t^{RS})_{t\in\mathbb{Z}} \in B\right) = E\Big[\sum_{k\in\mathbb{Z}} \frac{\|\Theta_k\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbb{1}_B\Big(\frac{(\Theta_{t+k})_{t\in\mathbb{Z}}}{\|\Theta_k\|}\Big)\Big]$$

for all cylinder sets B in $(\mathbb{R}^d)^{\mathbb{Z}}$. Alternatively, one could use these probabilities to fully characterize the distribution of the RS-transformation $(\Theta_t^{RS})_{t\in\mathbb{Z}}$. The RS-transform is only defined in terms of its distribution.

Theorem 2.2.5. *Let $(\Theta_t)_{t\in\mathbb{Z}}$ be an $\mathbb{R}^d$-valued time series which satisfies condition (SC). Then, $(\Theta_t)_{t\in\mathbb{Z}}$ has the property (TCF) if and only if*

$$(\Theta_t^{RS})_{t\in\mathbb{Z}} \stackrel{d}{=} (\Theta_t)_{t\in\mathbb{Z}}.$$

This invariance under the RS-transformation allows for some interpretation of the time change formula: the time change formula corresponds to a random shift and rescaling of the time series, which does not affect the distribution. The random shift is proportional to the magnitude of the time series at the respective lag.
Since the property (TCF) characterizes the class of admissible spectral tail processes, Theorem 2.2.5 now specifies this class under the additional assumption (SC), namely as all processes with distributions invariant under the RS-transformation.
The RS-transformation together with the previous theorem will be the basis for the deduction of the new projection based estimator $\hat{p}_{n,A}$ for the distribution of a spectral tail process in Chapter 5. With this, we end this short chapter about the basics for this thesis, in the next section only one proof is added. For more detailed information we refer to the references cited in this chapter. In the next chapter, we start with the development of new abstract limit theorems for estimators of rare events.

2.3 Proofs for Section 2.1

In this section, only the proof of the technical Lemma 2.1.9 is given. All other proofs for lemmas and theorems in this chapter can be found in the cited references. The idea of the proof of Lemma 2.1.9 is similar to the idea of the proof of Theorem 2.2 of Basrak et al. (2018), in particular it uses the same truncation arguments.

Proof of Lemma 2.1.9. With the Portemanteau Theorem (Billingsley (1968), Theorem 2.1) it suffices to show

$$E\left[g\Big((u_nc)^{-1}(X_t)_{|t|\le r_n},(u_nc)^{-1}(X_{t+j})_{|t|\le r_n}\Big)\,\Big|\,\|X_0\|>u_nc\right]\to E\left[g\left((Y_t)_{t\in\mathbb{Z}},(Y_{t+j})_{t\in\mathbb{Z}}\right)\right], \tag{2.3.1}$$

for all non-negative, bounded and uniformly continuous functions g on $l_\alpha\times l_\alpha$.
For $x\in l_\alpha$, we define the truncation at the level ξ by $x_\xi=(x_t\mathbb{1}_{\{\|x_t\|>\xi\}})_{t\in\mathbb{Z}}$. Then, it obviously holds that $\|x-x_\xi\|_\infty\le\xi$. In addition to g, we define the function g_ξ by $g_\xi(x,y):=g(x_\xi,y_\xi)$. Since g is uniformly continuous, for all $\varepsilon>0$ there exists a $\delta>0$ such that $\|(x,y)-(x',y')\|_\infty<\delta$ implies $|g(x,y)-g(x',y')|\le\varepsilon$. Hence, $\|g_\delta-g\|_\infty\le\varepsilon$ and it suffices to show (2.3.1) for g_ξ and all $\xi\in(0,1)$.
Fix $m\in\mathbb{N}$. For a sufficiently large $n\in\mathbb{N}$ such that $r_n>m$, one has

$$\left|E\left[g_\xi\left((u_nc)^{-1}(X_t)_{|t|\le r_n},(u_nc)^{-1}(X_{t+j})_{|t|\le r_n}\right)\mid\|X_0\|>u_nc\right]\right.$$

$$
\begin{aligned}
&\qquad - E\Big[g_\xi\Big((u_nc)^{-1}(X_t)_{|t|\le m},(u_nc)^{-1}(X_{t+j})_{|t|\le m}\Big) \mid \|X_0\|>u_nc\Big]\Big| \\
&= \frac{1}{P(\|X_0\|>u_nc)}\Big|E\Big[g_\xi\Big((u_nc)^{-1}(X_t)_{|t|\le r_n},(u_nc)^{-1}(X_{t+j})_{|t|\le r_n}\Big)\mathbb{1}_{\{\|X_0\|>u_nc\}}\Big] \\
&\qquad - E\Big[g_\xi\Big((u_nc)^{-1}(X_t)_{|t|\le m},(u_nc)^{-1}(X_{t+j})_{|t|\le m}\Big)\mathbb{1}_{\{\|X_0\|>u_nc\}}\Big]\Big| \\
&\le \frac{1}{P(\|X_0\|>u_nc)}\|g\|_\infty P\Big(\max_{m<|t|\le r_n+j}\|X_t\|>u_nc\xi,\|X_0\|>u_nc\Big) \\
&= \|g\|_\infty P\Big(\max_{m<|t|\le r_n+j}\|X_t\|>u_nc\xi \mid \|X_0\|>u_nc\Big).
\end{aligned}
$$

We obtain $\lim_{m\to\infty}\limsup_{n\to\infty}P\Big(\max_{m\le|t|\le r_n'}\|X_t\|>u_nc\xi\,\Big|\,\|X_0\|>u_nc\xi\Big)=0$ either since (AC) holds for $r_n'=r_n+j$ or since for $v_{n,c\xi}:=P(\|X_0\|>u_nc\xi)$ we have

$$
\begin{aligned}
&P\Big(\max_{m\le|t|\le r_n'}\|X_t\|>u_nc\xi\,\Big|\,\|X_0\|>u_nc\xi\Big) \\
&\le P\Big(\max_{m\le|t|\le r_n}\|X_t\|>u_nc\xi\,\Big|\,\|X_0\|>u_nc\xi\Big)+\sum_{r_n<|t|\le r_n+j}P\Big(\|X_t\|>u_nc\xi\,\Big|\,\|X_0\|>u_nc\xi\Big) \\
&\le P\Big(\max_{m\le|t|\le r_n}\|X_t\|>u_nc\xi\,\Big|\,\|X_0\|>u_nc\xi\Big)+2jv_{n,c\xi}+\frac{2j}{v_{n,c\xi}}\beta^X_{n,r_n},
\end{aligned}
$$

which converges to 0 as $n\to\infty$ and then $m\to\infty$ due to (2.1.2) and the mixing assumption. Thus, we conclude with $r_n'=r_n+j$

$$
\begin{aligned}
&\lim_{m\to\infty}\limsup_{n\to\infty}P\Big(\max_{m\le|t|\le r_n'}\|X_t\|>u_nc\xi\,\Big|\,\|X_0\|>u_nc\Big) \\
&\le \lim_{m\to\infty}\limsup_{n\to\infty}P\Big(\max_{m\le|t|\le r_n'}\|X_t\|>u_nc\xi\,\Big|\,\|X_0\|>u_nc\xi\Big)\frac{P(\|X_0\|>u_nc\xi)}{P(\|X_0\|>u_nc)}=0,
\end{aligned}
$$

where the last fraction converges to $\xi^{-\alpha}$, due to the regular variation of $(X_t)_{t\in\mathbb{Z}}$. Therefore,

$$
\begin{aligned}
\lim_{m\to\infty}\limsup_{n\to\infty}&\Big|E\Big[g_\xi\Big((u_nc)^{-1}(X_t)_{|t|\le r_n},(u_nc)^{-1}(X_{t+j})_{|t|\le r_n}\Big) \mid \|X_0\|>u_nc\Big] \\
&- E\Big[g_\xi\Big((u_nc)^{-1}(X_t)_{|t|\le m},(u_nc)^{-1}(X_{t+j})_{|t|\le m}\Big) \mid \|X_0\|>u_nc\Big]\Big|=0.
\end{aligned}
$$

Moreover, $\|Y_i\|\to 0$ a.s. for $|i|\to\infty$, due to (AC) (Basrak and Segers (2009), Proposition 4.2), which according to Lemma 2.2.3 implies $\sum_{h\in\mathbb{Z}}\|Y_h\|^\alpha<\infty$ a.s. Therefore, it follows that Y can only have a finite number of coordinates with a norm larger than ξ. Thus, $g_\xi((Y_t)_{|t|\le m},(Y_{t+j})_{|t|\le m})=g_\xi((Y_t)_{t\in\mathbb{Z}},(Y_{t+j})_{t\in\mathbb{Z}})$ for a random (depending on Y) and sufficiently large $m\in\mathbb{N}$. All in all, with the definition of the tail process, we conclude

$$
\begin{aligned}
&\lim_{n\to\infty}E\Big[g_\xi\Big((u_nc)^{-1}(X_t)_{|t|\le r_n},(u_nc)^{-1}(X_{t+j})_{|t|\le r_n}\Big) \mid \|X_0\|>u_nc\Big] \\
&= \lim_{m\to\infty}\limsup_{n\to\infty}E\Big[g_\xi\Big((u_nc)^{-1}(X_t)_{|t|\le m},(u_nc)^{-1}(X_{t+j})_{|t|\le m}\Big) \mid \|X_0\|>u_nc\Big] \\
&= \lim_{m\to\infty}E\Big[g_\xi\Big((Y_t)_{|t|\le m},(Y_{t+j})_{|t|\le m}\Big)\Big]
\end{aligned}
$$

$$= E\left[g_\xi\left((Y_t)_{t\in\mathbb{Z}}, (Y_{t+j})_{t\in\mathbb{Z}}\right)\right].$$

This completes the proof. □

Chapter 3

Functional limit theorem in an abstract setting

Throughout this chapter, we consider statistics for the dependence structure of extremes of stationary time series in a fairly abstract way. In the peak-over-threshold (POT) approach for extreme value statistics, such statistics can usually be defined block-wise. To be more specific, assume that, starting from a stationary $\mathbb{R}^d$-valued time series $(X_t)_{1\le t\le n}$, random variables $X_{n,t}$, $1 \le t \le n$, $n \in \mathbb{N}$, are defined, that in some sense capture its extreme value behavior. The most common example is $X_{n,t} := (X_t/u_n)\mathbb{1}_{(u_n,\infty)}(\|X_t\|)$ for some threshold u_n and some norm $\|\cdot\|$ on $\mathbb{R}^d$, but for certain applications $X_{n,t}$ may also depend on observations in the neighborhood of extreme observations. One typically considers statistics $g(W_{n,j})$ of blocks

$$W_{n,j} := (X_{n,j}, \dots, X_{n,j+s_n-1}) \tag{3.0.1}$$

of (possibly increasing) length s_n, starting with the j-th random variable. Estimators and statistics of interest can then be defined in terms of averages of such block statistics.

Examples for such block statistics are the block-wise estimators for the extremal index and cluster indexes in Chapter 4 and the empirical extremogram analyzed by Davis and Mikosch (2009*b*). Further examples are the forward and backward estimators of the distribution of the spectral tail process of a regularly varying time series examined by Drees et al. (2015) and Davis et al. (2018), see also Chapter 5, and the estimator of the cluster size distribution proposed by Hsing (1991).

In these references, the estimators are defined as averages of disjoint blocks statistics $g(W_{n,is_n+1})$, $0 \le i \le \lfloor n/s_n \rfloor - 1$. However, one could also define the estimators via sliding blocks statistics $g(W_{n,i})$, $1 \le i \le n - s_n + 1$, of overlapping sliding blocks, see e.g. the sliding blocks estimator in Section 4.2.1. The main difference is that sliding blocks use much more data but these blocks have a larger dependence, so it is unclear which method is more advantageous.

Another example for a block-wise defined estimator is the projection based estimator

$\hat{p}_{n,A}$ defined in Chapter 5. In the previous examples, one could choose whether sliding or disjoint blocks should be used and, in the references, the estimators were defined in a natural way with disjoint blocks. In contrast, the projection based estimator $\hat{p}_{n,A}$ includes sliding blocks which automatically emerge due to the given motivation. In this example, it would be an extra step to define the disjoint blocks estimator. Later on, in Chapter 5 we want to analyze the asymptotics of the projection based estimator based on sliding blocks.

Drees and Rootzén (2010) provided a general framework to analyze the asymptotic behavior of statistics which are based on averages of functionals of disjoint blocks from an absolutely regular time series. The sufficient conditions for the convergence of the empirical process of cluster functionals established there proved to be a powerful tool for establishing asymptotic normality of a range of estimators; see, e.g. Drees (2015), Davis et al. (2018), and Drees and Knezevic (2020). However, the setting considered by Drees and Rootzén (2010) is too restrictive to accommodate empirical processes based on sliding blocks and this setting could not be used for the asymptotic analysis in Chapter 5.

Therefore, the aim of this chapter is to establish a more general limit theorem for empirical processes based on sliding blocks statistics. In fact, we will treat an even more general and abstract setting for block-wise measurable statistics. In the setting of this limit theorem, the consideration of sliding blocks is possible as well as the consideration of disjoint blocks statistics. The setting introduced in this chapter gives a unifying framework which, in particular, allows a comparison between the asymptotic results of sliding blocks statistics and their disjoint blocks counterparts.

First, we will introduce the general setting. In Section 3.1.1 the convergence of finite dimensional marginal distributions (fidis) is considered and in Section 3.1.2 conditions for the process convergence will be established. Section 3.2 is devoted to the limit theorem for the special case of the sliding blocks statistics. Since the general framework introduced here provides a unifying setting for the analysis of disjoint and sliding blocks estimators, a comparison of both can be given in Section 3.3. All proofs are deferred to Section 3.4.

The main results from this chapter have already been published in advance in Section 2 and Appendix A of Drees and Neblung (2021).

3.1 General abstract setting

The general setting introduced here builds basically on the ideas of Drees and Rootzén (2010). The purpose is a general limit theorem for empirical processes, which can be used for statistical extreme value theoretical problems. In particular, statistics based on sliding blocks should be covered.

We start with the introduction of some objects and variables needed to define our general empirical process for which we want to develop a limit theorem. We consider a triangular array $(X_{n,i})_{1\leq i\leq n, n\in\mathbb{N}}$ of row-wise stationary random variables with values in a polish and normed vector space $(E, \|\cdot\|)$. We denote the null element (the zero vector w.r.t. the vector addition) of E by 0.

Example. The classical example, which is often used in applications, is $E \subset \mathbb{R}^d$ and $d \in \mathbb{N}$. In this case, a typical choice is $X_{n,i} = X_i/u_n$ or $X_{n,i} = (X_i/u_n)\mathbb{1}_{\{\|X_i\|>u_n\}}$ for a threshold u_n and an $\mathbb{R}^d$-valued stationary time series $(X_t)_{t\in\mathbb{N}}$. However, the theory developed here also applies to more general $X_{n,i}$.
For applications with sliding blocks, it makes sense to use whole blocks of the form $\tilde{X}_{n,i} = (X_{n,i}, ..., X_{n,i+s_n-1})$ instead of single observations $X_{n,i}$, where $s_n \in \mathbb{N}$ denotes the length of the sliding blocks. In this case, we can consider $\tilde{X}_{n,i}$ instead of $X_{n,i}$, where $\tilde{X}_{n,i}$ takes values in the space

$$\tilde{E} := \Big\{(x_k)_{k\in\mathbb{N}} \in E^{\mathbb{N}} \mid \exists K \in \mathbb{N} \text{ with } \forall k > K : x_k = 0\Big\}.$$

Here, we understand $\tilde{X}_{n,i} = (\tilde{X}_{n,i,k})_{k\in\mathbb{N}}$ with $\tilde{X}_{n,i,k} = 0$ for $k > s_n$ as an element of $E^{\mathbb{N}}$. Because of the flexibility in the interpretation of the space E and since there is no real extra effort, in this chapter we will consider a general polish space E which is not necessarily a subset of the simple case $\mathbb{R}^d$. ◇

The independence of the random variables $(X_{n,i})_{1\le i\le n}$ is not required here. However, a β-mixing condition is imposed, such that we can use approximating methods with iid random variables.

Definition 3.1.1 (β-mixing)**.** *For a family of E-valued random variables $(\Gamma_{n,i})_{1\le i\le m_n}$, the β-**mixing coefficient** is defined as*

$$\beta^{\Gamma}_{n,k} := \sup_{1\le l\le m_n-k-1} E\Big[\sup_{B\in\mathcal{B}^{\Gamma,m_n}_{n,l+k+1}} |\, P(B|\mathcal{B}^{\Gamma,l}_{n,1}) - P(B)|\Big], \tag{3.1.1}$$

where $\mathcal{B}^{\Gamma,j}_{n,i} = \sigma((\Gamma_{n,l})_{i\le l\le j})$ and $k \in \mathbb{N}$, $k \le m_n$.
*$(\Gamma_{n,i})_{1\le i\le n}$ is called β-**mixing** (or **absolutely regular**) if $\beta^{\Gamma}_{n,k_n} \to 0$ for a sequence $k_n \to \infty$ of natural numbers.*

Note that the inner supremum in (3.1.1) is measurable since E is a polish space.
For the following, fix sequences $r_n = o(n)$ and $s_n = o(r_n)$ of natural numbers. The general empirical process considered below is indexed by an arbitrary family of real-valued functionals $\mathcal{G}$. The aim is a limit theorem for the empirical process $(Z_n(g))_{g\in\mathcal{G}}$ defined by

$$Z_n(g) := \frac{1}{\sqrt{p_n}} \sum_{i=1}^{m_n} (V_{n,i}(g) - E[V_{n,i}(g)]), \qquad g \in \mathcal{G}. \tag{3.1.2}$$

Here $V_{n,i}$, $1 \le i \le m_n, n \in \mathbb{N}$, are some real-valued random processes, likewise indexed by $\mathcal{G}$, $m_n := \lfloor (n - s_n + 1)/r_n \rfloor$ and

$$p_n := P(\exists g \in \mathcal{G} : V_{n,1}(g) \neq 0) \to 0. \tag{3.1.3}$$

This last convergence is motivated by the extreme value theory and represents that we consider rare events, i.e. values different from 0 are unlikely.
Throughout this chapter, we assume that the set $\{\exists g \in \mathcal{G} : V_{n,1}(g) \neq 0\}$ is measurable. This holds under the following condition, which helps to avoid measurability problems.

(D0) The processes $(V_{n,1}(g))_{g\in\mathcal{G}}$, $n \in \mathbb{N}$, are separable.

Condition (D0) is in particular fulfilled if $\mathcal{G}$ is finite. If condition (D0) holds, then $\{\exists g \in \mathcal{G} : V_{n,1}(g) \neq 0\}$ can be represented as a countable union of measurable sets and the set is therefore measurable itself. The condition (D0) is needed to ensure that p_n is well defined. Condition (D0) also ensures that $\beta^V_{n,k}$ is well defined.
In this general setting, we will not specify $V_{n,i}$. The only requirements on the process $(V_{n,i}(g))_{1\leq i\leq n}$ are that it should be stationary for all $n \in \mathbb{N}$ and for all $g \in \mathcal{G}$ and that the random variables $V_{n,i}$ should be measurable with respect to $(X_{n,(i-1)r_n+1}, ..., X_{n,ir_n+s_n})$. Because of this measurablity, we will sometimes call any $V_{n,i}$ a block. To simplify the notation, we write V_n for an arbitrary block and due to stationarity, $V_n \overset{d}{=} V_{n,i}$ holds for all $1 \leq i \leq m_n$. The assumption of measurability of $V_{n,i}$ with respect to a block of the X_t's comes from the fact that we have mainly statistics in mind which deal with estimation problems based on the data $(X_{n,i})_{1\leq i\leq n}$.

Example. For estimators defined as the average of disjoint blocks statistics, a canonical example for the choice of $V_{n,i}$ is $V_{n,i} = m_n^{-1/2}\sum_{j=1}^{r_n} g(X_{n,(i-1)r_n+j})$ for some function g with $g(X_{n,i}) \neq 0$ if and only if $X_{n,i} \neq 0$. In this case, $p_n = \mathrm{O}(r_n P(X_{n,1} \neq 0))$, $m_n = \mathrm{O}(n/r_n)$, $s_n = 1$ and the normalization of the sum in the process Z_n is given by $\sqrt{p_n m_n} = \mathrm{O}(\sqrt{nP(X_{n,1} \neq 0)})$. This choice leads to the generalized tail array sums considered in Section 3 of Drees and Rootzén (2010).
The choice $V_{n,i}(g) = d_n(g)^{-1}\sum_{j=0}^{r_n/s_n-1} g(W_{n,(i-1)r_n+js_n+1})$ (assuming that r_n is a multiple of s_n) for suitable normalizing sequences $d_n(g)$ can also be used for the analysis of general sums of disjoint blocks statistics, see below for details.
The choice $V_{n,i}(g) = b_n(g)^{-1}\sum_{j=1}^{r_n} g(W_{n,(i-1)r_n+j})$ for a suitable normalizing sequence $b_n(g)$ leads to the sums of sliding blocks statistics which will be considered in Section 3.2. ◇

The sequence $(r_n)_{n\in\mathbb{N}}$ is $\mathbb{N}$-valued and models the length of the blocks of $(X_{n,i})$ with respect to which the $V_{n,i}$ are measurable. We usually assume $r_n \to \infty$ and $r_n = \mathrm{o}(n)$. The size s_n indicates the length of the overlap of the blocks of X-observations with respect to which the V_n are measurable. We assume $s_n = \mathrm{o}(r_n)$, such that the overlap of these X-blocks reaches only into the consecutive block and is asymptotical negligible.
In applications, V_n will often be given via $V_n = V_n'/b_n$ for another process V_n' and some standardizing $\mathbb{R}^+$-valued sequence $(b_n)_{n\in\mathbb{N}}$. Usually some standardization b_n is necessary for the conditions below to be satisfiable in applications (see also the previous example). Determining an appropriate normalization b_n such that the conditions are satisfied and the limit of the process convergence is non-trivial (i.e. not constant 0 or ∞) will be an

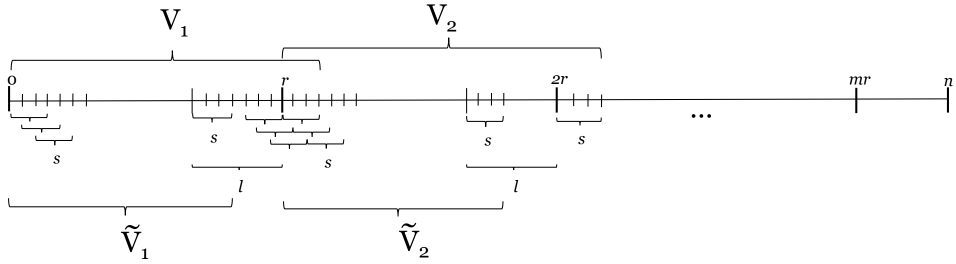

Figure 3.1: *Illustration of* $\tilde{V}_{n,i}$

important tasks for applications. It does not matter, whether the necessary normalization b_n appears explicitly in the theorems or is defined implicitly in V_n. To ease the notation, we omit the b_n in this section, it is only mentioned to draw attention to the need for appropriate normalization in V_n. However, it is not possible to omit the b_n in Section 3.2 and we will use the normalization there. We will restate the conditions including the normalizing constants there, such that one can see where and how the normalization has an impact on the conditions.

For the proofs in this section a *big block, small block* approach will be used. For the convergence of the finite-dimensional marginal distributions (fidis) of (3.1.2), the introduction of approximating random variables $\tilde{V}_{n,i}$ is necessary to bypass the dependence of the $V_{n,i}$ resulting from the overlaps of length s_n between the blocks. This stationary sequence of random processes $\tilde{V}_{n,i}$ approximates $V_{n,i}$ and is asymptotically independent for $1 \leq i \leq m_n$, more details on the approximation will be given later.

Example. An illustration of the idea of $\tilde{V}_{n,i}$ is given in Figure 3.1. In this figure, the axis represents the underlying process $(X_{n,i})_{1\leq i\leq n}$ and it is indicated to which random variables $V_{n,i}$ and $\tilde{V}_{n,i}$ refer. This figure shows one way how one could think of the $\tilde{V}_{n,i}$, but it is not entirely precise, since $\tilde{V}_{n,1}$ could be $(X_{n,1}, ..., X_{n,r_n+s_n})$-measurable. However, this figure shows how $\tilde{V}_{n,i}$ could be constructed as shortened blocks with gaps between the X-blocks w.r.t. which the $\tilde{V}_{n,i}$ are measurable. This is the way how $\tilde{V}_{n,i}$ is chosen for the sliding blocks in Section 3.2.

The random variables $V_{n,i} = m_n^{-1/2} \sum_{j=1}^{r_n} g(X_{n,(i-1)r_n+j})$ from the previous example can be approximated by $\tilde{V}_{n,i} = m_n^{-1/2} \sum_{j=1}^{r_n-l_n} g(X_{n,(i-1)r_n+j})$ for a suitable sequence $l_n = \mathrm{o}(r_n)$. $\diamond$

In the following section we will establish conditions for the weak convergence of the fidis of $(Z_n(g))_{g\in\mathcal{G}}$. Thereafter, conditions for the asymptotic tightness and asymptotic equicontinuity of the process are derived. The fidi convergence together with asymptotic equicontinuity or asymptotic tightness implies the weak convergence of $(Z_n(g))_{g\in\mathcal{G}}$ in $l^\infty(\mathcal{G})$. Here, $l^\infty(\mathcal{G})$ denotes the space of the real-valued bounded functions on $\mathcal{G}$. Up to this point and also in the following, many different sequences are used. An overview of numerous sequences used in this chapter, their interpretation, their typical behavior and their first occurrence is given in Table 3.1.

3.1.1 Convergence of fidis

In this section, assumptions for the convergence of the fidis of $(Z_n(g))_{g\in\mathcal{G}}$ defined in (3.1.2) are given. Fundamental are the following stationarity conditions, the dependency conditions in form of β -mixing conditions and the assumptions on the sizes of the emerging blocks. All these conditions were roughly sketched and motivated in the previous section.

(A1) $(X_{n,i})_{1\le i\le n}$ is stationary for all $n\in\mathbb{N}$ and the sequences $s_n, r_n\in\mathbb{N}$ satisfy $s_n = \mathrm{o}(r_n)$ and $r_n = \mathrm{o}(n)$.

(V) For all $n\in\mathbb{N}$, $1\le i\le m_n = \lfloor (n-s_n+1)/r_n\rfloor$, $V_{n,i}$ and $\tilde{V}_{n,i}$ are real-valued processes indexed by $\mathcal{G}$ that are measurable w.r.t. $(X_{n,(i-1)r_n+1},\dots,X_{n,ir_n+s_n-1})$ and $(V_{n,i},\tilde{V}_{n,i})_{1\le i\le m_n}$ is stationary.

(M$\tilde{\mathbf{V}}$) $m_n\beta^{\tilde{V}}_{n,0}\to 0$

(MX$_k$) $m_n\beta^X_{n,(k-1)r_n-s_n}\to 0$ for some $k\in\mathbb{N}$, $k\ge 2$.

The convergence of the overlap $s_n\to\infty$ is explicitly not required. This is possible but technically not necessary for the limit theorems below. The identity $s_n=1$, for all $n\in\mathbb{N}$, is also allowed, e.g. for the applications with disjoint blocks. Note that $s_n=\mathrm{o}(r_n)$ implies $r_n\to\infty$ since $s_n\ge 1$. The constraints on r_n are the usual conditions for block building for limit theorems which ensure that the blocks are not too large and the condition on s_n just states that the X-blocks may overlap, but asymptotically the overlap is negligible.
The mixing condition (M$\tilde{V}$) ensures the asymptotic independence of the approximating random variables $\tilde{V}_{n,i}$. The mixing condition (MX$_k$) is needed for the dependence between the $X_{n,i}$, and, therefore, also between the $V_{n,i}$, to be sufficiently weak. It enables us to replace $X_{n,i}$ by independent copies. One has $\beta^X_{n,p}\le\beta^X_{n,\tilde{p}}$ for $p>\tilde{p}$ and therefore, condition (MX$_k$) is the weaker the larger k can be chosen. In many applications, such as in the examples from Drees and Rootzén (2010), it can be chosen as $k=2$ for simplicity. Depending on the shape of the mixing coefficient, the choice of k may also have impact on the rate with which $(\beta^X_{n,k})_{n\in\mathbb{N}}$ tends to 0 for increasing n. Therefore, the variable k introduces some additional flexibility and may cover more cases than a fixed k. This Condition (MX$_k$) also implies β-mixing of $V_{n,i}$ as obtained by the next lemma.

Lemma 3.1.2. *Condition (MX$_k$) implies $m_n\beta^V_{n,k-1}\to 0$.*

This β-mixing property of $V_{n,i}$ would be sufficient for Theorems 3.1.7 and 3.1.9 below and is less restrictive than (MX$_k$).
The proof of the fidis convergence of $(Z_n(g))_{g\in\mathcal{G}}$ uses a technique which works similar to the *big block, small block*-method. Since V_n is interpreted as a block but, other than in

the given examples, must not be directly recognizable as a block, it is not immediately clear what the small block should look like. This role will be taken by the difference

$$\Delta_{n,i} := V_{n,i} - \tilde{V}_{n,i},$$

which is why the approximating $\tilde{V}_{n,i}$ are needed. To simplify the notation, Δ_n stands for any $\Delta_{n,i}$, i.e. $\Delta_n \overset{d}{=} \Delta_{n,i}$ holds. The random variable $\tilde{V}_{n,i}$ introduced above should approximate $V_{n,i}$ sufficiently accurately, i.e. the approximating error should be asymptoticly negligible; the concrete meaning of this is determined by condition (Δ) below. More details on this approximation are given in the proof to Lemma 3.4.1. Condition $(\mathrm{M}\tilde{V})$ formalizes that the $\tilde{V}_{n,i}$ are almost independent. This is needed, in order for a central limit theorem for independent random variables to be applicable.
The usage of the approximating random variables $\tilde{V}_{n,i}$ facilitates a true generalization opposed to a setting where the mixing condition in $(\mathrm{M}\tilde{V})$ is directly asked to be satisfied by V_n. Here, $V_{n,i}$ may have a stronger dependency.

Example. For $V_{n,i} = m_n^{-1/2} \sum_{j=1}^{r_n} g(X_{n,(i-1)r_n+j})$, as considered in the previous example, one can choose $\tilde{V}_{n,i} = m_n^{-1/2} \sum_{j=1}^{r_n-l_n} g(X_{n,(i-1)r_n+j})$ for some sequence $l_n \to \infty$ with $l_n = \mathrm{o}(r_n)$ and $s_n = 1$. Here, $\tilde{V}_{n,i}$ is even $(X_{n,(i-1)r_n+1}, ..., X_{n,ir_n-l_n})$-measurable. In this case, $m_n\beta^X_{n,l_n-s_n} \to 0$ implies $m_n\beta^{\tilde{V}}_{n,0} \to 0$, the proof of this is similar to the proof of Lemma 3.1.2. Since $l_n = \mathrm{o}(r_n)$, $m_n\beta^X_{n,r_n-s_n} \to 0$ is implied by $m_n\beta^X_{n,l_n-s_n} \to 0$. So, the condition $(\mathrm{M}\tilde{V})$ is not needed if a corresponding β-mixing condition on $(X_{n,i})_{1\le i\le n}$, $n \in \mathbb{N}$, is imposed. In particular, the dependency conditions can be set at the level of $X_{n,i}$ if a stricter measurability assumption holds. ◇

For the convergence of the fidis, a Lindeberg condition (L) and the convergence of the covariance function (C) are necessary, in order for a known central limit theorem to be applicable. More precisely, (L) and (C) imply convergence for the sum of independent copies of $V_n(g)$.

(Δ) (i) $E\left[(\Delta_n(g) - E[\Delta_n(g)])^2 \mathbb{1}_{\{|\Delta_n(g)-E[\Delta_n(g)]|\le\sqrt{p_n}\}}\right] = \mathrm{o}\left(\frac{p_n}{m_n}\right) \qquad \forall g \in \mathcal{G}$

(ii) $P\left(|\Delta_n(g) - E[\Delta_n(g)]| > \sqrt{p_n}\right) = \mathrm{o}\left(\frac{1}{m_n}\right) \qquad \forall g \in \mathcal{G}.$

(iii) There exists $\tau > 0$ such that

$$E\left[(\Delta_n(g) - E[\Delta_n(g)]) \mathbb{1}_{\{|\Delta_n(g)-E[\Delta_n(g)]|\le\tau\sqrt{p_n}\}}\right] = \mathrm{o}\left(\frac{\sqrt{p_n}}{m_n}\right) \qquad \forall g \in \mathcal{G}.$$

(L) $E\left[(V_n(g) - E[V_n(g)])^2 \mathbb{1}_{\{|V_n(g)-E[V_n(g)]|>\varepsilon\sqrt{p_n}\}}\right] = \mathrm{o}\left(\frac{p_n}{m_n}\right) \quad \forall g \in \mathcal{G}, \varepsilon > 0$

(C) There exists a function $c : \mathcal{G}^2 \to \mathbb{R}$ such that

$$\frac{m_n}{p_n} Cov\left(V_n(g), V_n(h)\right) \to c(g,h), \qquad \forall\, g,h \in \mathcal{G}.$$

It is also possible to use the following condition (L*) instead of (L). The Lyapunov condition (L*) implies the Lindeberg condition (L) and can be used as a slightly stronger condition which is easier to verify.

$$\textbf{(L*)}\quad \exists \delta > 0 : \frac{m_n}{(\sqrt{p_n})^{2+\delta}} E\left[(V_n(g))^{2+\delta}\right] \to 0 \qquad\qquad \forall g \in \mathcal{G}.$$

Condition (Δ) is implied by a simpler condition.

Lemma 3.1.3. *Suppose (A1), (V) and (D0) are met. Then, Condition (Δ) is fulfilled if*

$$E\left[(\Delta_n(g))^2\right] = o\left(\frac{p_n}{m_n}\right) \qquad \forall g \in \mathcal{G}, \tag{3.1.4}$$

or if $Var\left(\Delta_n(g)\right) = o\left(p_n/m_n\right)$, respectively.

For part (i) and (ii) of (Δ), this lemma can be verified by direct calculations. For part (iii), a more sophisticated argument is needed. In fact, for our applications we will always check (3.1.4) rather than (Δ).
All these conditions introduced and discussed so far will be used to establish fidis convergence of the empirical process $(Z_n(g))_{g\in\mathcal{G}}$.

Theorem 3.1.4. *Suppose the conditions (A1), (V), ($M\tilde{V}$), (MX_k) for some $k \in \mathbb{N}$, $k \geq 2$, (D0), (Δ), (L) and (C) are satisfied. Then the fidis of the empirical process $(Z_n(g))_{g\in\mathcal{G}}$ converge weakly to the fidis of a Gaussian process with covariance function c.*
The statement remains true if (Δ) is replaced by (3.1.4).

3.1.2 Process convergence

Convergence of the whole process $(Z_n(g))_{g\in\mathcal{G}}$ can be concluded if in addition to fidis convergence the process is asymptotically tight or asymptotically equicontinuous.
First, we consider conditions that ensure the asymptotic tightness of the empirical process. Later on in this section, we also discuss asymptotic equicontinuity. As usual in empirical process theory outer probabilities and outer expectation are denoted by P^* and E^*, respectively.

Definition 3.1.5 (Asymptotic Tightness)**.** *The sequence $(Z_n)_{n\in\mathbb{N}}$ is **asymptotically tight** if for all $\varepsilon > 0$ there exists a compact set $K \subset l^\infty(\mathcal{G})$ such that*

$$\limsup_{n\to\infty} P^*(Z_n \notin K^\delta) < \varepsilon \qquad\qquad \forall \delta > 0.$$

Here $K^\delta := \left\{f \in l^\infty(\mathcal{G}) | \inf_{k\in K} d_{l^\infty(\mathcal{G})}(f,k) \leq \delta\right\}$ denotes the set of all elements of $l^\infty(\mathcal{G})$ with a distance less than or equal to δ to K with respect to the metric $d_{l^\infty(\mathcal{G})}$ on $l^\infty(\mathcal{G})$.

The following conditions are needed to show the asymptotic tightness of $(Z_n(g))_{g\in\mathcal{G}}$.

(B) $E[|V_n(g)|^2] < \infty$ for all $g \in \mathcal{G}$, $n \geq 1$ and the paths of the process $(V_n(g))_{g\in\mathcal{G}}$ are bounded, i.e.

$$V_n(\mathcal{G}) := \sup_{g\in\mathcal{G}} |V_n(g)| < \infty \qquad \text{a.s.}$$

(L1) $\dfrac{m_n}{\sqrt{p_n}} E^*\left[V_n(\mathcal{G})\mathbb{1}_{\{V_n(\mathcal{G})>\varepsilon\sqrt{p_n}\}}\right] \to 0 \qquad \forall \varepsilon > 0$

(D1) There exists a semi-metric ρ on $\mathcal{G}$ such that $\mathcal{G}$ is totally bounded (i.e. for all $\varepsilon > 0$ the set $\mathcal{G}$ can be covered by finitely many balls with radius ε w.r.t. ρ) and

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \sup_{g,h\in\mathcal{G},\rho(g,h)<\delta} \frac{m_n}{p_n} E\left[(V_n(g) - V_n(h))^2\right] = 0 \tag{3.1.5}$$

(D2)

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \int_0^\delta \sqrt{\log N_{[\cdot]}(\varepsilon, \mathcal{G}, L_2^n)}\, d\varepsilon = 0,$$

where $N_{[\cdot]}(\varepsilon, \mathcal{G}, L_2^n)$ denotes the ε-bracketing number of $\mathcal{G}$ w.r.t. L_2^n, i.e. the smallest number N_ε such that there exists a partition $(\mathcal{G}^\varepsilon_{n,k})_{1\leq k\leq N_\varepsilon}$ of $\mathcal{G}$ for each $n \in \mathbb{N}$ satisfying

$$\frac{m_n}{p_n} E^*\left[\sup_{g,h\in\mathcal{G}^\varepsilon_{n,k}} (V_n(g) - V_n(h))^2\right] \leq \varepsilon^2, \qquad \forall 1 \leq k \leq N_\varepsilon.$$

Condition (B) and (L1) essentially bound the size of the set $\mathcal{G}$. Roughly speaking, condition (D1) ensures the continuity of the paths of the limit process w.r.t. ρ. Condition (D2) is most restrictive and limits the complexity of $\mathcal{G}$ in terms of the bracketing number, which is an entropy measure, see e.g. Van der Vaart and Wellner (1996), Section 2.11 for details.

If for all $\varepsilon > 0$ the partition of $\mathcal{G}$ in (D2) does not depend on n, then the condition (3.1.5) in (D1) can be omitted. More precisely: if for all $\varepsilon > 0$ there exist a partition $(\mathcal{G}^\varepsilon_k)_{1\leq k\leq \tilde{N}_\varepsilon}$ of $\mathcal{G}$ which does not depend on n and which satisfies

$$\frac{m_n}{p_n} E^*\left[\sup_{g,h\in\mathcal{G}^\varepsilon_k} (V_n(g) - V_n(h))^2\right] \leq \varepsilon^2, \qquad \forall 1 \leq k \leq \tilde{N}_\varepsilon,$$

then (D1) and (D2) can be replaced by the following simpler condition: There exists a semi-metric ρ on $\mathcal{G}$ such that $\mathcal{G}$ is totally bounded and

$$\int_0^\delta \sqrt{\log \tilde{N}_\varepsilon}\, d\varepsilon < \infty$$

for some $\delta > 0$ (cf. Theorem 2.11.9 in Van der Vaart and Wellner (1996)).

Instead of (L1), the more restrictive condition (L2), which essentially restricts the size of $V_n(\mathcal{G})$, can be used.

(L2)

$$\frac{m_n}{p_n} E^* \left[(V_n(\mathcal{G}))^2 \mathbb{1}_{\{V_n(\mathcal{G}) > \varepsilon\sqrt{p_n}\}} \right] \to 0 \qquad \forall \varepsilon > 0.$$

An advantage of condition (L2) is that it also implies (L). (L1) is a weaker condition than (L2) and is sufficient to proof tightness in Theorem 3.1.7 below. For the proof of the equicontinuity in Theorem 3.1.9 considered below we will need the stronger condition (L2) anyway, which is why we introduce this condition here.

Lemma 3.1.6. *Condition (L2) implies (L) and (L1).*

These conditions allow us to conclude asymptotic tightness of $(Z_n(g))_{g\in\mathcal{G}}$.

Theorem 3.1.7. *Suppose the conditions (A1), (V), (MX$_k$) for some $k \in \mathbb{N}$, $k \geq 2$, (D0), (B), (L1), (D1) and (D2) are satisfied. Then, the process $(Z_n(g))_{g\in\mathcal{G}}$ is asymptotically tight.*

Alternatively to the asymptotic tightness of the process, one can consider asymptotic equicontinuity to achieve process convergence.

Definition 3.1.8 (Asymptotic Equicontinuity)**.** *The process $(Z_n(g))_{g\in\mathcal{G}}$ is **asymptotically equicontinuous** with respect to a semi-metric ρ if*

$$\forall\, \varepsilon > 0,\, \eta > 0\, \exists\, \delta > 0 : \limsup_{n\to\infty} P^* \left(\sup_{g,h\in\mathcal{G}, \rho(g,h)<\delta} |Z_n(g) - Z_n(h)| > \varepsilon \right) < \eta.$$

We use the following condition to establish asymptotic equicontinuity of $(Z_n(g))_{g\in\mathcal{G}}$.

(D3) Denote by $N(\varepsilon, \mathcal{G}, d_n)$ the ε-covering number of $\mathcal{G}$ w.r.t. the random semi-metric

$$d_n(g,h) = \left(\frac{1}{p_n} \sum_{i=1}^{m_n} (V^*_{n,i}(g) - V^*_{n,i}(h))^2 \right)^{1/2}$$

with $V^*_{n,i}$, $1 \leq i \leq m_n$, being independent copies of $V_{n,1}$, i.e. $N(\varepsilon, \mathcal{G}, d_n)$ is the smallest number of balls with d_n-radius ε which is needed to cover $\mathcal{G}$. We assume

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} P^* \left(\int_0^\delta \sqrt{\log(N(\varepsilon, \mathcal{G}, d_n))} d\varepsilon > \tau \right) = 0, \qquad \forall \tau > 0.$$

The condition (D3) ensures that the parameter set $\mathcal{G}$ is not too complex by restricting the covering number, which is an entropy measure. A, possibly, simpler criterion to verify condition (D3) can be given using the Vapnik-Cervonenkis (VC) theory. The condition (D3) is satisfied if $\mathcal{G}$ is a so-called VC-class or a VC-hull class (cf. Van der Vaart and Wellner (1996), Section 2.6, or Drees and Rootzén (2010), Remark 2.11).
Technically, the following measurability condition is also required in the proof of the next theorem.

(D4) For all $\delta > 0, n \in \mathbb{N}, (e_i)_{1\le i\le \lceil m_n/k\rceil} \in \{-1,0,1\}^{\lceil m_n/k\rceil}$ and $a \in \{1,2\}$ the random variable

$$\sup_{g,h\in\mathcal{G},\rho(g,h)<\delta} \sum_{j=1}^{\lceil m_n/k\rceil} e_j(V^*_{n,j}(g) - V^*_{n,j}(h))^a$$

is measurable, where $V^*_{n,i}$, $1 \le i \le m_n$ are independent copies of V_n and k is determined such that condition (MX_k) holds.

However, this measurability condition (D4) is implied by the simpler condition (D0) introduced above. This is due to the fact that separability implies the measurability of the supremum in (D4). Therefore, we do not need to assume the condition (D4) separately.

Theorem 3.1.9. *Suppose the conditions (A1), (V), (MX_k) for some $k \in \mathbb{N}$, $k \ge 2$, (D0), (B), (L2), (D1) and (D3) are satisfied. Then the process $(Z_n(g))_{g\in\mathcal{G}}$ is asymptotically equicontinuous.*

The previous theorems on convergence of fidis and on asymptotic equicontinuity or asymptotic tightness can be summarized to one theorem about the process convergence of $(Z_n(g))_{g\in\mathcal{G}}$.

Theorem 3.1.10. *If one of the two sets of conditions*

(i) (A1), (MX_k), ($M\tilde{V}$), (D0), (Δ), (C), (B), (L2), (D1) and (D3),

(ii) (A1), (MX_k), ($M\tilde{V}$), (D0), (Δ), (L), (C), (B), (L1), (D1) and (D2)

are satisfied, then the empirical process $(Z_n(g))_{g\in\mathcal{G}}$ converges weakly to a Gaussian process $(Z(g))_{g\in\mathcal{G}}$ with covariance function c.
The statement remains true, if one replaces condition (Δ) by condition (3.1.4)*.*

A special case of the theory considered in this chapter is the more specific situation in Drees and Rootzén (2010) with cluster functionals. With the special choice $V_{n,i}(f) = (p_n/(nv_n))^{1/2} f(X_{n,(i-1)r_n+1}, \ldots . X_{n,ir_n})$, for a measurable cluster functional f, where $v_n = P(X_{n,1} \neq 0)$, we are exactly in the situation from Drees and Rootzén (2010). All examples listed there are therefore examples for the application of the theory derived here.
In contrast to the typical PoT-settings in the literature, in the setting considered here, extreme value theoretical situations can also be taken into account, in which also small (not extreme) observations are included in the statistics. An application of this feature is given by the projection based estimator in Chapter 5. Often it is required that there are only a few extreme observations that are not equal to 0 and only extreme observations can have an impact on the statistics considered. In our setting, only $p_n \to 0$ is required, i.e. most blocks are equal to 0, but in the non-zero blocks many non extreme observations can be considered and can have an impact on statistics. Also in this respect the theory of this chapter offers an important generalization.
So far, we introduced fairly general and abstract limit theorems for the empirical process in (3.1.2). In the next section, we analyze the more specific (but still quite general) case of sums of sliding blocks statistics.

3.2 Sliding blocks limit theorem

Sums of sliding blocks statistics are a special form of the random variables V_n as introduced before. They can be used for the analysis of statistics based on overlapping blocks as introduced in the beginning of this chapter. We will call the following $V_{n,i}$ a *sliding-blocks-sum* or, more shortly, *sliding-blocks*:

$$V_{n,i}(g) = \frac{1}{b_n(g)} \sum_{j=1}^{r_n} g(X_{n,(i-1)r_n+j}, ..., X_{n,(i-1)r_n+j+s_n}) = \frac{1}{b_n(g)} \sum_{j=1}^{r_n} g(W_{n,(i-1)r_n+j}), \quad (3.2.1)$$

where $b_n(g) > 0$ is a normalization constant for each $g \in \mathcal{G}$ with $b_n(g) \to \infty$ for $n \to \infty$. For simplicity of notation, recall $W_{n,j} = (X_{n,j}, ..., X_{n,j+s_n-1})$ from (3.0.1). Here again, r_n denotes a sequence that grows faster than s_n but slower than n. Furthermore, r_n is chosen such that it is unlikely to have any extreme value in a sequence of r_n consecutive observations of $X_{n,i}$.
In this setting, one has

$$p_n = P\Big(\exists g \in \mathcal{G} : V_n(g) \neq 0\Big) \leq P\Big(\exists g \in \mathcal{G}, i \in \{1, ..., r_n\} : g(W_{n,i}) \neq 0\Big).$$

In usual statistical applications one has $g \geq 0$ for all $g \in \mathcal{G}$. Then, the inequality is sharp. In this section, the normalization is given by $b_n(g)$. In order to see where the standardization has an impact on the processes and conditions of the previous section one has to replace $V_{n,i}(g)$ by $b_n(g)^{-1} \sum_{j=1}^{r_n} g(W_{n,(i-1)r_n+j})$. The use of normalizations $b_n(g)$ depending on $g \in \mathcal{G}$ allows that limit results can be achieved for different normalizations, which increases the flexibility of the results w.r.t. $\mathcal{G}$. An example of an application where this is necessary is the sliding blocks estimator for the extremal index in Section 4.2 where numerator and denominator are sliding blocks statistics with different normalizations but a joint convergence is needed.
For the following more specific limit results for the sliding blocks sums, we require $p_n \to 0$ and $\sqrt{p_n} b_n(g) \to \infty$ for all $g \in \mathcal{G}$. The previous condition implies $b_n(g) \to \infty$, since $p_n \to 0$. The normalization of the empirical process $Z_n(g)$ is $\sqrt{p_n} b_n(g)$ and should be proportional to the square root of the expected number of observations $X_{n,j}$ for which an $i \in \{j, ..., j + r_n - 1\}$ exists such that $g(W_{n,i}) \neq 0$, i.e. the number of observations which have an impact on $V_{n,i}(g)$. In order to expect a convergence to a Gaussian process, the common assumption seems necessary that the normalization converges to ∞.
A family of functions $\mathcal{G}$ with $g : E^{\mathbb{N}} \to \mathbb{R}$ for all $g \in \mathcal{G}$ is considered. We use the usual embedding of E^n, $n \in \mathbb{N}$, in $E^{\mathbb{N}}$ by the map $E^n \ni (x_i)_{1 \leq i \leq n} \mapsto (x_i)_{i \in \mathbb{N}} \in E^{\mathbb{N}}$ with $x_i = 0$ for $i > n$. Instead of p_n, the normalizations

$$q_{g,n} := P(g(W_{n,1}) \neq 0) \text{ and } q_{\mathcal{G},n} := P(\exists g \in \mathcal{G} : g(W_{n,1}) \neq 0) \quad (3.2.2)$$

are also used and, as for p_n, we require $q_{\mathcal{G},n} \to 0$. For $g \geq 0$ one has $q_{\mathcal{G},n} \leq p_n$, such that

$q_{\mathcal{G},n} \to 0$ already follows from $p_n \to 0$.
For most statistical applications, functions $g \geq 0$ that satisfy $g(0) = 0$ are of interest where 0 represents the null element $0 := (0, ..., 0) \in E^n$ for all $n \in \mathbb{N}$. This means that g maps to 0 whenever no observation different from 0 appears in the block of length s_n. This condition on the function g is weaker than the conditions of cluster functionals, since all observations, including those with value 0 after the last or before the first extreme observation, can have an effect if there is only one extreme observation in the block. An application where this is needed is the projection based estimator in Chapter 5.
In this section, we use the approximating random variables

$$\tilde{V}_{n,i}(g) := \frac{1}{b_n(g)} \sum_{j=1}^{r_n - l_n} g(W_{n,(i-1)r_n+j}),$$

where $l_n = \mathrm{o}(r_n)$ and $l_n \to \infty$. In particular, $\tilde{V}_{n,i}$ is $(X_{n,(i-1)r_n+1}, ..., X_{n,(i-1)r_n-l_n+s_n})$-measurable. The sequence l_n describes the length of the *small block* $\Delta_n = V_n - \tilde{V}_n$, which is used in the *big block, small block* arguments in the abstract theory. Here, $s_n \leq l_n$ is additionally required, such that the cut block of length l_n ensures that the gap of observations between two blocks $(X_1, ..., X_{n,r_n+s_n-l_n})$ and $(X_{r_n+1}, ..., X_{2r_n+s_n-l_n})$ exists and is not entirely covered by the overlap of length s_n.
In the previous paragraphs we have imposed some conditions on sequences and their convergence rates which are collected along with other assumptions in the following conditions.

(A) $(X_{n,i})_{1 \leq i \leq n}$ is stationary for all $n \in \mathbb{N}$.

(A2) The sequences $l_n, r_n, s_n \in \mathbb{N}$, p_n, and $b_n(g) > 0$, $g \in \mathcal{G}$, satisfy $s_n \leq l_n = \mathrm{o}(r_n)$, $r_n = \mathrm{o}(n)$, $p_n \to 0$ and $r_n = \mathrm{o}\Big(\sqrt{p_n} \inf_{g \in \mathcal{G}} b_n(g)\Big)$.

For the above defined $\tilde{V}_{n,i}$ and $V_{n,i}$ we can summarize the mixing condition (M$\tilde{V}$) and (MX$_k$) by one mixing condition (cf. Lemma 3.1.2):

(MX) $(n/r_n)\beta^X_{n,l_n-s_n} \to 0$.

To recap all the sequences that appear and in order to give a compact overview, numerous sequences are summarized in Table 3.1.
Formally, the condition $s_n \leq l_n$ is enough to prove the following results. However, the smaller $l_n - s_n$, the stronger is the β-mixing condition (MX). Therefore, it seems appropriate to choose s_n such that $l_n - s_n \to \infty$ to weaken the mixing condition.

Example. Consider the typical case $X_{n,i} = (X_i/u_n)\mathbb{1}_{\{\|X_i\| > u_n\}}$ for a stationary time series $(X_t)_{t \in \mathbb{Z}}$ and some sequence of thresholds $u_n \to \infty$. Then, only the extreme values of X, i.e. those with $\|X_i\| > u_n$, have an influence on the value of $g(W_{n,1})$. In this case the β-mixing condition (MX) is satisfied if $(n/r_n)\tilde{\beta}^X_{n,l_n-s_n} \to 0$ where $\tilde{\beta}$ are the β-mixing coefficients of the time series $(X_t)_{t \in \mathbb{Z}}$. However, the weaker condition (MX), where the

	interpretation	$\to$	*main constraints*	*typ. behavior*	*first use*
n	number of observations	∞			p.19
s_n	length of sliding blocks			$s_n \to \infty$	(3.0.1)
r_n	length of big block	∞	Sect. 3: $r_n = \mathrm{o}(n)$, $r_n = \mathrm{o}\big(\sqrt{p_n}\inf_{g\in\mathcal{G}} b_n(g)\big)$ Sect. 4.2: $r_n v_n \to 0$, $r_n = \mathrm{o}\big(\sqrt{n v_n}\big)$		(3.1.2)
l_n	length of small block	∞	$s_n \le l_n = \mathrm{o}(r_n)$		(A2)
m_n	number of big blocks	∞	$m_n \asymp n/r_n$		(3.1.2)
u_n	threshold for X to be large	∞			p.19
v_n	$P(X_{n,1} \neq 0)$	0	$n v_n \to \infty$		p.40
p_n	$P(\exists 1 \le i \le r_n : X_{n,i} \neq 0)$	0	$r_n = \mathrm{o}\big(\sqrt{p_n}\inf_{g\in\mathcal{G}} b_n(g)\big)$	$p_n \asymp r_n v_n$	(3.1.3)
$q_{\mathcal{G},n}$	$P(\exists g \in \mathcal{G} : g(W_{n,1}) \neq 0)$	0	$r_n q_{\mathcal{G},n} = \mathrm{O}(p_n)$	$q_{\mathcal{G},n} \asymp v_n$ or $q_{\mathcal{G},n} \asymp s_n v_n$	(3.2.2)
$b_n(g)$	normalizing constant	∞	$\sqrt{p_n} b_n(g) \to \infty$	$b_n(g) \asymp \sqrt{m_n}$ or $b_n(g) \asymp \sqrt{m_n} s_n$	(3.2.1)
a_n	normalization in Section 3.3			$a_n \asymp 1$	(3.3.1)
k_n	restricts β-mixing in Thm. 3.3.6	∞	$k_n = \mathrm{o}(r_n a_n)$		Thm. 3.3.6

Table 3.1: *Overview of sequences occurring in Chapter 3.*

β-mixing coefficients corresponds to $(X_{n,i})_{1\le i\le n}$, suffices, i.e. only the mixing behavior of the extreme part of the distribution of the time series is relevant, not the mixing properties of the whole (non-extreme) time series. Condition (MX) is always implied by the stronger corresponding mixing property of the underlying time series $(X_t)_{t\in\mathbb{Z}}$.

This choice of $X_{n,i}$ is used in the applications dealing with the estimation of cluster indexes in Chapter 4. ◇

In this sliding blocks setting, we can specify

$$Z_n(g) := \frac{1}{\sqrt{p_n}} \sum_{i=1}^{m_n} (V_{n,i}(g) - E[V_{n,i}(g)])$$
$$= \frac{1}{\sqrt{p_n} b_n(g)} \sum_{j=1}^{m_n r_n} (g(W_{n,j}) - E[g(W_{n,j})]), \qquad g \in \mathcal{G},$$

with $m_n = \lfloor (n - s_n + 1)/r_n \rfloor$ and

$$\Delta_n(g) = V_n(g) - \tilde{V}_n(g) = \frac{1}{b_n(g)} \sum_{j=r_n-l_n+1}^{r_n} g(W_{n,j}) \overset{d}{=} \frac{1}{b_n(g)} \sum_{j=1}^{l_n} g(W_{n,j}).$$

Furthermore, we define

$$\bar{Z}_n(g) := \frac{1}{\sqrt{p_n} b_n(g)} \sum_{j=1}^{n-s_n+1} (g(W_{n,j}) - E[g(W_{n,j})]), \qquad g \in \mathcal{G}. \tag{3.2.3}$$

We will see below that under suitable conditions the last $n - s_n + 1 - m_n r_n < r_n$ summands in definition (3.2.3) are asymptotically negligible.
The more specific setting allows us to derive simplified conditions for the convergence of the empirical process Z_n. Such a constellation is considered in the following theorem. Beforehand we rewrite condition (Δ), (L), (C) and (D1), (D2) for the more specific setting here, in particular, including the normalization constants $b_n(g)$ and the more concrete form of V_n and Δ_n. (Here, only the conditions are displayed where something changes compared to the previous section. Note that, with slight abuse of notation, we name these conditions as before even if they are less general. If, in the following sections, these conditions are cited, they refer to the conditions from the previous section. In the context of sliding blocks $V_{n,i}(g)$ defined in (3.2.1) is used.)

($\boldsymbol{\Delta}$) (i) It holds for all $g \in \mathcal{G}$ that

$$E\Big[\Big(\sum_{j=1}^{l_n}(g(W_{n,j}) - E[g(W_{n,1})])\Big)^2 \mathbb{1}_{\left\{|\sum_{j=1}^{l_n}(g(W_{n,j})-E[g(W_{n,1})])| \le \sqrt{p_n} b_n(g)\right\}}\Big] = \mathrm{o}\Big(\frac{p_n b_n(g)^2}{m_n}\Big).$$

(ii) $$P\Big(\Big|\sum_{j=1}^{l_n}(g(W_{n,j}) - E[g(W_{n,1})])\Big| > \sqrt{p_n} b_n(g)\Big) = \mathrm{o}\Big(\frac{1}{m_n}\Big) \qquad \forall g \in \mathcal{G}.$$

(iii) There exists $\tau > 0$ such that

$$E\Big[\Big(\sum_{j=1}^{l_n}(g(W_{n,j}) - E[g(W_{n,1})])\Big)\mathbb{1}_{\big\{|\sum_{j=1}^{l_n}(g(W_{n,j})-E[g(W_{n,1})])|\leq\tau\sqrt{p_n}b_n(g)\big\}}\Big] = \mathrm{o}\Big(\frac{\sqrt{p_n}b_n(g)}{m_n}\Big) \qquad \forall g\in\mathcal{G}.$$

(L) It holds for all $g \in \mathcal{G}$ and all $\varepsilon > 0$ that

$$E\Big[\Big(\sum_{j=1}^{r_n}(g(W_{n,j}) - E[g(W_{n,1})])\Big)^2\mathbb{1}_{\big\{|\sum_{j=1}^{r_n}(g(W_{n,j})-E[g(W_{n,1})])|>\varepsilon\sqrt{p_n}b_n(g)\big\}}\Big] = \mathrm{o}\Big(\frac{p_n b_n(g)^2}{m_n}\Big).$$

(C) There exists a covariance function $c : \mathcal{G}^2 \to \mathbb{R}$ such that

$$\frac{m_n}{p_n b_n(g)b_n(h)}Cov\Big(\sum_{i=1}^{r_n} g(W_{n,i}), \sum_{j=1}^{r_n} h(W_{n,j})\Big) \to c(g,h) \qquad \forall g,h\in\mathcal{G}.$$

(D1) There exists a semi-metric ρ on $\mathcal{G}$ such that $\mathcal{G}$ is totally bounded and

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\sup_{g,h\in\mathcal{G},\rho(g,h)<\delta}\frac{m_n}{p_n}E\Big[\Big(\sum_{j=1}^{r_n}\Big(\frac{1}{b_n(g)}g(W_{n,j}) - \frac{1}{b_n(h)}h(W_{n,j})\Big)\Big)^2\Big] = 0$$

holds.

(D2) It holds that

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\int_0^\delta\sqrt{\log N_{[\cdot]}(\varepsilon,\mathcal{G},L_2^n)}d\varepsilon = 0,$$

where $N_{[\cdot]}(\varepsilon,\mathcal{G},L_2^n)$ is the bracketing number, i.e. the smallest number N_ε such that for each $n \in \mathbb{N}$ there exists a partition $(\mathcal{G}^\varepsilon_{n,k})_{1\leq k\leq N_\varepsilon}$ of $\mathcal{G}$ such that

$$\frac{m_n}{p_n}E^*\Big[\sup_{g,h\in\mathcal{G}^\varepsilon_{n,k}}\Big(\sum_{j=1}^{r_n}\Big(\frac{g(W_{n,j})}{b_n(g)} - \frac{h(W_{n,j})}{b_n(h)}\Big)\Big)^2\Big] \leq \varepsilon^2 \qquad \forall 1\leq k\leq N_\varepsilon.$$

In the first theorem of this section, we consider the special situation of uniformly bounded functionals g, i.e. there exists a function $g_{\max} := \sup_{g\in\mathcal{G}}|g|$, which is bounded. We will refer to this theorem as sliding blocks limit theorem for bounded g, one central result of this chapter. The basic idea of this theorem and the subsequent corollary is comparable with Drees and Rootzén (2010), Corollary 3.6. However, this theorem holds in a much more general setting.

Theorem 3.2.1. *(a) Let $g_{\max} = \sup_{g\in\mathcal{G}}|g|$ be bounded and measurable. Suppose (A),* *(A2), (MX) and (D0) are satisfied. In addition, assume*

$$E\Big[\Big(\sum_{j=1}^{r_n} \mathbb{1}_{\{g(W_{n,j})\neq 0\}}\Big)^2\Big] = O\Big(\frac{p_n b_n(g)^2}{m_n}\Big) \qquad \forall g \in \mathcal{G}. \tag{3.2.4}$$

Then, the conditions (3.1.4)*, (B) and (L2) are satisfied. Moreover, it holds that*

$$\sup_{g\in\mathcal{G}} |Z_n(g) - \bar{Z}_n(g)| \xrightarrow{P^*} 0. \tag{3.2.5}$$

If, in addition, condition (C) is fulfilled, then the fidis of each of the empirical processes $(Z_n(g))_{g\in\mathcal{G}}$ *and* $(\bar{Z}_n(g))_{g\in\mathcal{G}}$ *converge weakly to the fidis of a Gaussian process* $(Z(g))_{g\in\mathcal{G}}$ *with covariance function c.*

(b) If, in addition to all assumptions in part (a), one of the sets of conditions

(i) (D1) and (D3) or

(ii) (D1) and (D2)

is fulfilled, then each of the empirical processes $(Z_n(g))_{g\in\mathcal{G}}$ *and* $(\bar{Z}_n(g))_{g\in\mathcal{G}}$ *converges weakly to a Gaussian process with covariance function c.*

If for all $\varepsilon > 0$ a partition of $\mathcal{G}$ in condition (D2) is used which does not depend on n, then the condition (3.1.5) in (D1) can be omitted (cf. Van der Vaart and Wellner (1996), Theorem 2.11.9, and the discussion directly after the definition of condition (D2)). Then, to verify the condition set *(ii)* in the previous theorem, it suffices to show (D2) with a partition independent of n and the total boundedness of $\mathcal{G}$.
If $r_n q_{\mathcal{G},n} = \mathrm{O}(p_n)$ is valid, then p_n can be replaced by $r_n q_{\mathcal{G},n}$ in any condition of the previous theorem. In this case, the new condition (3.2.4) can be verified at the level of the functions g alone, without taking $(V_n(g))_{g\in\mathcal{G}}$ into account.
The idea of the proof of Theorem 3.2.1 is based on the sum structure of $V_n(g)$ and on the fact that Δ_n is a shorter sum of the same form as $V_{n,i}$. The choice of $b_n(g)$ for each $g \in \mathcal{G}$ is central for condition (3.2.4).
In the previous theorem, we required the conditions (B), (L1), (D1), (D2) and (D3) for the whole set $\mathcal{G}$. Consider the special case of only finitely many different normalizations $b_1, ..., b_K$, i.e. $b_n(g) \in \{b_1, ..., b_K\}$ for all $g \in \mathcal{G}$. We denote the subsets of $\mathcal{G}$ which contains all functions g with the same normalization by $\mathcal{G}_i := \{g \in \mathcal{G} : b_n(g) = b_i\}$ for $i = 1, ..., K$. By the usual arguments for tightness and equicontinuity it is enough to establish the conditions (B), (L1), (D1), (D2) and (D3) for each $\mathcal{G}_i$, $i = 1, ..., K$, separately. Then, the asymptotic tightness or asymptotic equicontinuity holds for $\mathcal{G}$ as the finite union of these families $\mathcal{G}_i$, $i = 1, ..., K$.

Example. To give an idea how g and the corresponding normalization $b_n(g)$ could look like, we give a short example in anticipation of Section 4.2. There, we consider the bounded functions

$$g_1(x_1,\dots,x_{s_n}) := \mathbb{1}_{\{\max_{1\le i\le s_n} x_i>1\}}, \qquad g_2(x_1,\dots,x_{s_n}) := \mathbb{1}_{\{x_1>1\}}$$

to analyze the sliding blocks estimator of the extremal index. Here appropriate normalizing sequences are $b_n(g_1)=\sqrt{m_n}s_n$ and $b_n(g_2)=\sqrt{m_n}$. Note that already in this rather simple example, the normalizing sequences converge at a different rate for different functions. Indeed, it is somewhat archetypical that the event $\{g(W_{n,1})\neq 0\}$ either depends on all observations of the block $W_{n,1}$ (as for $g=g_1$), or it only depends on a single fixed observation $X_{n,i}$ (as for $g=g_2$); usually, the normalizing factor $b_n(g)$ is larger by the factor s_n in the former case. ◇

So far, we have only discussed the case of bounded functions g, which is sometimes too restrictive. This assumption can be dropped if the moment condition (3.2.4) is strengthened. For simplicity's sake, in the case of unbounded functions and for the process convergence (not for the fidi convergence), we assume that all functionals are normalized in the same way. The conclusion of the next theorem about unbounded $g_{\max}$ is basically the same as of Theorem 3.2.1. Hence, we will refer to this theorem as sliding blocks limit theorem for unbounded functions.

Theorem 3.2.2. *(a) Suppose (A), (A2), (MX), (D0) and (C) are met and $g_{\max}$ is not necessarily bounded. In addition, let $m_n l_n P(V_n(|g|)\neq 0)=o(r_n b_n^2(g)p_n)$ for all $g\in\mathcal{G}$ and*

$$E\Big[\Big(\sum_{i=1}^{r_n}|g(W_{n,i})|\Big)^{2+\delta}\Big]=O\Big(\frac{p_n b_n^2(g)}{m_n}\Big),\quad \forall\, g\in\mathcal{G}, \tag{3.2.6}$$

for some $\delta>0$. Then the conditions (3.1.4), (B) and (L) are satisfied. Moreover, the fidis of $(Z_n(g))_{g\in\mathcal{G}}$ and of $(\bar Z_n(g))_{g\in\mathcal{G}}$ converge to the fidis of the Gaussian process $(Z(g))_{g\in\mathcal{G}}$ defined in Theorem 3.2.1.

(b) If, in addition, $b_n(g)=b_n>0$ is the same for all $g\in\mathcal{G}$ and $n\in\mathbb{N}$ and for some positive sequence $(b_n)_{n\in\mathbb{N}}$, (3.2.6) holds for $g=g_{\max}$ and the conditions (i) or (ii) of Theorem 3.2.1 are fulfilled, then the processes $(Z_n(g))_{g\in\mathcal{G}}$ and $(\bar Z_n(g))_{g\in\mathcal{G}}$ converge weakly to $(Z(g))_{g\in\mathcal{G}}$ uniformly.

Note that usually $P(V_n(|g|)\neq 0)=O(p_n)$; in particular, this holds true if g has constant sign. Then the condition $m_n l_n P(V_n(|g|)\neq 0)=\mathrm{o}(r_n b_n^2(g)p_n)$ is fulfilled for the typical behavior of the sequences outlined in Table 3.1. As mentioned above, usually it suffices to consider just two different normalizing sequences, say $b_{n,1}$ and $b_{n,2}$. In this case, one may apply Theorem 3.2.2 separately to $(Z_n(g))_{g\in\mathcal{G}_i}$ for $i\in\{1,2\}$ with $\mathcal{G}_i:=\{g\in\mathcal{G}|b_n(g)=b_{n,i},\ \forall\, n\in\mathbb{N}\}$ to conclude that both processes are asymptotically tight. This in turn implies the asymptotic tightness of $(Z_n(g))_{g\in\mathcal{G}}$ and thus, in view of part (a), its convergence to $(Z(g))_{g\in\mathcal{G}}$. Hence, in fact the extra condition on b_n in part (b) does not further restrict the setting in the vast majority of applications.

The condition (3.2.6) is not always easy to check for $\delta > 0$ and could be weakened to $\delta = 0$. In this case, one has to assume condition (L) additionally and one can omit the weak technical condition $m_n l_n P(V_n(|g|) \neq 0) = o(r_n b_n^2(g) p_n)$. This leads to a slightly modified version of Theorem 3.2.2.

Theorem 3.2.3. *(a) Suppose (A), (A2), (MX), (D0) and (C) are met and $g_{\max}$ is not necessarily bounded. In addition, assume condition (L) is satisfied and*

$$E\bigg[\Big(\sum_{i=1}^{r_n} |g(W_{n,i})|\Big)^2\bigg] = O\Big(\frac{p_n b_n^2(g)}{m_n}\Big), \quad \forall\, g \in \mathcal{G}. \tag{3.2.7}$$

Then, the conditions (3.1.4) and (B) are satisfied and the fidis of $(Z_n(g))_{g\in\mathcal{G}}$ and of $(\bar{Z}_n(g))_{g\in\mathcal{G}}$ converge to the fidis of the Gaussian process $(Z(g))_{g\in\mathcal{G}}$ defined in Theorem 3.2.1.

(b) If, in addition, $g_{\max} = \sup_{g\in\mathcal{G}} |g|$ is measurable, $b_n(g) = b_n > 0$ is the same for all $g \in \mathcal{G}$ and $n \in \mathbb{N}$, (3.2.7) holds for $g = g_{\max}$ and the conditions (i) or (ii) of part (b) in Theorem 3.2.1 are fulfilled, then the processes $(Z_n(g))_{g\in\mathcal{G}}$ and $(\bar{Z}_n(g))_{g\in\mathcal{G}}$ converge weakly to $(Z(g))_{g\in\mathcal{G}}$ uniformly.

Conditions (3.2.4) and (C) (or, in the general setting, (Δ), (L) and (C)) are not always easy to check. In fact, (Δ) (or alternatively (3.1.4)) in particular causes difficulties in the application. Therefore, we will state three lemmas, which can be used to verify these conditions under some stronger assumptions in the sliding blocks setting. For sliding blocks sums, the following criterion is often useful to verify condition (3.2.4) and thereby (Δ):

(S) For all $g \in \mathcal{G}$ and $n \in \mathbb{N}$ one has

$$\sum_{k=1}^{r_n} P\left(g(W_{n,1}) \neq 0, g(W_{n,k}) \neq 0\right) = \mathrm{O}\Big(\frac{p_n b_n(g)^2}{n}\Big).$$

Lemma 3.2.4. *Suppose $g_{\max}$ is bounded. If condition (S) is satisfied, then (3.2.4) holds. In particular, (3.1.4) is satisfied.*

The statement of Lemma 3.2.4 can be reformulated with a condition based on conditional probabilities. For this, the additional assumption $r_n q_{g,n} = \mathrm{O}\left(p_n\right)$ is necessary. From the definition of p_n and $q_{\mathcal{G},n}$, it directly follows that

$$\begin{aligned} p_n &= P\left(\exists g \in \mathcal{G} : V_n(g) = \frac{1}{b_n(g)} \sum_{i=1}^{r_n} g(W_{n,i}) \neq 0\right) \\ &\leq P\Big(\exists g \in \mathcal{G}, i \in \{1, ..., r_n\} : g(W_{n,i}) \neq 0\Big) = \mathrm{O}(r_n q_{\mathcal{G},n}). \end{aligned}$$

If $g \geq 0$, the inequality becomes an equality. The assumption $r_n q_{\mathcal{G},n} = \mathrm{O}(p_n)$ therefore seems quite natural and reasonable.

(S*) For all $g \in \mathcal{G}$, $n \in \mathbb{N}$ and $k \in \{1, ..., r_n\}$ there exists

$$e_{g,n}(k) \geq P\left(g(W_{n,k}) \neq 0 \mid g(W_{n,1}) \neq 0\right)$$

such that $\sum_{k=1}^{r_n} e_{g,n}(k) = \mathrm{O}\left(b_n(g)^2/m_n\right)$.

Lemma 3.2.5. *Suppose $g_{\max}$ is bounded. If condition (S*) is satisfied and $r_n q_{g,n} = O(p_n)$ for all $g \in \mathcal{G}$, then* (3.2.4) *holds. In particular,* (3.1.4) *is satisfied.*

Example. This result somehow generalizes the proof of condition (C) in Drees et al. (2015). In the cited paper it is $p_n = r_n q_{g,n}$, $b_n(g) = \sqrt{m_n}$ for all $g \in \mathcal{G}$ and $g(W_{n,1}) \neq 0$ is equivalent to $X_{n,1} \neq 0$. With these conditions, (S*) matches the condition in Drees et al. (2015). ◇

Under some more restrictive conditions, but still quite general, one can show, that the modified condition (S*) also implies condition (C).

Lemma 3.2.6. *Suppose $g_{\max}$ is bounded and $\lim_{n\to\infty}(r_n q_{g,n})/p_n$ exists for all $g \in \mathcal{G}$ and let condition (S*) be satisfied. Assume $\lim_{n\to\infty} \sum_{k=1}^{r_n} e_{g,n}(k) = \sum_{k=1}^{\infty} \lim_{n\to\infty} e_{g,n}(k) < \infty$ and $b_n(g) = \sqrt{m_n}$ for all $g \in \mathcal{G}$.*

(a) If $E[g(W_{n,k})h(W_{n,1}) | h(W_{n,1}) \neq 0]$ converges for all $g, h \in \mathcal{G}$ and all $k \in \mathbb{N}$, then condition (C) is satisfied with covariance function given by

$$\left(\lim_{n\to\infty} \frac{q_{h,n} r_n}{p_n}\right) \sum_{k\in\mathbb{Z}} \lim_{n\to\infty} E[g(W_{n,0})h(W_{n,k}) \mid g(W_{n,0}) \neq 0].$$

(b) If $(X_t)_{t\in\mathbb{Z}}$ is regularly varying, $X_{n,i} = (X_i/u_n)\mathbb{1}_{\{\|X_i\|>u_n\}}$, $P(g(W_{n,1}) \neq 0)$ is of the same order as $P(\|X_1\| > u_n)$, and $g(W_{n,1}) \neq 0$ implies $\|X_1\| > u_n$, then the limit can be specified in terms of the tail process $Y = (Y_t)_{t\in\mathbb{Z}}$ as

$$\left(\lim_{n\to\infty} \frac{q_{h,n} r_n}{p_n} \frac{P(\|X_1\| > u_n)}{P(g(W_{n,1}) \neq 0)}\right) \sum_{k\in\mathbb{Z}} E[g((Y_t)_{t\in\mathbb{Z}})h((Y_{t+k})_{t\in\mathbb{Z}})].$$

The condition that $P(g(W_{n,1}) \neq 0)$ is of the same order as $P(\|X_1\| > u_n)$ and that $g(W_{n,1}) \neq 0$ implies $\|X_1\| > u_n$ is a real restriction. However, this lemma covers functions g including $\mathbb{1}_{\{\|X_1\|>u_n\}}$. In particular, so-called runs estimators and sliding blocks estimators with such indicators are still covered. Some examples for the application of this lemma are given by the runs estimators introduced in Chapter 4 and the estimators in Chapter 5. The advantage of the previous lemma is that one does not have to treat the covariance of sums of length r_n of sliding blocks, but one can consider only the convergence of expected values of single, shifted sliding blocks. With this, we close the discussion of the conditions for the sliding blocks limit theorems.

3.3 Sliding versus disjoint blocks

The previous section was devoted to general limit theorems for sliding blocks statistics. In this section, we want to compare the asymptotic variance of a sliding blocks statistic for a single functional g with that of the corresponding disjoint blocks statistic.

As mentioned in the beginning of Chapter 3, in extreme value statistics, one may average either statistics $g(W_{n,is_n+1})$, $0 \le i \le \lfloor n/s_n \rfloor - 1$, of disjoint blocks or statistics $g(W_{n,i})$, $1 \le i \le n - s_n + 1$, of overlapping sliding blocks. The main difference is that sliding blocks use much more data, but these blocks are also much more dependent, so it is unclear which method is more advantageous. It has been suggested in the literature that the latter approach may often be more efficient, see, e.g., Beirlant et al. (2004), p. 390, for such a statement about blocks estimators of the extremal index. However, the asymptotic performance of both approaches has been compared only for a couple of estimators, while general results showing the superiority of the sliding blocks estimators are not yet known in the peak-over-threshold (POT) setting. Robert et al. (2009) have shown such an advantage first in the literature. They proved that for some specific type of estimators of the extremal index (a different estimator than considered in Chapter 4) the version using sliding blocks has a strictly smaller asymptotic variance than the one based on disjoint blocks, while the bias is asymptotically the same. In a block maxima setting, Zou et al. (2021) proved that, under quite general conditions, an estimator of the extreme value copula of multivariate stationary time series is more efficient if it is based on sliding rather than disjoint blocks. The same observation in the block maxima approach has been made in Bücher and Segers (2018*a*), Bücher and Segers (2018*b*) for the maximum likelihood estimator of the parameters of a Fréchet distribution based on maxima of sliding or disjoint blocks of a stationary time series with marginal distribution in the maximum domain of attraction of this Fréchet distribution. Also, Bücher and Jennessen (2020*b*) observed in the block maxima setting for their blocks estimators for the limiting cluster size distribution that the sliding blocks estimator outperforms the disjoint blocks version.

Apart from the mentioned examples, so far, the consideration of sliding blocks is not very common in the literature. This could be due to the fact that there is not much theory for the asymptotic analysis of sliding blocks. E.g. Northrop (2015) did not consider any asymptotics for his sliding blocks estimator due to the complex theory. However, now the theory in Section 3.2 gives a new tool to derive such asymptotics in the POT setting.

We will derive a general result for the comparison of the performance of sliding and disjoint blocks statistics in the POT setting. The results are the counterpart to a result of Zou et al. (2021), who recently established a general result including conditions when sliding blocks are better than disjoint blocks in the block maxima setting.

For this purpose we consider the estimation of a quantity ξ which depends on the time series $(X_t)_{t\in\mathbb{Z}}$. We will compare the asymptotic variance of the disjoint blocks estimator with the asymptotic variance of the corresponding sliding blocks estimator. The requirement for this is that both estimators have asymptotic normal distributions, or at least belong to the same scale family of distributions and are unbiased. We will focus on the case of asymptotic normal distributions. For the disjoint blocks estimator such a limit distribution can be derived with the abstract setting from Section 3.1. For the sliding blocks estimator such a limit distribution can be derived with the theory in Section 3.2. The use of the theory from Chapter 3 allows the derivation of comparable conditions for the asymptotic normality of both estimators.

Note that, here, we use a different parametrization of the normalization constants, partly because the probability p_n used in the normalization above refers to the whole process and seems inappropriate in the present context, partly to facilitate the comparison of the asymptotic variances, a more detailed explanation is given below.

We start with the consideration of the disjoint blocks statistic

$$T_n^d(g) := \frac{1}{nv_n a_n} \sum_{i=1}^{\lfloor n/s_n \rfloor} g(X_{n,(i-1)s_n+1}, ..., X_{n,is_n-1}). \tag{3.3.1}$$

Here we define $v_n := P(X_{n,0} \neq 0) \to 0$. Recall, that a typical choice for $X_{n,i}$ would be $X_{n,i} := (X_i/u_n)\mathbb{1}_{\{\|X_i\|>u_n\}}$ for a sequence u_n with $v_n = P(\|X_1\| > u_n) \to 0$, but the $X_{n,i}$ are more general here.

The sequence a_n is an additional normalization sequence which ensures that the expected value $E[T_n^d(g)]$ converges in $\mathbb{R}$. This sequence a_n increases the flexibility of the result presented here. In some application, e.g. for the extremal index in Section 4.2, we have $a_n = 1$ for all $n \in \mathbb{N}$, but $a_n \to \infty$ is also possible.

The corresponding sliding blocks statistic is

$$T_n^s(g) := \frac{1}{nv_n s_n a_n} \sum_{i=1}^{n-s_n+1} g(X_{n,i}, ..., X_{n,i+s_n-1}). \tag{3.3.2}$$

Again, we will use the notation $W_{n,i} = (X_{n,i}, ..., X_{n,i+s_n-1})$ from (3.0.1). Both statistics shall be estimators for the same value $\xi \in \mathbb{R}$, which depends on the distribution of the time series $(X_t)_{t\in\mathbb{Z}}$.

The normalization of $T_n^s(g)$ is larger than the normalization of $T_n^d(g)$ by the factor s_n. This is necessary such that one can expect convergence of the expectation. If the expectation of the sliding blocks statistic converges to some value $\xi \in \mathbb{R}$, i.e.

$$E\left[T_n^s(g)\right] = \frac{1}{s_n v_n a_n} E\left[g(W_{n,1})\right] \frac{n - s_n + 1}{n} \to \xi, \tag{3.3.3}$$

then also

$$E[T_n^d(g)] = \frac{1}{v_n a_n} E\left[g(W_{n,1})\right] \frac{\lfloor n/s_n \rfloor}{n} = \frac{1}{s_n v_n a_n} E\left[g(W_{n,1})\right] \frac{\lfloor n/s_n \rfloor s_n}{n} \to \xi.$$

Moreover, the difference between both expectations is asymptotically negligible if

$$\left|E\left[T_n^d(g) - T_n^s(g)\right]\right| = \frac{1}{s_n v_n a_n}\left|E\left[g(W_{n,1})\right]\right| \cdot \left|\frac{s_n}{n}\left\lfloor\frac{n}{s_n}\right\rfloor - \frac{n-s_n+1}{n}\right| = \mathrm{O}(s_n/n) \quad (3.3.4)$$

is of smaller order than $(nv_n)^{-1/2}$ (cf. (3.3.5), (3.3.6)), which, in particular, holds under the basic condition $s_n v_n \to 0$. Thus, if the bias for one of the statistics is negligible, then $T_n^s(g)$ will be a more efficient (i.e. has a smaller asymptotic mean squared error) estimator than $T_n^d(g)$ if its asymptotic variance is smaller.

If the normalization of the sliding blocks statistics was of another order than s_n times the normalization of the disjoint blocks statistic, then the expectation of one estimator would no longer converge to ξ and both statistics would estimate different things. Since we want to compare the corresponding statistics for the same estimation problem, this normalization is the only possibility.

For the comparison, we need the asymptotic variances of both statistics. These asymptotic variances are given by the variances of the corresponding asymptotic normal distributions. For the sliding blocks statistics, the weak convergence

$$\begin{aligned}&\sqrt{nv_n}(T_n^s(g) - E[T_n^s(g)]) \\ &= \frac{1}{\sqrt{nv_n}s_n a_n}\sum_{i=1}^{n-s_n}(g(W_{n,i}) - E[g(W_{n,1})]) \to \mathcal{N}(0, c^{(s)})\end{aligned} \quad (3.3.5)$$

can be proved with Theorem 3.2.1, part (a). In the setting of Section 3.2, we have $m_n = \lfloor (n - s_n + 1)/r_n \rfloor$, $p_n = P(\sum_{i=1}^{r_n} g(W_{n,i}) \neq 0)$, and we choose $b_n(g) = b_n = (nv_n/p_n)^{1/2} a_n s_n$ with $v_n = P(X_{n,1} \neq 0)$. The precise conditions are stated in the next corollary.

Corollary 3.3.1. *Suppose the following conditions are fulfilled:*

(i) $(X_{n,i})_{1 \le i \le n}$ *is stationary for all* $n \in \mathbb{N}$.

(ii) The sequences $l_n, r_n, s_n \in \mathbb{N}$, a_n *and* p_n *satisfy* $s_n \le l_n = o(r_n)$, $r_n = o(n)$, $p_n \to 0$, $r_n = o\left(\sqrt{nv_n}s_n a_n\right)$ *and* $(n/r_n)\beta^X_{n,l_n - s_n} \to 0$.

(iii) g *is measurable and bounded.*

(iv) $E\left[\left(\sum_{j=1}^{r_n} \mathbb{1}_{\{g(W_{n,j}) \neq 0\}}\right)^2\right] = O\left(r_n v_n a_n^2 s_n^2\right)$.

(v) $c^{(s)}$ *as defined in* (3.3.8) *below exists in* $[0, \infty)$.

Then, convergence (3.3.5) *holds.*

To prove this corollary, simply insert the setting here into the conditions of Theorem 3.2.1, part (a), no additional calculations are required. One may drop the assumption that g is bounded if condition (iv) is adapted in the same way as in Theorem 3.2.2.

For the disjoint blocks statistic,

$$
\begin{aligned}
&\sqrt{nv_n}\left(T_n^d(g) - E[T_n^d(g)]\right) \\
&= \frac{1}{\sqrt{nv_n}a_n} \sum_{i=1}^{\lfloor n/s_n \rfloor} (g(W_{n,(i-1)s_n+1}) - E[g(W_{n,1})]) \to \mathcal{N}(0, c^{(d)})
\end{aligned} \tag{3.3.6}
$$

holds under suitable conditions thanks to Theorem 3.1.4. In the corresponding setting of this theorem we use $V_{n,i}(g) = (p_n/(nv_n a_n^2))^{1/2} \sum_{j=1}^{r_n/s_n} g(W_{n,(j-1)s_n+(i-1)r_n+1})$, $1 \le i \le m_n$ where $p_n = P(\sum_{j=1}^{r_n/s_n} g(W_{n,(j-1)s_n+1}) \neq 0)$, assuming that r_n is a multiple of s_n. Moreover, we choose a sequence l_n, $n \in \mathbb{N}$, of multiples of s_n and we define $\tilde{V}_{n,i}(g) = (p_n/(nv_n a_n^2))^{1/2} \sum_{j=1}^{(r_n-l_n)/s_n} g(W_{n,(j-1)s_n+(i-1)r_n+1})$ as the approximating sums.
The precise conditions under which (3.3.6) holds in the setting of Section 3.1.1 are given in the next corollary, again the proof is only a direct application of the theory in Section 3.1.1.

Corollary 3.3.2. *Suppose that, in addition to (i) and (iii) of Corollary 3.3.1, the following conditions are satisfied:*

(ii)* *For the sequences* $l_n, r_n, s_n \in \mathbb{N}$ *we have that* $l_n = o(r_n)$, $r_n = o(n)$, l_n *and* r_n *are multiples of* s_n, $p_n \to 0$, $r_n = o\left(\sqrt{nv_n}\right)$ *and* $(n/r_n)\beta^X_{n,l_n-s_n} \to 0$.

(iv)* $E\left[\left(\sum_{j=1}^{l_n/s_n} \mathbb{1}_{\{g(W_{n,j})\neq 0\}}\right)^2\right] = o\left(r_n v_n a_n^2\right)$.

(v)* $c^{(d)}$ *as defined in* (3.3.7) *below exists in* $[0, \infty)$.

(vi)* $E\left[\sum_{j=1}^{r_n/s_n} \left(g(W_{n,(j-1)s_n+1}) - Eg(W_{n,1})\right)^2 \mathbb{1}_{\{|\sum_{j=1}^{r_n/s_n}(g(W_{n,(j-1)s_n+1})-Eg(W_{n,1}))|>\varepsilon\sqrt{nv_n}a_n\}}\right]$
$= o(r_n v_n a_n^2)$ *for all* $\varepsilon > 0$.

Then, convergence (3.3.6) *holds.*

Condition (iv*) could be weakened to condition (Δ). Alternatively, one could prove the asymptotic normality of $T_n^d(g)$ using Theorem 2.3 of Drees and Rootzén (2010) with r_n replaced by s_n, but the following representation of the asymptotic variance $c^{(d)}$ simplifies the comparison with $c^{(s)}$.
Recall that the sequence r_n is only needed in the proofs which use the *big blocks, small blocks* technique, i.e. it has no operational meaning, but it must be chosen such that the conditions of Theorem 3.2.1 and Theorem 3.1.4, respectively, are met. According to the next lemma, we may assume w.l.o.g. that r_n is a multiple of s_n, where the multiplicity depends on n. Note that r_n/s_n must tend to ∞ if Theorem 3.2.1 shall be applied. The assumption $p_n \asymp r_n v_n$ is not strong, it holds true in all examples considered in this thesis and all examples known to the author.

Lemma 3.3.3. *Suppose $p_n \asymp r_n v_n$ and the conditions of Theorem 3.2.1, Theorem 3.2.2 and Theorem 3.1.4 with $V_{n,i}(g) = (p_n/(nv_n a_n^2))^{1/2} \sum_{j=1}^{\lfloor r_n/s_n \rfloor} g(W_{n,(j-1)s_n+(i-1)r_n+1})$ are met by some sequence r_n. Then, these conditions are also fulfilled for $r_n^* := \lfloor r_n/s_n \rfloor s_n$.*

If the convergences (3.3.5) and (3.3.6) hold by the Corollaries 3.3.1 and 3.3.2, respectively, then the asymptotic covariances can be calculated as

$$c^{(d)} = \lim_{n\to\infty} \frac{1}{r_n v_n a_n^2} Var\left(\sum_{i=1}^{r_n/s_n} g(W_{n,is_n+1})\right), \tag{3.3.7}$$

$$c^{(s)} = \lim_{n\to\infty} \frac{1}{r_n v_n s_n^2 a_n^2} Var\left(\sum_{i=1}^{r_n} g(W_{n,i})\right). \tag{3.3.8}$$

In this section, we changed the parametrization of the normalization. Namely, we use $\sqrt{nv_n} s_n a_n$ instead of $\sqrt{p_n} b_n(g)$ for the sliding blocks in (3.3.5) and $\sqrt{nv_n} a_n$ instead of $\sqrt{p_n m_n} a_n$ for the disjoint blocks in (3.3.6). This has mainly two reasons: First, the normalization applied here increases the comparability of the normalization for $T_n^s(g)$ and $T_n^d(g)$. With this notation the difference in the normalizations is obvious. Second, $T_n^s(g)$ and $T_n^d(g)$ are only pseudo-estimators, since they depend on v_n which in turn depends on $(X_t)_{t\in\mathbb{Z}}$. Later on, we will replace this v_n by an estimator. With the parametrization of the previous section, we would have to estimate p_n (or $q_{g,n}$) which depends on g. This would make it difficult to compare the result for different g and it would make the analysis of the variances much more sophisticated.

Note that for the disjoint blocks it would also be possible, and somehow more natural, to apply the theory of Section 3.1.1 with $V_{n,i}(g) = (p_n/(nv_n a_n^2))^{-1/2} g(W_{n,(i-1)s_n+1})$. But in this case, we had $p_n = \mathrm{O}(s_n v_n)$ and $m_n = \lfloor n/s_n \rfloor$. For the comparability of the convergence results for the sliding and disjoint blocks statistics, we have chosen the above representation for the disjoint blocks, such that p_n and m_n have the same meaning for both statistics. Moreover, this demonstrates that an unified framework is used, with comparable handling for disjoint and sliding blocks.

The following lemma shows that the asymptotic variance of the sliding blocks statistic is never greater than that of the disjoint blocks counterpart.

Lemma 3.3.4. *If conditions (A),* (3.3.7) *and* (3.3.8) *hold and $r_n/s_n \in \mathbb{N}$ for all $n \in \mathbb{N}$, then $c^{(s)} \le c^{(d)}$.*

The conditions of the lemma are not enough to establish the convergences for the disjoint and sliding blocks statistics in (3.3.6) and (3.3.5), i.e. under the conditions of the lemma $c^{(s)}$ and $c^{(d)}$ are not necessarily the asymptotic variances. If, in addition to the conditions of the lemma, the conditions of Corollary 3.3.1 and Corollary 3.3.2 hold, then $c^{(s)}$ and $c^{(d)}$ are the asymptotic variances of the disjoint and sliding blocks estimator. The condition that $c^{(d)}$ and $c^{(s)}$ exist is exactly the condition (C) in the previous section.

The statistics in (3.3.1) and (3.3.2) include the factor v_n^{-1}. This is an important normalization, since we consider only extreme events with $v_n \to 0$. Without this normalization, the estimator would converge to 0 in probability. The statistics considered above are not actual estimators, since the probability v_n that a single observation $X_{n,1}$ does not vanish is typically unknown. This v_n must be estimated as well, e.g. by the empirical version $\sum_{i=1}^n \mathbb{1}_{\{X_{n,i}\neq 0\}}/n$. In what follows, we thus analyze versions of our statistics where v_n is replaced with this empirical version. This results in the disjoint blocks estimator

$$\tilde{T}_n^d(g) := \frac{a_n^{-1}\sum_{i=1}^{\lfloor n/s_n\rfloor} g(X_{n,(i-1)s_n+1},...,X_{n,is_n})}{\sum_{i=1}^{n-s_n+1}\mathbb{1}_{\{X_{n,i}\neq 0\}}} \tag{3.3.9}$$

and the sliding blocks estimator

$$\tilde{T}_n^s(g) := \frac{(s_n a_n)^{-1}\sum_{i=1}^{n-s_n+1} g(X_{n,i},...,X_{n,i+s_n-1})}{\sum_{i=1}^{n-s_n+1}\mathbb{1}_{\{X_{n,i}\neq 0\}}} \tag{3.3.10}$$

for the estimation of ξ. Note that we chose the denominators to be the same for both statistics.

In order to prove convergence of these estimators, one needs the joint convergence of the numerator and denominator. This can again be concluded from Theorem 3.1.4 or part (a) of Theorem 3.2.1, respectively, now applied with $\mathcal{G} = \{g,h\}$ and $h(x_1,...,x_{s_n}) = \mathbb{1}_{\{x_1\neq 0\}}$. For the disjoint blocks estimator, we obtain

$$\begin{pmatrix} (\sqrt{nv_n}a_n)^{-1}\sum_{i=1}^{\lfloor n/s_n\rfloor}(g(W_{n,(i-1)s_n+1}) - E[g(W_{n,1})]) \\ (nv_n)^{-1/2}\sum_{i=1}^{n-s_n+1}\left(\mathbb{1}_{\{X_{n,i}\neq 0\}} - P(X_{n,1}\neq 0)\right)\end{pmatrix} \xrightarrow{w} \begin{pmatrix} Z^D \\ Z^N \end{pmatrix} \sim \mathcal{N}_2\left(0, \begin{pmatrix} c^{(d)} & c^{(d,v)} \\ c^{(d,v)} & c^{(v)}\end{pmatrix}\right) \tag{3.3.11}$$

under suitable conditions. Sufficient conditions are those of Corollary 3.3.2, provided $c^{(d,v)}$ and $c^{(v)}$ exists. For the sliding blocks estimator,

$$\begin{pmatrix} (\sqrt{nv_n}s_n a_n)^{-1}\sum_{i=1}^{n-s_n+1}(g(W_{n,i}) - E[g(W_{n,1})]) \\ (nv_n)^{-1/2}\sum_{i=1}^{n-s_n+1}\left(\mathbb{1}_{\{X_{n,i}\neq 0\}} - P(X_{n,1}\neq 0)\right)\end{pmatrix} \xrightarrow{w} \begin{pmatrix} Z^S \\ Z^N \end{pmatrix} \sim \mathcal{N}_2\left(0, \begin{pmatrix} c^{(s)} & c^{(s,v)} \\ c^{(s,v)} & c^{(v)}\end{pmatrix}\right) \tag{3.3.12}$$

holds under the conditions of Corollary 3.3.1, provided that $c^{(v)}$ and $c^{(s,v)}$ exist. Note that the same result holds if the sum in the second component goes up to n instead of $n - s_n + 1$ (cf. (3.2.5)).

Here, the asymptotic covariances are given by

$$c^{(d,v)} := \lim_{n\to\infty}\frac{1}{r_n v_n a_n} Cov\Big(\sum_{j=1}^{r_n/s_n} g(W_{n,(j-1)s_n+1}), \sum_{i=1}^{r_n}\mathbb{1}_{\{\|X_{n,i}\neq 0\}}\Big),$$

$$c^{(s,v)} := \lim_{n\to\infty} \frac{1}{r_n v_n s_n a_n} Cov\Big(\sum_{i=1}^{r_n} g(W_{n,i}), \sum_{i=1}^{r_n} \mathbb{1}_{\{X_{n,i}\neq 0\}}\Big),$$
$$c^{(v)} := \lim_{n\to\infty} \frac{1}{r_n v_n} E\Big[\Big(\sum_{j=1}^{r_n} \mathbb{1}_{\{X_{n,j}\neq 0\}}\Big)^2\Big].$$

The asymptotic normality of the estimators (3.3.9) and (3.3.10) follows from the following Lemma, if (3.3.11) and (3.3.12) and an appropriate bias condition holds.

Lemma 3.3.5. *Suppose $nv_n \to \infty$ and that the weak convergence*

$$\sqrt{nv_n}\left(\begin{pmatrix} Z_n^1 \\ Z_n^2 \end{pmatrix} - \begin{pmatrix} E[Z_n^1] \\ E[Z_n^2] \end{pmatrix}\right) \xrightarrow{w} \begin{pmatrix} Z^1 \\ Z^2 \end{pmatrix} \sim \mathcal{N}_2\left(0, \begin{pmatrix} Var(Z^1) & Cov(Z^1,Z^2) \\ Cov(Z^1,Z^2) & Var(Z^2) \end{pmatrix}\right)$$

holds for some processes (Z_n^1, Z_n^2) , $n \in \mathbb{N}$ and a limit process (Z^1, Z^2). Let $E[Z_n^2] \to b$ hold for a constant $b \in \mathbb{R}\backslash\{0\}$ and assume that the bias condition

$$\sqrt{nv_n}\Big(E[Z_n^1] - \xi E[Z_n^2]\Big) \to 0$$

holds for some constant $\xi \in \mathbb{R}$. Then,

$$\sqrt{nv_n}\Big(\frac{Z_n^1}{Z_n^2} - \xi\Big) \xrightarrow{w} \frac{1}{b}(Z^1 - \xi Z^2)$$
$$\sim \mathcal{N}\Big(0, \frac{1}{b^2}\Big(Var(Z^1) + \xi^2 Var(Z^2) - 2\xi Cov(Z^1, Z^2)\Big)\Big).$$

Thus, if, in addition to the conditions of Corollary 3.3.2, $c^{(v)}$ and $c^{(d,v)}$ exists and the bias condition

$$E[g(W_{n,1})]/s_n v_n a_n - \xi = \mathrm{o}\Big((nv_n)^{-1/2}\Big)$$

is satisfied, then Lemma 3.3.5 implies the asymptotic normal distribution of the estimation error for the disjoint blocks estimator (3.3.9):

$$\sqrt{nv_n}\Big(\tilde{T}_n^d(g) - \xi\Big) \xrightarrow{w} \mathcal{N}(0, \tilde{c}^{(d)})$$

with $\tilde{c}^{(d)} := c^{(d)} + \xi^2 c^{(v)} - 2\xi c^{(d,v)}$.

Analogously, under the same bias condition, Lemma 3.3.5 implies the asymptotic normality of the sliding blocks estimator (3.3.9) if the conditions of Corollary 3.3.1 are fulfilled and $c^{(v)}$ and $c^{(s,v)}$ exist:

$$\sqrt{nv_n}\Big(\tilde{T}_n^s(g) - \xi\Big) \xrightarrow{w} \mathcal{N}(0, \tilde{c}^{(s)})$$

with $\tilde{c}^{(s)} := c^{(s)} + \xi^2 c^{(v)} - 2\xi c^{(s,v)}$.

To this end, we let $Z_n^2 = (nv_n)^{-1}\sum_{i=1}^{n-s_n+1} \mathbb{1}_{\{X_{n,i}\neq 0\}}$ with $E[Z_n^2] = (n-s_n)v_n/(nv_n) \to 1$ and $Z_n^1 = (ns_n v_n a_n)^{-1}\sum_{i=1}^{n-s_n+1} g(W_{n,i})$ with $E[Z_n^1] = (n-s_n)(ns_n v_n a_n)E[g(W_{n,1})]$ in the sliding blocks case, for the disjoint blocks case analogously.

The following theorem states that, under rather mild conditions, again the asymptotic variance of the sliding blocks estimator (3.3.10) is not greater than the variance of the disjoint blocks estimator (3.3.9), at least if g has a uniform sign. This is the central result about the comparison of sliding and disjoint blocks estimators.

Theorem 3.3.6. *Suppose the conditions of Lemma 3.3.4 are satisfied,* (3.3.3) *holds, the function* g *is bounded and does not change its sign,* $s_n = o(r_n a_n)$ *and* $s_n v_n \to 0$. *In addition, assume there exists a sequence* $k_n = o(r_n a_n)$ *of natural numbers such that the* β*-mixing coefficients of* $(X_{n,i})_{1\le i\le n}$ *satisfy* $\sum_{i=k_n}^{r_n} \beta_{n,i}^X = o(r_n v_n a_n)$. *Then* $\tilde{c}^{(s)} \le \tilde{c}^{(d)}$.

If the above mentioned conditions under which the asymptotic normality of $\tilde{T}_n^d(g)$ and $\tilde{T}_n^s(g)$ are fulfilled, then $\tilde{c}^{(d)}$ and $\tilde{c}^{(s)}$ are the asymptotic variances. The condition that $c^{(d,v)}$, $c^{(s,v)}$ and $c^{(v)}$ exist is the condition (C) in the previous section.
The β-mixing condition in the theorem is satisfied e.g. if the β-mixing coefficients decrease exponentially fast, i.e. geometrically with $\beta_{n,k}^X \le t\eta^k$ for some constants $\eta \in (0,1)$ and $t > 0$. In this case, provided $\log n = \mathrm{o}(r_n a_n)$, the sequence $k_n = \lfloor c\log(n)\rfloor$ with sufficiently large constant $c > 0$ fulfills the conditions of Theorem 3.3.6. This is e.g. the case for some solutions to stochastic recurrence equations (cf. Doukhan (1994), Corollary 2.4.1), see also the β-mixing arguments in Section 5.5.2.
This theorem shows that a sliding blocks estimator of the form above, where the unknown v_n is estimated by $\sum_{i=1}^n \mathbb{1}_{\{X_{n,i}\neq 0\}}/n$, is always at least as efficient as the corresponding disjoint blocks estimator. This result in this general setting seems to be shown here for the first time. One implication of this is that, for application with disjoint blocks estimators, usually, the corresponding sliding blocks estimator should be preferred.
In fact, the difference $\tilde{c}^{(d)} - \tilde{c}^{(s)}$ for the true estimators $\tilde{T}_n^d(g)$ and $\tilde{T}_n^s(g)$ is the same as for the pseudo estimators $T_n^d(g)$ and $T_n^s(g)$, i.e $\tilde{c}^{(d)} - \tilde{c}^{(s)} = c^{(d)} - c^{(s)}$.

Corollary 3.3.7. *Suppose the conditions of Theorem 3.3.6 are fulfilled. Then,* $\tilde{c}^{(d)} - \tilde{c}^{(s)} = c^{(d)} - c^{(s)} \ge 0$.

The previous corollary shows that there is generally only one possible cause for the smaller variance of the sliding blocks estimator, namely that the variance of the numerator $c^{(s)}$ is smaller than $c^{(d)}$. The covariance between numerator and denominator $c^{(s,v)}$ and $c^{(d,v)}$ are identical for sliding and disjoint blocks estimators. An intuitive explanation for the smaller variance could be that, for disjoint blocks, the variance $c^{(d)}$ consists only of the variance of a single block $g(W_{n,1})$. In the case of sliding blocks, the variance $c^{(s)}$ of the numerator is essentially the mean of the covariances of all overlapping blocks, i.e. the covariances of $g(W_{n,1})$ with $g(W_{n,k})$, $k = 1, ..., s_n$. Since the overlap gets smaller with increasing k, it seems plausible that the covariance gets smaller with increasing k and, therefore, the mean value of these covariances is smaller than the variance of the disjoint blocks. This intuition seems to be the reason for the smaller sliding bocks variance in Robert et al. (2009) and Bücher and Segers (2018*a*).

So far we have considered a single function g. Indeed, one can even prove a multivariate version of Lemma 3.3.4. In the following corollary we want to generalize Lemma 3.3.4 to a finite family of functions $\mathcal{G}$. For this, we consider the Loewner order for the asymptotic covariance matrices, denoted by $\leq_L$. The Loewner order is a semi order defined on the vector space of the symmetric real-valued $n \times n$ matrices. It is defined by $A \leq_L B$ if and only if $x^\top(B-A)x \geq 0$ for all $x \in \mathbb{R}^m$, i.e. $(B-A)$ is positive semi-definite.
Fix some finite set $\mathcal{G}$ of functions g of the type considered before in this section, i.e. we assume $|\mathcal{G}| < \infty$. If all functions in $\mathcal{G}$ fulfill the conditions of Corollary 3.3.1 and 3.3.2, respectively, then

$$\begin{aligned}&\left(\sqrt{nv_n}(T_n^d(g) - E[T_n^d(g)])\right)_{g\in\mathcal{G}} \\ &= \left(\frac{1}{\sqrt{nv_n}a_n}\sum_{i=1}^{\lfloor n/s_n\rfloor}(g(W_{n,(i-1)s_n+1}) - E[g(W_{n,1})])\right)_{g\in\mathcal{G}} \xrightarrow{w} \mathcal{N}_{|\mathcal{G}|}(0, C^{(d)}).\end{aligned} \tag{3.3.13}$$

and

$$\begin{aligned}&(\sqrt{nv_n}(T_n^s(g) - E[T_n^s(g)]))_{g\in\mathcal{G}} \\ &= \left(\frac{1}{\sqrt{nv_n}s_n a_n}\sum_{i=1}^{n-s_n}(g(W_{n,i}) - E[g(W_{n,1})])\right)_{g\in\mathcal{G}} \xrightarrow{w} \mathcal{N}_{|\mathcal{G}|}(0, C^{(s)})\end{aligned} \tag{3.3.14}$$

with $C^{(d)} = (c^{(d)}(g,h))_{g,h\in\mathcal{G}}$, $C^{(s)} = (c^{(s)}(g,h))_{g,h\in\mathcal{G}}$, provided that

$$\begin{aligned}c^{(d)}(g,h) &:= \lim_{n\to\infty}\frac{1}{r_n v_n a_n^2}Cov\left(\sum_{i=1}^{\lfloor r_n/s_n\rfloor} g(W_{n,is_n+1}), \sum_{i=1}^{\lfloor r_n/s_n\rfloor} h(W_{n,is_n+1})\right), \\ c^{(s)}(g,h) &:= \lim_{n\to\infty}\frac{1}{r_n v_n s_n^2 a_n^2}Cov\left(\sum_{i=1}^{r_n} g(W_{n,i}), \sum_{i=1}^{r_n} h(W_{n,i})\right)\end{aligned}$$

for all $g, h \in \mathcal{G}$ exists. Under the given conditions, the convergence is an immediate consequence of Theorem 3.1.4 or Theorem 3.2.1, respectively. The conditions of the next corollary are similar to the conditions of Lemma 3.3.4, only adapted for $|\mathcal{G}| > 1$.

Corollary 3.3.8. *Assume that condition (A) holds, the limiting covariance matrices $C^{(d)}$ and $C^{(s)}$ exists, $r_n/s_n \in \mathbb{N}$ for all $n \in \mathbb{N}$ and $|\mathcal{G}| < \infty$. Then,*

$$C^{(s)} \leq_L C^{(d)}.$$

The conditions of the corollary do not imply the convergence for the disjoint and sliding blocks statistics in (3.3.13) and (3.3.14). For this, one would have to assume additionally the conditions of Corollary 3.3.2 and 3.3.1, respectively. In this case, the sliding blocks statistic is at least as efficient as the disjoint blocks analogue in terms of the asymptotic covariance matrix in the Loewner order.
Note that the assertion of the corollary is equivalent to the statement that for all linear

combinations h of functions in $\mathcal{G}$ the asymptotic variance of $T_n^s(h)$ is not greater than the corresponding asymptotic variance of $T_n^d(h)$.
This corollary cannot easily be generalized for the estimators with estimated v_n as in Theorem 3.3.6. In the proof of the theorem, a fundamental argument is based on the constant sign of the function g. For the Loewner order, the covariance inequality in the previous proof must be valid for all $w \in \mathbb{R}^{|\mathcal{G}|}$, thus also for some $w_i < 0$ and some $w_j > 0$. Then, $\tilde{g}_w \geq 0$ (or $\tilde{g}_w \leq 0$, see proofs) can no longer be ensured. Therefore, the corollary is not directly transferable for Theorem 3.3.6.
This concludes our general consideration of disjoint and sliding blocks statistics, and it also concludes our discussion of the abstract limit theorems. In the following chapters, we will apply the theory developed here to prove asymptotic statements for estimators of the extremal dependence of $(X_t)_{t\in\mathbb{Z}}$. In the next Chapter 4, we will consider sliding and disjoint blocks and runs estimators for cluster indexes as an exemplary application of the theory developed so far. In particular, we will see that, in these examples, the variances of sliding and disjoint blocks estimators are the same.

3.4 Proofs

In this section, all proofs of theorems, lemmas and corollaries of this chapter are given.

3.4.1 Proofs for Section 3.1.1

First, we prove Lemma 3.1.2 about the β-mixing condition.

Proof of Lemma 3.1.2. For this proof we use the following alternative representation of the β- mixing coefficients: Denote with $P_{l,n,k}^V$ the product measure on $\mathcal{B}_{n,1}^{V,l} \otimes \mathcal{B}_{n,l+k+1}^{V,m_n}$ with $P_{l,n,k}^V(A \times B) = P(A)P(B)$ for $A \in \mathcal{B}_{n,1}^{V,l}$ and $B \in \mathcal{B}_{n,l+k+1}^{V,m_n}$. Then,

$$\beta_{n,k}^V = \sup_{1\leq l\leq m_n-k-1} \sup_{D\in\mathcal{B}_{n,1}^{V,l}\otimes\mathcal{B}_{n,l+k+1}^{V,m_n}} |P(D) - P_{l,n,k}^V(D)|$$

(see Volkonskii and Rozanov (1959), p. 179 or Doukhan (1994), Section 1.1). Likewise we define $P_{l,n,k}^X$ as product measure on $\mathcal{B}_{n,1}^{X,l} \otimes \mathcal{B}_{n,l+k+1}^{X,n}$.
Due to the $(X_{n,(i-1)r_n+1}, ..., X_{n,ir_n+s_n-1})$-measurability of $V_{n,i}$ for all $1 \leq i \leq m_n$, we obtain $\mathcal{B}_{n,i}^{V,j} \subset \mathcal{B}_{n,(i-1)r_n+1}^{X,jr_n+s_n-1}$ for all $1 \leq i \leq j \leq m_n$. In particular, $\mathcal{B}_{n,1}^{V,l} \subset \mathcal{B}_{n,1}^{X,lr_n+s_n-1}$ and $\mathcal{B}_{n,l+k+1}^{V,m_n} \subset \mathcal{B}_{n,lr_n+s_n-k}^{X,n}$ for n large enough such that $s_n - k \leq kr_n + 1$. This inclusions lead to $P_{lr_n+s_n-1,n,k}^X|_{\mathcal{B}_{n,1}^{V,l}\otimes\mathcal{B}_{n,l+k+1}^{V,m_n}} = P_{l,n,k}^V$. Using this representation of $\beta_{n,k}^V$ leads to

$$\begin{aligned} m_n\beta_{n,k}^V &= m_n \sup_{1\leq l\leq m_n-k-1} \sup_{D\in\mathcal{B}_{n,1}^{V,l}\otimes\mathcal{B}_{n,l+k+1}^{V,m_n}} |P(D) - P_{l,n,k}^V(D)| \qquad (3.4.1) \\ &\leq m_n \sup_{1\leq l\leq m_n-k-1} \sup_{D\in\mathcal{B}_{n,1}^{X,lr_n+s_n-1}\otimes\mathcal{B}_{n,(l+k)r_n+1}^{X,m_nr_n+s_n-1}} |P(D) - P_{lr_n+s_n-1,n,k}^X(D)| \end{aligned}$$

$$\leq m_n \sup_{1\leq l\leq n-(kr_n-s_n)-1} \sup_{D\in\mathcal{B}^{X,l}_{n,1}\otimes\mathcal{B}^{X,n}_{n,l+(kr_n-s_n)+1}} |P(D)-P^X_{l,n,k}(D)|$$
$$= m_n\beta^X_{n,kr_n-s_n} \to 0. \qquad \square$$

Next, we prove Theorem 3.1.4. In a first step, we show that for the proof of convergence of the fidis of $(Z_n(g))_{g\in\mathcal{G}}$ it suffices to consider independent copies of $V_{n,i}$.

Lemma 3.4.1. *Suppose the conditions (A1), (V), (M$\tilde{V}$), (MX$_k$) for some $k\in\mathbb{N}$, $k\geq 2$, (D0) and (Δ) are satisfied. Let*

$$Z^*_n(g) := \frac{1}{\sqrt{p_n}}\sum_{i=1}^{m_n}\left(V^*_{n,i}(g)-E[V^*_{n,i}(g)]\right), \qquad g\in\mathcal{G},$$

where $V^*_{n,i}$ are independent copies of $V_{n,i}$, $1\leq i\leq m_n$. Then the fidis of $(Z_n(g))_{g\in\mathcal{G}}$ converge weakly if and only if the fidis of $(Z^*_n(g))_{g\in\mathcal{G}}$ converge weakly and if so, the limits coincide.

Proof of Lemma 3.4.1. Let $\Delta^*_{n,i}$ be independent copies of $\Delta_{n,i}$, so that $\Delta^*_{n,i}(g)\overset{d}{=}\Delta_n(g)$. The condition ($\Delta$) (i) and (ii) correspond to the conditions of Theorem 1 in Section IX.1 of Petrov (1975), if $X_{nk}=p_n^{-1/2}\Delta^*_{n,k}(g)$ is inserted there and if the stationarity of $\tilde{V}_{n,i}$ and $V_{n,i}$ is taken into account (which holds due to condition (V)). This law of large numbers returns, provided part (i) and (ii) of (Δ) hold,

$$\frac{1}{\sqrt{p_n}}\sum_{i=1}^{m_n}\left(\Delta^*_{n,i}(g)-E[\Delta^*_{n,i}(g)]\right)-\tau_n=\mathrm{o}_P(1), \qquad \forall\, g\in\mathcal{G}, \tag{3.4.2}$$

with

$$\tau_n := \frac{m_n}{\sqrt{p_n}}E\Big[(\Delta_n(g)-E[\Delta_n(g)])\mathbb{1}_{\{|\Delta_n(g)-E[\Delta_n(g)]|\leq\tau\sqrt{p_n}\}}\Big] \tag{3.4.3}$$

for some $\tau>0$. Due to condition (Δ) (iii), we have $\tau_n\to 0$, which leads to

$$\frac{1}{\sqrt{p_n}}\sum_{i=1}^{m_n}\left(\Delta^*_{n,i}(g)-E[\Delta^*_{n,i}(g)]\right)=\mathrm{o}_P(1), \qquad \forall\, g\in\mathcal{G}. \tag{3.4.4}$$

Analogously, the statement follows for the partial sums that contain only every k-th summand, where k is the number for which condition (MX$_k$) is satisfied., i.e.

$$\frac{1}{\sqrt{p_n}}\sum_{j=1}^{m_{n,k,i}}\left(\Delta^*_{n,jk-i}(g)-E[\Delta^*_{n,jk-i}(g)]\right)=\mathrm{o}_P(1), \qquad \forall\, g\in\mathcal{G}, \tag{3.4.5}$$

for $i=0,...,k-1$. Here $m_{n,k,i}:=\lfloor m_n/k\rfloor+\mathbb{1}_{\{k\lfloor m_n/k\rfloor+(k-i)\leq m_n\}}=\lfloor (m_n+i)/k\rfloor\leq m_n$. Because of $m_n/k=\mathrm{O}(m_n)$, (Δ) (i) and (ii) are the correct conditions for this application of the law of large numbers from Petrov (1975).
Since $\tilde{V}_{n,i}$ and $V_{n,i}$ are $(X_{n,(i-1)r_n+1},...,X_{n,ir_n+s_n-1})$-measurable, $\Delta_{n,i}$ is also measurable with respect to $(X_{n,(i-1)r_n+1},...,X_{n,ir_n+s_n-1})$. The blocks $\Delta_{n,jk-i}$ are separated by $(k-$

$1)r_n - s_n$ X-observations for different j. Hence, by (MX$_k$) and an inequality by Eberlein (1984), the total variation distance between the joint distribution of $\Delta_{n,jk-i}$, $1 \leq j \leq m_{n,k,i}$, and that of $\Delta^*_{n,jk-i}$, $1 \leq j \leq m_{n,k,i}$, converges to 0:

$$\begin{aligned} &\|P^{(\Delta^*_{n,jk-i})_{1\leq j\leq m_{n,k,i}}} - P^{(\Delta_{n,jk-i})_{1\leq j\leq m_{n,k,i}}}\|_{TV} \\ &\leq m_{n,k,i}\beta^{\Delta}_{n,k-1} \leq m_{n,k,i}\beta^{X}_{n,(k-1)r_n-s_n} \to 0. \end{aligned} \tag{3.4.6}$$

In the first step, the inequality for β-mixing coefficients from Eberlein (1984) was applied and in the last step, the condition (MX$_k$) was used. The second inequality follows from the measurability of $\Delta_{n,i}$ with respect to $(X_{n,(i-1)r_n+1}, ..., X_{n,ir_n+s_n})$ for all $1 \leq i \leq m_n$. In particular, from this measurability it follows that $\mathcal{B}^{\Delta,j}_{n,i} \subset \mathcal{B}^{X,jr_n+s_n}_{n,(i-1)r_n}$ for all $1 \leq i \leq j \leq m_n$, and, thereby with the same arguments as in (3.4.1),

$$\beta^{\Delta}_{n,k} \leq \beta^{X}_{n,kr_n-s_n} \to 0.$$

Define the set

$$B_{n,i,\varepsilon,g} := \Big\{(y_j)_{1\leq j\leq m_{n,k,i}} \in (\mathbb{R}^d)^{m_{n,k,i}} : \Big|\frac{1}{\sqrt{p_n}}\sum_{i=j}^{m_{n,k,i}}(y_j - E[\Delta_{n,jk-i}(g)])\Big| > \varepsilon\Big\}.$$

Then,

$$\begin{aligned} &P\Big(\Big|\frac{1}{\sqrt{p_n}}\sum_{i=j}^{m_{n,k,i}}(\Delta_{n,jk-i}(g) - E[\Delta_{n,jk-i}(g)])\Big| > \varepsilon\Big) \\ &= P^{(\Delta_{n,jk-i})_{1\leq j\leq m_{n,k,i}}}(B_{n,i,\varepsilon,g}) \\ &\leq P^{(\Delta^*_{n,jk-i})_{1\leq j\leq m_{n,k,i}}}(B_{n,i,\varepsilon,g}) + \|P^{(\Delta^*_{n,jk-i})_{1\leq j\leq m_{n,k,i}}} - P^{(\Delta_{n,jk-i})_{1\leq j\leq m_{n,k,i}}}\|_{TV} \to 0, \end{aligned}$$

where the first summand converges to 0 due to (3.4.5) and the second due to (3.4.6). Therefore, it applies to all $i = 0, ..., k-1$ that

$$\frac{1}{\sqrt{p_n}}\sum_{j=1}^{m_{n,k,i}}(\Delta_{n,jk-i}(g) - E[\Delta_{n,jk-i}(g)]) = \mathrm{o}_P(1), \qquad \forall\, g \in \mathcal{G}.$$

Summing over all $i \in \{0, ..., k-1\}$, it follows that

$$\frac{1}{\sqrt{p_n}}\sum_{j=1}^{m_n}(\Delta_{n,j}(g) - E[\Delta_{n,j}(g)]) = \mathrm{o}_P(1), \qquad \forall\, g \in \mathcal{G}. \tag{3.4.7}$$

With (3.4.7), the weak convergence of the fidis of $(Z_n(g))_{g\in\mathcal{G}}$ follows if and only if the fidis of $(\bar{Z}_n(g))_{g\in\mathcal{G}}$ converge weakly, with

$$\bar{Z}_n(g) := Z_n(g) - \frac{1}{\sqrt{p_n}}\sum_{j=1}^{m_n}(\Delta_{n,j}(g) - E[\Delta_{n,j}(g)])$$

$$= \frac{1}{\sqrt{p_n}} \sum_{i=1}^{m_n} \left(\tilde{V}_{n,i}(g) - E[\tilde{V}_{n,i}(g)] \right).$$

In the case of convergence, the limit distributions are the same. Analogously, with (3.4.4), the fidis convergence of $(Z_n^*(g))_{g\in\mathcal{G}}$ follows if and only if the fidis of $(\bar{Z}_n^*(g))_{g\in\mathcal{G}}$ converge, where $\bar{Z}_n^*(g)$ is defined analogously to $\bar{Z}_n(g)$ with independent copies.
Another application of the inequality from Eberlein (1984) to $\tilde{V}_{n,i}$ yields

$$\|P^{(\tilde{V}_{n,j}^*)_{1\le j\le m_n}} - P^{(\tilde{V}_{n,j})_{1\le j\le m_n}}\|_{TV} \le m_n \beta_{n,0}^{\tilde{V}} \to 0,$$

where $\tilde{V}_{n,j}^*$ are independent copies of $\tilde{V}_{n,j}$. Here, condition (M$\tilde{V}$) was applied and the blocks $\tilde{V}_{n,j}^*$ are separated by zero X-observations for different j.
Since the total variation distance converges to 0, it follows that the fidis of $(\bar{Z}_n(g))_{g\in\mathcal{G}}$ converge weakly if and only if those of $(\bar{Z}_n^*(g))_{g\in\mathcal{G}}$ converge weakly and, in that case, the limit distributions coincide. Overall, the fidis of $(Z_n(g))_{g\in\mathcal{G}}$ converge if and only if the fidis of $(Z_n^*(g))_{g\in\mathcal{G}}$ converge. □

Before we give the proof of Theorem 3.1.4, the following proof shows that condition (Δ) can be replaced by the simpler but stronger condition (3.1.4). This is needed for the last assertion of Theorem 3.1.4.

Proof of Lemma 3.1.3. It directly holds

$$\begin{aligned} E\Big[(\Delta_n(g) - E[\Delta_n(g)])^2 \mathbb{1}_{\{|\Delta_n(g)-E[\Delta_n(g)]|\le\sqrt{p_n}\}}\Big] &\le Var\Big(\Delta_n(g)\Big) \\ &\le E\left[(\Delta_n(g))^2\right] = \mathrm{o}\left(\frac{p_n}{m_n}\right) \end{aligned}$$

and, with Chebyshev's inequality, this yields

$$P\Big(|\Delta_n(g) - E[\Delta_n(g)]| > \sqrt{p_n}\Big) \le \frac{Var\,(\Delta_n(g))}{p_n} \le \frac{E\,[(\Delta_n(g))^2]}{p_n} = \frac{\mathrm{o}\,(p_n/m_n)}{p_n} = \mathrm{o}\Big(\frac{1}{m_n}\Big).$$

Thus, part (i) and (ii) of (Δ) are direct consequences of (3.1.4). Now, we turn to part (iii) of (Δ). For this, we show that under (3.1.4) we have $\tau_n \to 0$ where τ_n is defined in (3.4.3).
Note that we already established (Δ) (i) and (ii) and therefore (3.4.2) holds by the same arguments as in the previous proof. This implies the convergence

$$\frac{1}{\sqrt{p_n}} \sum_{i=1}^{m_n} \left(\Delta_{n,i}^*(g) - E[\Delta_{n,i}^*(g)] \right) \to \lim_{n\to\infty} \tau_n, \qquad \forall\, g \in \mathcal{G},$$

in probability, if the limit exists. The expectation of the left hand side equals 0. Thus,

$\lim_{n\to\infty}\tau_n = 0$ follows, if the left hand side is uniformly integrable for $n \in \mathbb{N}$, i.e. if

$$\lim_{M\to\infty}\sup_{n\in\mathbb{N}} E\Big[\frac{1}{\sqrt{p_n}}\sum_{i=1}^{m_n}\Big|\Delta^*_{n,i}(g) - E[\Delta^*_{n,i}(g)]\Big|\,\mathbb{1}_{\{p_n^{-1/2}\sum_{i=1}^{m_n}|\Delta^*_{n,i}(g)-E[\Delta^*_{n,i}(g)]|>M\}}\Big] = 0.$$

This is true, if the uniform moment bound (and Lyapunov-type condition)

$$\sup_{n\in\mathbb{N}} E\Big[\Big(\frac{1}{\sqrt{p_n}}\sum_{i=1}^{m_n}\Big(\Delta^*_{n,i}(g) - E[\Delta^*_{n,i}(g)]\Big)\Big)^2\Big] < \infty$$

holds. This, in turn, is implied by stationarity, the independence of $\Delta^*_{n,i}$ and (3.1.4):

$$\begin{aligned}
&\sup_{n\in\mathbb{N}} E\Big[\Big(\frac{1}{\sqrt{p_n}}\sum_{i=1}^{m_n}\Big(\Delta^*_{n,i}(g) - E[\Delta^*_{n,i}(g)]\Big)\Big)^2\Big] \\
&= \sup_{n\in\mathbb{N}}\Big(\frac{1}{p_n}2\sum_{i=1}^{m_n} E\Big[\Big(\Delta^*_{n,i}(g) - E[\Delta^*_{n,i}(g)]\Big)\sum_{j=1}^{i-1}\Big(\Delta^*_{n,j}(g) - E[\Delta^*_{n,j}(g)]\Big)\Big] \\
&\qquad\qquad + \frac{m_n}{p_n}E\Big[(\Delta^*_n(g) - E[\Delta^*_n(g)])^2\Big]\Big) \\
&= \sup_{n\in\mathbb{N}}\Big(\frac{1}{p_n}2\sum_{i=1}^{m_n} E\Big[\Big(\Delta^*_{n,i}(g) - E[\Delta^*_{n,i}(g)]\Big)\Big]\sum_{j=1}^{i-1} E\Big[\Big(\Delta^*_{n,j}(g) - E[\Delta^*_{n,j}(g)]\Big)\Big] \\
&\qquad\qquad + \frac{m_n}{p_n}E\Big[(\Delta^*_n(g) - E[\Delta^*_n(g)])^2\Big]\Big) \\
&= \sup_{n\in\mathbb{N}}\Big(\frac{1}{p_n}m_n(m_n-1)\cdot 0 + \frac{m_n}{p_n}Var(\Delta^*_n(g))\Big) \\
&\le \sup_{n\in\mathbb{N}}\frac{m_n}{p_n}E\Big[(\Delta_n(g))^2\Big] < \infty.
\end{aligned}$$

Thus, with the uniform integrability and the convergence to a constant in probability above, the convergence of the expectation follows:

$$\frac{1}{\sqrt{p_n}}\sum_{i=1}^{m_n}\Big(\Delta^*_{n,i}(g) - E[\Delta^*_{n,i}(g)]\Big) \xrightarrow{P} \lim_{n\to\infty} E\Big[\frac{1}{\sqrt{p_n}}\sum_{i=1}^{m_n}\Big(\Delta^*_{n,i}(g) - E[\Delta^*_{n,i}(g)]\Big)\Big] = 0,$$

i.e. $\lim_{n\to\infty}\tau_n = 0$.
Since the left hand side of condition (Δ) (iii) is just $\tau_n\sqrt{p_n}/m_n$, this proves (Δ) (iii) and thereby the assertion. □

Proof of Theorem 3.1.4. According to Lemma 3.4.1, the fidis of $(Z_n(g))_{g\in\mathcal{G}}$ converge if and only if the fidis of $(Z^*_n(g))_{g\in\mathcal{G}}$ converge. The convergence for $(Z^*_n(g))_{g\in\mathcal{G}}$ follows from the multivariate central limit theorem for triangular schemes of row-wise independent random vectors by Lindeberg-Feller. The conditions (L) and (C) assure that the conditions of the Lindeberg-Feller theorem are satisfied.
The last assertion of the theorem is a direct consequence of Lemma 3.1.3, since (3.1.4) implies (Δ) and, thereby, Lemma 3.4.1 remains true with (Δ) replaced by (3.1.4). (In the proof of Lemma 3.4.1 condition (Δ) (iii) was only used to show $\tau_n \to 0$ which is

established under (3.1.4) in the proof of Lemma 3.1.3.) □

3.4.2 Proofs for Section 3.1.2

The first proof in this section shows that (L2) implies (L) and (L1).

Proof of Lemma 3.1.6. Observe that $\mathbb{1}_{\{V_n(\mathcal{G})>\sqrt{p_n}\varepsilon\}} \leq \mathbb{1}_{\{V_n(\mathcal{G})>\sqrt{p_n}\varepsilon\}} V_n(\mathcal{G})^2/(\varepsilon^2 p_n)$. Using this, by the Cauchy-Schwarz inequality and Chebyshev's inequality, it follows

$$\begin{aligned}
&E^*\left[V_n(\mathcal{G})\mathbb{1}_{\{V_n(\mathcal{G})>\sqrt{p_n}\varepsilon\}}\right] \leq \left(E^*\left[V_n(\mathcal{G})^2\mathbb{1}_{\{V_n(\mathcal{G})>\sqrt{p_n}\varepsilon\}}\right] E^*\left[\mathbb{1}_{\{V_n(\mathcal{G})>\sqrt{p_n}\varepsilon\}}\right]\right)^{1/2} \\
&\leq \left(\frac{\left(E^*\left[V_n(\mathcal{G})^2\mathbb{1}_{\{V_n(\mathcal{G})>\sqrt{p_n}\varepsilon\}}\right]\right)^2}{\varepsilon^2 p_n}\right)^{1/2} \\
&= \mathrm{o}\left(\frac{p_n^2}{p_n m_n^2}\right)^{1/2} = \mathrm{o}\left(\frac{\sqrt{p_n}}{m_n}\right),
\end{aligned}$$

where the penultimate step holds because of (L2). Hence, (L1) is met.
Applying (L2) also yields

$$E\left[(V_n(g))^2\mathbb{1}_{\{|V_n(g)|>\sqrt{p_n}\varepsilon\}}\right] = \mathrm{o}\left(\frac{p_n}{m_n}\right) \tag{3.4.8}$$

for all $\varepsilon > 0$ and $g \in \mathcal{G}$. Thus,

$$E\left[\left(\frac{V_n(g)}{\sqrt{p_n}}\right)^2\right] \leq \frac{1}{p_n}E\left[V_n(g)^2\mathbb{1}_{\{|V_n(g)|>\sqrt{p_n}\varepsilon\}}\right] + \varepsilon^2 = \mathrm{o}\left(\frac{1}{m_n}\right) + \varepsilon^2$$

for all $\varepsilon > 0$ and, therefore, $E[V_n(g)] = o(\sqrt{p_n})$. Together with (3.4.8), this implies (L), since $E[V_n(g)]$ has no impact in the indicator for large n and $E[V_n(g)]^2 \leq E[V_n(g)^2]$. More formally, we may conclude for sufficiently large n that

$$\begin{aligned}
&E\left[(V_n(g) - EV_n(g))^2\mathbb{1}_{\{|V_n(g)-EV_n(g)|>\varepsilon\sqrt{p_n}\}}\right] \\
&\leq 2E\left[\left((V_n(g))^2 + (EV_n(g))^2\right)\mathbb{1}_{\{|V_n(g)|>\varepsilon\sqrt{p_n}/2\}}\right] \\
&\leq 2E^*\left[(V_n(\mathcal{G}))^2\mathbb{1}_{\{V_n(\mathcal{G})>\varepsilon\sqrt{p_n}/2\}}\right] + \mathrm{o}(p_n)P\left(|V_n(g)| > \varepsilon\sqrt{p_n}/2\right) \\
&\leq 4E^*\left[(V_n(\mathcal{G}))^2\mathbb{1}_{\{V_n(\mathcal{G})>\varepsilon\sqrt{p_n}/2\}}\right] \\
&= \mathrm{o}\left(\frac{p_n}{m_n}\right),
\end{aligned}$$

i.e. (L) holds. □

The following proofs for Theorem 3.1.7 and 3.1.9 are inspired by the corresponding Theorems 2.8 and 2.10, respectively, in Drees and Rootzén (2010). The notation and setting, however, are much more general here.

Proof of Theorem 3.1.7. Let k be the number for which condition (MX_k) is satisfied. Again, we denote $m_{n,k,i} = \lfloor (m_n + i)/k \rfloor$. If the processes $(Z_n^{(i)})_{n\in\mathbb{N}}$ with

$$Z_n^{(i)}(g) = \frac{1}{\sqrt{p_n}} \sum_{j=1}^{m_{n,k,i}} (V_{n,kj-i}(g) - E[V_{n,kj-i}(g)]) \tag{3.4.9}$$

is asymptotically tight for all $i \in \{0, ..., k-1\}$, then the process $Z_n = \sum_{i=1}^{k} Z_n^{(i)}$ is itself asymptotically tight as a sum of finitely many asymptotically tight processes. The processes $Z_n^{(i)}$ are the partial sums of Z_n including each k-th summand, starting with summand i. Since the blocks $(V_{n,jk-i})_{1\le j\le m_{n,k,i}}$ are measurable w.r.t. some X-blocks, which have a distance of $(k-1)r_n - s_n$ observations, the inequality for β-mixing coefficients from Eberlein (1984) implies

$$\|P^{(V^*_{n,jk-i})_{1\le j\le m_{n,k,i}}} - P^{(V_{n,jk-i})_{1\le i\le m_{n,k,i}}}\|_{TV} \le m_{n,k,i}\beta^X_{n,(k-1)r_n-s_n} \to 0, \tag{3.4.10}$$

where $(V^*_{n,jk-i})_{1\le j\le m_{n,k,i}}$ are independent copies of $(V_{n,jk-i})_{1\le j\le m_{n,k,i}}$. Because

$$\begin{aligned}\limsup_{n\to\infty} & P^*(Z_n^{(i)} \notin K^\delta) \\ & \le \limsup_{n\to\infty} P^*(Z_n^{(i)*} \notin K^\delta) + \|P^{(V^*_{n,jk-i})_{1\le j\le m_{n,k,i}}} - P^{(V_{n,jk-i})_{1\le i\le m_{n,k,i}}}\|_{TV}\end{aligned}$$

for each set K, and the second summand converges to 0, the process $Z_n^{(i)}(g)$ is asymptotically tight if and only if

$$Z_n^{(i)*}(g) = \frac{1}{\sqrt{p_n}} \sum_{j=1}^{m_{n,k,i}} \left(V^*_{n,kj-i}(g) - E[V^*_{n,kj-i}(g)]\right)$$

is asymptotically tight. Note that $Z_n^{(i)*}(g)$ is the sum of independent blocks $V^*_{n,jk-i}$. The conditions (B), (L1), (D1) and (D2) directly imply the conditions of Theorem 2.11.9 in Van der Vaart and Wellner (1996). The application of this theorem provides the asymptotic tightness of $Z_n^{(i)*}(g)$ and, therefore, of $Z_n^{(i)}(g)$. This holds for all $i = 0, ..., k-1$ and, thus, the assertion is proven. □

The ideas of the next proof are similar to the concept of the previous proof.

Proof of Theorem 3.1.9. Let k be the number for which (MX_k) is satisfied. Because

$$\begin{aligned}P^*\Big(\sup_{g,h\in\mathcal{G},\rho(g,h)<\delta} |Z_n(g) - Z_n(h)| > \varepsilon\Big) &\le P^*\Big(\sum_{i=0}^{k-1} \sup_{g,h\in\mathcal{G},\rho(g,h)<\delta} |Z_n^{(i)}(g) - Z_n^{(i)}(h)| > \varepsilon\Big) \\ &\le \sum_{i=0}^{k-1} P^*\Big(\sup_{g,h\in\mathcal{G},\rho(g,h)<\delta} |Z_n^{(i)}(g) - Z_n^{(i)}(h)| > \frac{\varepsilon}{k}\Big)\end{aligned}$$

with $Z_n^{(i)}(g)$ from (3.4.9) for $0 \le i \le k-1$, the process $(Z_n(g))_{g\in\mathcal{G}}$ is asymptotically equicontinuous if the processes $(Z_n^{(i)}(g))_{g\in\mathcal{G}}$ for $0 \le i \le k-1$ are asymptotically equicon-

tinuous. The asymptotic equicontinuity of $(Z_n^{(i)}(g))_{g\in\mathcal{G}}$ will be shown in the following. The process $(Z_n^{(i)}(g))_{g\in\mathcal{G}}$ is a partial sum of the process $(Z_n(g))_{g\in\mathcal{G}}$, where only each k-th summand occurs. This ensures that the distance between the blocks, with respect to which the individual summands can be measured, becomes larger. Due to the β-mixing condition, these summands can now be approximated by sums of independent processes, to which the result from Van der Vaart and Wellner (1996) can be applied. This idea is formally implemented in the following:
Let $V^*_{n,jk-i}$ be independent copies of $(V_{n,jk-i})$ for $1 \le j \le m_{n,k,i}$. Then, the equation (3.4.10) applies here as well. For the independent copies $(V^*_{n,jk-i})_{1\le j\le m_{n,k,i}}$, only partial sums of the empirical process $(Z_n^*(g))_{g\in\mathcal{G}}$ are considered and all the conditions of Theorem 2.11.1 in Van der Vaart and Wellner (1996) follow from the conditions (D0) (which implies (D4)), (B), (L2), (D1) and (D3). (In fact, the conditions just arise from the application of this Theorem 2.11.1.) The condition (D3) implies the entropy condition required here, since $m_{n,k,i} = \mathrm{O}(m_n)$ and the metric considered for this situation is less than or equal to the metric in (D3). Thus, all conditions of the cited theorem for $(V^*_{n,jk-i})_{1\le j\le m_{n,k,i}}$ are satisfied. Thus, the asymptotic equicontinuity of $(Z_n^{(i)*}(g))_{g\in\mathcal{G}}$ follows with

$$Z_n^{(i)*}(g) = \frac{1}{\sqrt{p_n}} \sum_{j=1}^{m_{n,k,i}} \left(V^*_{n,kj-i}(g) - E[V^*_{n,kj-i}(g)] \right).$$

Hence, the asymptotic equicontinuity of $(Z_n^{(i)}(g))_{g\in\mathcal{G}}$ follows from

$$\begin{aligned} &P^*\Big(\sup_{g,h\in\mathcal{G}:\rho(g,h)<\delta} |Z_n^{(i)}(g) - Z_n^{(i)}(h)| > \varepsilon \Big) \\ &\le P^*\Big(\sup_{g,h:\rho(g,h)<\delta} |Z_n^{(i)*}(g) - Z_n^{(i)*}(h)| > \varepsilon \Big) + \|P^{(V^*_{n,jk-i})_{1\le j\le m_{n,k,i}}} - P^{(V_{n,jk-i})_{1\le i\le m_{n,k,i}}}\|_{TV} \end{aligned}$$

and (3.4.10). This shows the assertion. □

Finally, the proof of process convergence is a simple corollary of the previous theorems.

Proof of Theorem 3.1.10. The weak convergence of an empirical process follows from the weak convergence of the fidis and asymptotic tightness or asymptotic equicontinuity. Thus, the statement of this theorem follows directly from Theorems 3.1.4, 3.1.7 and 3.1.9. It should be noted that condition (L) is implied by (L2) due to Lemma 3.1.6. □

3.4.3 Proofs for Section 3.2

This section starts with the proofs of the sliding blocks limit theorems for bounded and unbounded functions $g_{\max}$, Theorem 3.2.1 and Theorem 3.2.2, respectively.

Proof of Theorem 3.2.1. We start with the proof of part (a). Since $g_{\max}$ is bounded, for each $n \in \mathbb{N}$, it holds

$$V_n(\mathcal{G}) = \sup_{g\in\mathcal{G}} \frac{1}{b_n(g)} \sum_{j=1}^{r_n} g(W_{n,(i-1)r_n+j}) \leq r_n \|g_{\max}\|_\infty \frac{1}{\inf_{g\in\mathcal{G}} b_n(g)} < \infty.$$

Here we used $\inf_{g\in\mathcal{G}} b_n(g) > 0$, which holds by assumption (A2). Therefore, (B) is satisfied. Since $r_n = \mathrm{o}(\sqrt{p_n} \inf_{g\in\mathcal{G}} b_n(g))$, eventually $V_n(\mathcal{G}) \leq \sup_{g\in\mathcal{G}} \|g_{\max}\|_\infty r_n/b_n(g) \leq \sqrt{p_n}\varepsilon$ follows for $\varepsilon > 0$ and a sufficiently large n. In particular this implies $\mathbb{1}_{\{V_n(\mathcal{G})>\sqrt{p_n}\varepsilon\}} = 0$ for a sufficiently large n. Thus, condition (L2) follows, since the indicator on the left hand side equals 0 for a sufficiently large n. Recall that (L2) implies (L) and (L1) by Lemma 3.1.6. By direct calculation,

$$\begin{aligned} E\Big[\Big(\sum_{i=1}^{r_n} \mathbb{1}_{\{g(W_{n,j})\neq 0\}}\Big)^2\Big] &\geq E\Big[\sum_{j=1}^{\lfloor r_n/l_n\rfloor} \Big(\sum_{i=1}^{l_n} \mathbb{1}_{\{g(W_{n,(j-1)l_n+i})\neq 0\}}\Big)^2\Big] \\ &= \Big\lfloor \frac{r_n}{l_n} \Big\rfloor E\Big[\Big(\sum_{i=1}^{l_n} \mathbb{1}_{\{g(W_{n,j})\neq 0\}}\Big)^2\Big]. \end{aligned}$$

Using the row-wise stationarity of $(X_{n,i})_{1\leq i\leq n}$, condition (3.2.4) and $l_n = \mathrm{o}(r_n)$ we may conclude (3.1.4):

$$\begin{aligned} E\big[(\Delta_n(g))^2\big] &\leq \frac{1}{b_n(g)^2} E\Big[\Big(\sum_{j=1}^{l_n} \|g_{\max}\|_\infty \mathbb{1}_{\{g(W_{n,j})\neq 0\}}\Big)^2\Big] \\ &= \frac{1}{b_n(g)^2} \|g_{\max}\|_\infty^2 E\Big[\Big(\sum_{j=1}^{l_n} \mathbb{1}_{\{g(W_{n,j})\neq 0\}}\Big)^2\Big] \\ &= \mathrm{O}\Big(\frac{l_n}{r_n b_n(g)^2} E\Big[\Big(\sum_{j=1}^{r_n} \mathbb{1}_{\{g(W_{n,j})\neq 0\}}\Big)^2\Big]\Big) \\ &= \mathrm{O}\Big(\frac{l_n}{r_n} \frac{p_n b_n(g)^2}{m_n b_n(g)^2}\Big) = \mathrm{o}\Big(\frac{p_n}{m_n}\Big). \end{aligned} \tag{3.4.11}$$

Furthermore, (3.2.5) holds due to assumption (A2):

$$\begin{aligned} &E^*\Big[\sup_{g\in\mathcal{G}} (Z_n(g) - \bar{Z}_n(g))^2\Big] \\ &= E^*\Big[\sup_{g\in\mathcal{G}} \Big(\frac{1}{\sqrt{p_n} b_n(g)} \sum_{j=r_n m_n+1}^{n-s_n+1} \big(g(W_{n,j}) - E(g(W_{n,j}))\big)\Big)^2\Big] \\ &\leq E^*\Big[\sup_{g\in\mathcal{G}} \Big(\frac{1}{\sqrt{p_n} b_n(g)} \sum_{j=r_n m_n+1}^{n-s_n+1} (|g(W_{n,j})| + E[|g(W_{n,j})|])\Big)^2\Big] \\ &\leq \frac{1}{p_n \inf_{g\in\mathcal{G}} b_n(g)^2} E\Big[\Big(\sum_{j=r_n m_n+1}^{n-s_n+1} (|g_{\max}(W_{n,j})| + E[|g_{\max}(W_{n,j})|])\Big)^2\Big] \\ &\leq \frac{1}{p_n \inf_{g\in\mathcal{G}} b_n(g)^2} r_n^2 \|g_{\max}\|_\infty^2 \to 0. \end{aligned}$$

In the last step it was applied that $r_n = \mathrm{o}(\sqrt{p_n} \inf_{g\in\mathcal{G}} b_n(g))$.
Since $(k-1)r_n > l_n$ for $k \geq 2$, it follows $\beta^X_{n,(k-1)r_n-s_n} \leq \beta^X_{n,l_n-s_n}$ and thus (MX$_k$) is implied by (MX). By Lemma 3.1.2, this, in turn, implies $n/r_n\beta^V_{n,(k-1)} \to 0$. Due to the form of $\tilde{V}_{n,i}$ and the $(X_{n,(i-1)r_n+1}, ..., X_{n,(i-1)r_n-l_n+s_n})$-measurability, (M$\tilde{V}$) follows from (MX) with the same arguments as in the proof of Lemma 3.1.2 (cf. (3.4.1)).
The statement about the fidis convergence follows from Theorem 3.1.4. The statement in part (b) about the process convergence follows from Theorem 3.1.10. Note that the (D)-conditions for asymptotic tightness or asymptotic equicontinuity are explicitly assumed in this theorem and all other conditions are shown above. The matching limit of $(Z_n(g))_{g\in\mathcal{G}}$ and $(\bar{Z}_n(g))_{g\in\mathcal{G}}$ is a consequence of (3.2.5). □

Proof of Theorem 3.2.2. We start with part (a) and want to apply Theorem 3.1.4. To this end, we have to show that the conditions (L) and (3.1.4) hold under the new conditions, while the remaining assumptions of Theorem 3.1.4 can be verified as before in the proof of Theorem 3.2.1.
By the Hölder inequality, the generalized Markov inequality and (3.2.6), for all $g \in \mathcal{G}$, we obtain

$$\begin{aligned}
&E^*\left[(V_n(g)^2 \mathbb{1}_{\{|V_n(g)|>\sqrt{p_n}\varepsilon\}}\right]\\
&= \frac{1}{b_n^2(g)} E^*\left[\Big(\sum_{i=1}^{r_n} g(W_{n,i})\Big)^2 \mathbb{1}_{\{|\sum_{i=1}^{r_n} g(W_{n,i})|>\sqrt{p_n}b_n(g)\varepsilon\}}\right]\\
&\leq \frac{1}{b_n^2(g)} E^*\left[\Big(\Big|\sum_{i=1}^{r_n} g(W_{n,i})\Big|\Big)^{2+\delta}\right]^{2/(2+\delta)} E^*\left[\mathbb{1}_{\{|\sum_{i=1}^{r_n} g(W_{n,i})|>\sqrt{p_n}b_n(g)\varepsilon\}}\right]^{1-2/(2+\delta)}\\
&= \frac{1}{b_n^2(g)} \mathrm{O}\left(\frac{p_n b_n^2(g)}{m_n}\right)^{2/(2+\delta)} \left(\frac{E^*\left[(|\sum_{i=1}^{r_n} g(W_{n,i})|)^{2+\delta}\right]}{(\sqrt{p_n}b_n(g)\varepsilon)^{2+\delta}}\right)^{1-2/(2+\delta)}\\
&= \frac{1}{b_n^2(g)} \mathrm{O}\left(\frac{p_n b_n^2(g)}{m_n}\right)^{2/(2+\delta)} \mathrm{O}\left(\frac{p_n b_n^2(g)}{m_n}\right)^{1-2/(2+\delta)} \left(\frac{1}{(\sqrt{p_n}b_n(g)\varepsilon)^{2+\delta}}\right)^{1-2/(2+\delta)}\\
&= \frac{1}{b_n^2(g)} \mathrm{O}\left(\frac{p_n b_n^2(g)}{m_n} \frac{1}{(\sqrt{p_n}b_n(g))^{\delta}}\right)\\
&= \frac{1}{b_n^2(g)} \mathrm{o}\left(\frac{p_n b_n^2(g)}{m_n}\right) = \mathrm{o}\left(\frac{p_n}{m_n}\right).
\end{aligned}$$

The penultimate equality holds because of $\sqrt{p_n}b_n(g) \to \infty$ by assumption (A2). With the same arguments as in the proof of Lemma 3.1.6, it follows that condition (L) is satisfied. Furthermore,

$$\begin{aligned}
E\left[\Big(\sum_{i=1}^{r_n} |g(W_{n,i})|\Big)^2\right] &\geq \sum_{j=1}^{\lfloor r_n/l_n\rfloor} E\left[\Big(\sum_{i=1}^{l_n} |g(W_{n,(j-1)l_n+i})|\Big)^2\right]\\
&= \lfloor r_n/l_n\rfloor E\left[\Big(\sum_{i=1}^{l_n} |g(W_{n,i})|\Big)^2\right]
\end{aligned}$$

and, thus, by (3.2.6),

$$\begin{aligned}
E[\Delta_n(g)^2] &\leq \frac{1}{b_n^2(g)} E\Big[\Big(\sum_{i=1}^{l_n} |g(W_{n,i})|\Big)^2\Big] \\
&\leq \frac{1}{b_n^2(g)\lfloor r_n/l_n \rfloor} E\Big[\Big(\sum_{i=1}^{r_n} |g(W_{n,i})|\Big)^2\Big] \\
&\leq \frac{1}{b_n^2(g)\lfloor r_n/l_n \rfloor} E\Big[\Big(\sum_{i=1}^{r_n} |g(W_{n,i})|\Big)^{2+\delta} \mathbb{1}_{\left\{\sum_{i=1}^{r_n} |g(W_{n,i})|>1\right\}} + \mathbb{1}_{\left\{0<\sum_{i=1}^{r_n} |g(W_{n,i})|\leq 1\right\}}\Big] \\
&\leq \frac{1}{b_n^2(g)\lfloor r_n/l_n \rfloor} E\Big[\Big(\sum_{i=1}^{r_n} |g(W_{n,i})|\Big)^{2+\delta} + \mathbb{1}_{\left\{\sum_{i=1}^{r_n} |g(W_{n,i})|\neq 0\right\}}\Big] \\
&= \mathrm{O}\bigg(\frac{l_n}{r_n b_n^2(g)}\bigg(\frac{p_n b_n^2(g)}{m_n} + P(V_n(|g|)\neq 0)\bigg)\bigg) = \mathrm{o}\bigg(\frac{p_n}{m_n}\bigg) \qquad (3.4.12)
\end{aligned}$$

where in the last step we have used $m_n l_n P(V_n(|g|) \neq 0) = \mathrm{o}(r_n b_n^2(g) p_n)$. Hence, Condition (3.1.4) holds. From here, the same reasoning as in the proof of Theorem 3.2.1 can be used to show the weak convergence of $(Z_n(g))_{g\in\mathcal{G}}$ to a Gaussian process with covariance function c.

Similarly,

$$\begin{aligned}
E\Big[(\bar{Z}_n(g) - Z_n(g))^2\Big] &= Var\bigg(\frac{1}{\sqrt{p_n} b_n(g)} \sum_{j=r_n m_n+1}^{n-s_n} g(W_{n,j})\bigg) \\
&\leq \frac{1}{p_n b_n^2(g)} E\Big[\Big(\sum_{j=r_n m_n+1}^{n-s_n} |g(W_{n,j})|\Big)^2\Big] \\
&= \mathrm{O}\bigg(\frac{1}{m_n} + \frac{P(V_n(|g|)\neq 0)}{p_n b_n^2(g)}\bigg) \to 0,
\end{aligned}$$

because $p_n b_n^2(g) \to \infty$ by assumption (A2), so that the fidi-convergence of $(\bar{Z}_n(g))_{g\in\mathcal{G}}$ follows, too.

Next, we prove part (b), the process convergence. Note that for this part of the assertion we assume $b_n(g) = b_n$. We want to apply Theorem 3.1.10 and all conditions apart from (B), (L2) and (3.1.4) can be shown as in the proof of Theorem 3.2.1.

With similar calculations as above, using $V_n(\mathcal{G}) = b_n^{-1} \sum_{i=1}^{r_n} g_{\max}(W_{n,i})$, the Hölder inequality, the generalized Markov inequality and (3.2.6), we obtain

$$\begin{aligned}
& E^*\left[(V_n(\mathcal{G}))^2 \mathbb{1}_{\left\{V_n(\mathcal{G})>\sqrt{p_n}\varepsilon\right\}}\right] \\
&= \frac{1}{b_n^2} E^*\left[\Big(\sum_{i=1}^{r_n} g_{\max}(W_{n,i})\Big)^2 \mathbb{1}_{\left\{\sum_{i=1}^{r_n} g_{\max}(W_{n,i})>\sqrt{p_n} b_n \varepsilon\right\}}\right] \\
&\leq \frac{1}{b_n^2} E^*\left[\Big(\sum_{i=1}^{r_n} g_{\max}(W_{n,i})\Big)^{2+\delta}\right]^{2/(2+\delta)} E^*\left[\mathbb{1}_{\left\{\sum_{i=1}^{r_n} g_{\max}(W_{n,i})>\sqrt{p_n} b_n \varepsilon\right\}}\right]^{1-2/(2+\delta)} \\
&= \frac{1}{b_n^2} \mathrm{O}\left(\frac{p_n b_n^2}{m_n}\right)^{2/(2+\delta)} \left(\frac{E^*\left[(\sum_{i=1}^{r_n} g_{\max}(W_{n,i}))^{2+\delta}\right]}{(\sqrt{p_n} b_n \varepsilon)^{2+\delta}}\right)^{1-2/(2+\delta)}
\end{aligned}$$

$$= \frac{1}{b_n^2} \mathrm{O}\left(\frac{p_n b_n^2}{m_n}\right)^{2/(2+\delta)} \mathrm{O}\left(\frac{p_n b_n^2}{m_n}\right)^{1-2/(2+\delta)} \left(\frac{1}{(\sqrt{p_n} b_n \varepsilon)^{2+\delta}}\right)^{1-2/(2+\delta)}$$
$$= \frac{1}{b_n^2} \mathrm{O}\left(\frac{p_n b_n^2}{m_n} \frac{1}{(\sqrt{p_n} b_n)^{\delta}}\right) = \frac{1}{b_n^2} \mathrm{o}\left(\frac{p_n b_n^2}{m_n}\right) = \mathrm{o}\left(\frac{p_n}{m_n}\right).$$

Thus, conditions (B) and (L2) are satisfied. The same calculations as those leading to (3.4.12) with $g_{\max}$ instead of g yield (3.1.4). Now, the convergence of the fidis of $(Z_n(g))_{g\in\mathcal{G}}$ follows from Theorem 3.1.10.
With similar arguments as in (3.4.12), it follows

$$\begin{aligned}
E^*\Big[\sup_{g\in\mathcal{G}}(\bar{Z}_n(g) - Z_n(g))^2\Big] &= E^*\Bigg[\sup_{g\in\mathcal{G}}\Big|\frac{1}{\sqrt{p_n} b_n(g)} \sum_{j=r_n m_n+1}^{n-s_n+1} (g(W_{n,j}) - E[g(W_{n,j})])\Big|^2\Bigg] \\
&\le E\Bigg[\Big(\frac{1}{\sqrt{p_n} b_n} \sum_{j=1}^{n-s_n+1} g_{\max}(W_{n,j}) + E[g_{\max}(W_{n,j})]\Big)^2\Bigg] \\
&\le \frac{4}{p_n b_n^2} E\Bigg[\Big(\sum_{j=1}^{r_n} g_{\max}(W_{n,j})\Big)^2\Bigg] \\
&= \mathrm{O}\Big(\frac{1}{p_n b_n^2}\Big(\frac{p_n b_n^2}{m_n} + \frac{p_n b_n^2}{b_n^2}\Big)\Big) = \mathrm{O}\Big(\frac{1}{m_n} + \frac{1}{b_n^2}\Big) \to 0
\end{aligned}$$

because of $b_n \to \infty$ by assumption (A2) (since $r_n \to \infty$, $p_n \to 0$ and $r_n = \mathrm{o}(\sqrt{p_n} b_n)$). Thus, (3.2.5) holds and the process convergence of $(\bar{Z}_n(g))_{g\in\mathcal{G}}$ follows, which completes the proof. □

The proof of the modified sliding blocks limit theorem (Theorem 3.2.3) is basically the same as the previous proof of Theorem 3.2.2. Therefore, we show only the differences.

Proof of Theorem 3.2.3. We start with part (a). We want to apply Theorem 3.1.4. To this end, we only have to verify condition (3.1.4), because (L) is assumed and the remaining conditions follow as in the proof of Theorem 3.2.1. Condition (B) is obviously fulfilled by (3.2.7) and since we consider only fidis for this part.
It is clear that

$$\begin{aligned}
E\Bigg[\Big(\sum_{i=1}^{r_n} |g(W_{n,i})|\Big)^2\Bigg] &\ge \sum_{j=1}^{\lfloor r_n/l_n\rfloor} E\Bigg[\Big(\sum_{i=1}^{l_n} |g(W_{n,(j-1)l_n+i})|\Big)^2\Bigg] \\
&= \lfloor r_n/l_n\rfloor E\Bigg[\Big(\sum_{i=1}^{l_n} |g(W_{n,i})|\Big)^2\Bigg]
\end{aligned}$$

and, thus, by (3.2.7),

$$\begin{aligned}
E(\Delta_n(g)^2) &\le \frac{1}{b_n^2(g)} E\Bigg[\Big(\sum_{i=1}^{l_n} |g(W_{n,i})|\Big)^2\Bigg] \le \frac{1}{b_n^2(g)\lfloor r_n/l_n\rfloor} E\Bigg[\Big(\sum_{i=1}^{r_n} |g(W_{n,i})|\Big)^2\Bigg] \\
&= \mathrm{O}\Big(\frac{l_n}{r_n b_n^2(g)} \frac{p_n b_n^2(g)}{m_n}\Big) = \mathrm{o}\Big(\frac{p_n}{m_n}\Big), \qquad (3.4.13)
\end{aligned}$$

where we used $l_n = o(r_n)$. Hence, condition (3.1.4) holds, which, in turn, implies (Δ) (cf. Lemma 3.1.3). From here, the same reasoning as in the proof of Theorem 3.2.1 can be used to show the weak convergence of $(Z_n(g))_{g\in\mathcal{G}}$ to a Gaussian process with covariance function c.
Similarly,

$$E\Big[(\bar{Z}_n(g) - Z_n(g))^2\Big] = Var\bigg(\frac{1}{\sqrt{p_n}b_n(g)}\sum_{j=r_nm_n+1}^{n-s_n} g(W_{n,j})\bigg)$$
$$\le \frac{1}{p_n b_n^2(g)} E\bigg[\Big(\sum_{j=r_nm_n+1}^{n-s_n} |g(W_{n,j})|\Big)^2\bigg]$$
$$= O\Big(\frac{1}{m_n}\Big) \to 0,$$

so that the fidi-convergence of $(\bar{Z}_n(g))_{g\in\mathcal{G}}$ follows, too.
Next we prove part (b), the process convergence. Note that for this part of the assertion we have $b_n(g) = b_n$. We want to apply Theorem 3.1.10 and all conditions apart from (B), (L) and (3.1.4) can be shown as in the proof of Theorem 3.2.1. Note that (L) holds by assumption and (B) is obviously satisfied by (3.2.7) and the measurability assumption. The same calculations as those leading to (3.4.13) with $g_{\max}$ instead of g yield (3.1.4). Now, the convergence of the fidis of $(Z_n(g))_{g\in\mathcal{G}}$ follows from Theorem 3.1.10.
With similar arguments as before, it follows

$$E^*\Big[\sup_{g\in\mathcal{G}}(\bar{Z}_n(g) - Z_n(g))^2\Big] \le \frac{1}{p_n b_n^2} E\bigg[\Big(\sum_{j=1}^{r_n} g_{\max}(W_{n,j})\Big)^2\bigg] = O\Big(\frac{1}{m_n}\Big) \to 0.$$

Thus, the process convergence of $(\bar{Z}_n(g))_{g\in\mathcal{G}}$ follows, which completes the proof. □

It remains to prove the three lemmas about the verification of conditions (3.2.4) and (C).

Proof of Lemma 3.2.4. Since $g_{\max}$ is bounded, condition (3.1.4) follows from condition (3.2.4) according to Theorem 3.2.1. This condition is verified below. One has

$$E\bigg[\Big(\sum_{j=1}^{r_n} \mathbb{1}_{\{g(W_{n,j})\neq 0\}}\Big)^2\bigg] = \sum_{i=1}^{r_n}\sum_{j=1}^{r_n} E\Big[\mathbb{1}_{\{g(W_{n,i})\neq 0\}}\mathbb{1}_{\{g(W_{n,j})\neq 0\}}\Big]$$
$$\le 2r_n \sum_{k=1}^{r_n}\Big(1 - \frac{k-1}{r_n}\Big) P\left(g(W_{n,1}) \neq 0, g(W_{n,k}) \neq 0\right)$$
$$= 2r_n O\Big(\frac{p_n b_n(g)^2}{n}\Big) = O\Big(\frac{p_n b_n(g)^2}{m_n}\Big).$$

This is true for all $g \in \mathcal{G}$ and, thus, concludes the proof. □

Proof of Lemma 3.2.5. Since $g_{\max}$ is bounded, according to Theorem 3.2.1, condition

(3.1.4) is implied by (3.2.4). This condition is verified below. By stationarity

$$\begin{aligned}
E\Big[\Big(\sum_{j=1}^{r_n}\mathbb{1}_{\{g(W_{n,j})\neq 0\}}\Big)^2\Big] &= \sum_{i=1}^{r_n}\sum_{j=1}^{r_n} E\left[\mathbb{1}_{\{g(W_{n,i})\neq 0\}}\mathbb{1}_{\{g(W_{n,j})\neq 0\}}\right]\\
&\leq 2r_n\sum_{k=1}^{r_n}\Big(1-\frac{k-1}{r_n}\Big)P\left(g(W_{n,1})\neq 0, g(W_{n,k})\neq 0\right)\\
&= 2r_n q_{g,n}\sum_{k=1}^{r_n-1}\Big(1-\frac{k-1}{r_n}\Big)P\left(g(W_{n,k})\neq 0 \mid g(W_{n,1})\neq 0\right).
\end{aligned}$$

Moreover,

$$\begin{aligned}
&\sum_{k=1}^{r_n}\Big(1-\frac{k-1}{r_n}\Big)P\left(g(W_{n,k})\neq 0 \mid g(W_{n,1})\neq 0\right)\\
&\leq \sum_{k=1}^{r_n}P\left(g(W_{n,k})\neq 0 \mid g(W_{n,1})\neq 0\right) = \mathrm{O}\Big(\frac{b_n(g)^2}{m_n}\Big).
\end{aligned}$$

Therefore, it follows

$$E\Big[\Big(\sum_{j=1}^{r_n}\mathbb{1}_{\{g(W_{n,j})\neq 0\}}\Big)^2\Big] = r_n q_{g,n}\mathrm{O}\Big(\frac{b_n(g)^2}{m_n}\Big) = \mathrm{O}\Big(\frac{p_n b_n(g)^2}{m_n}\Big).$$

This is true for all $g \in \mathcal{G}$ and, thus, concludes the proof. □

Proof of Lemma 3.2.6. Let $g, h \in \mathcal{G}$. Using the stationarity of $(X_t)_{t\in\mathbb{Z}}$ and $b_n(g) = \sqrt{m_n}$ leads to

$$\begin{aligned}
&\frac{m_n}{p_n}Cov(V_n(g), V_n(h)) = \frac{m_n}{p_n b_n(g) b_n(h)}Cov\Big(\sum_{i=1}^{r_n} g(W_{n,i}), \sum_{j=1}^{r_n} h(W_{n,i})\Big)\\
&= \frac{1}{p_n}E\Big[\sum_{i=1}^{r_n}\sum_{j=1}^{r_n} g(W_{n,i})h(W_{n,j})\Big] - \frac{1}{p_n}E\Big[\sum_{i=1}^{r_n} g(W_{n,i})\Big]E\Big[\sum_{j=1}^{r_n} h(W_{n,i})\Big]\\
&= \frac{1}{p_n}\sum_{k=-r_n+1}^{r_n-1}(r_n-|k|)E[g(W_{n,0})h(W_{n,k})] - \frac{r_n q_{g,n}}{p_n} r_n q_{h,n}\|g_{\max}\|_\infty^2\\
&= \frac{q_{h,n} r_n}{p_n}\sum_{k=-r_n+1}^{r_n-1}\Big(1-\frac{|k|}{r_n}\Big)E[g(W_{n,0})h(W_{n,k}) \mid h(W_{n,0})\neq 0] + \mathrm{o}(1).
\end{aligned}$$

Moreover,

$$\begin{aligned}
E[g(W_{n,0})h(W_{n,k}) \mid h(W_{n,0})\neq 0] &\leq \|h\|_\infty\|g\|_\infty P\left(h(W_{n,|k|})\neq 0 \mid h(W_{n,0})\neq 0\right)\\
&\leq \|h\|_\infty\|g\|_\infty e_{n,h}(|k|),
\end{aligned}$$

where for $k < 0$ stationarity is applied. By the assumptions and with Pratt's Lemma (Pratt, 1960), we achieve convergence of this expression for the covariance to

$$\Big(\lim_{n\to\infty}\frac{q_{h,n} r_n}{p_n}\Big)\sum_{k\in\mathbb{Z}}\lim_{n\to\infty} E[g(W_{n,0})h(W_{n,k}) \mid g(W_{n,0})\neq 0],$$

where the first limit exists due to the assumptions.
If $(X_t)_{t\in\mathbb{Z}}$ is regularly varying, $X_{n,i} = X_i/u_n \mathbb{1}_{\{\|X_i\|>u_n\}}$, $P(g(W_{n,0}) \neq 0)$ is of the same order as $P(\|X_0\| > u_n)$, and $g(W_{n,0}) \neq 0$ implies $\|X_0\| > u_n$, then

$$E[g(W_{n,0})h(W_{n,k}) \mid g(W_{n,0}) \neq 0] = E[g(W_{n,0})h(W_{n,k}) \mid \|X_0\| > u_n] \frac{P(\|X_0\| > u_n)}{P(g(W_{n,0}) \neq 0)}.$$

Therefore, the limit can be stated more explicitly in terms of the tail process $Y = (Y_t)_{t\in\mathbb{Z}}$ as

$$\Big(\lim_{n\to\infty} \frac{q_{h,n} r_n}{p_n} \frac{P(\|X_0\| > u_n)}{P(g(W_{n,0}) \neq 0)}\Big) \sum_{k\in\mathbb{Z}} E[g(Y)h(B^k Y)]$$

where B denotes the shift operator, i.e. $B^k Y = (Y_{t+k})_{t\in\mathbb{Z}}$ and the limit exists due to the assumptions. □

3.4.4 Proofs for Section 3.3

The first proof shows that $r_n/s_n \in \mathbb{N}$ can be assumed w.l.o.g.

Proof of Lemma 3.3.3. Suppose that, for a given sequence $(s_n)_{n\in\mathbb{N}}$, the sequence $(r_n)_{n\in\mathbb{N}}$ is given such that the assumptions of Theorem 3.2.1 are satisfied. Let $r_n^* := \lfloor r_n/s_n \rfloor s_n \sim r_n$, so that $l_n = \mathrm{o}(r_n^*)$, $r_n^* = \mathrm{o}(n)$ and $m_n^* := \lfloor (n - s_n + 1)/r_n^* \rfloor \sim m_n$. The proof of Theorem 3.2.1 (cf. (3.4.11)) shows that for

$$V_{n,1}^*(g) := \frac{1}{b_n(g)} \sum_{j=1}^{r_n^*} g(W_{n,j}), \quad \text{and} \quad p_n^* := P(\exists g \in \mathcal{G} : V_{n,1}^*(g) \neq 0),$$

one has

$$E\Big[(V_{n,1}^*(g) - V_{n,1}(g))^2\Big] = E\Big[\Big(\frac{1}{b_n(g)} \sum_{j=r_n^*+1}^{r_n} g(W_{n,j})\Big)^2\Big] = \mathrm{o}\Big(\frac{p_n}{m_n}\Big),$$
$$|p_n^* - p_n| \le s_n v_n.$$

Hence, due to $p_n \asymp r_n v_n$, we have $p_n^* \sim p_n$ and the conditions of Theorem 3.2.1 are still fulfilled if one replaces r_n with r_n^*. The same arguments can be used in the case of unbounded functions with Theorem 3.2.2 and in the case of disjoint block sums in Theorem 3.1.4. This concludes the proof. □

The proof of Lemma 3.3.4 about the variance inequalities for the pseudo estimators make use of methods which were used similarly by Zou et al. (2021), Lemma A.10.

Proof of Lemma 3.3.4. In the following, we examine the pre-asymptotic variances. Note that, due to $r_n/s_n \in \mathbb{N}$, it holds $\lfloor r_n/s_n \rfloor = r_n/s_n$. For $c^{(d)}$, it holds that

$$
\begin{aligned}
c^{(d)} &\leftarrow \frac{1}{r_n v_n a_n^2} Var\Big(\sum_{i=1}^{\lfloor r_n/s_n \rfloor} g(W_{n,is_n+1}) \Big) \\
&= \frac{1}{r_n v_n a_n^2} E\Big[\sum_{i=1}^{r_n/s_n} \sum_{j=1}^{r_n/s_n} g(W_{n,js_n+1}) g(W_{n,is_n+1}) \Big] - \frac{1}{r_n v_n a_n^2} \Big(\frac{r_n}{s_n} E[g(W_{n,0})] \Big)^2 \\
&= \frac{1}{r_n v_n a_n^2} \sum_{k=-r_n/s_n+1}^{r_n/s_n-1} \Big(\frac{r_n}{s_n} - \frac{|k| s_n}{s_n} \Big) E\left[g(W_{n,ks_n}) g(W_{n,0}) \right] - \frac{r_n}{s_n^2 v_n a_n^2} E[g(W_{n,0})]^2 \\
&= \frac{1}{r_n v_n a_n^2} \sum_{k=-r_n+1}^{r_n-1} \mathbb{1}_{\{k \bmod s_n = 0\}} \Big(\frac{r_n}{s_n} - \frac{|k|}{s_n} \Big) E\left[g(W_{n,k}) g(W_{n,0}) \right] - \frac{r_n E[g(W_{n,0})]^2}{s_n^2 v_n a_n^2} \\
&= \frac{1}{s_n v_n a_n^2} \sum_{k=-r_n+1}^{r_n-1} \mathbb{1}_{\{k \bmod s_n = 0\}} \Big(1 - \frac{|k|}{r_n} \Big) E\left[g(W_{n,k}) g(W_{n,0}) \right] - \frac{r_n}{s_n^2 v_n a_n^2} E[g(W_{n,0})]^2.
\end{aligned}
$$

For the variance $c^{(s)}$, we obtain

$$
\frac{1}{r_n v_n s_n^2 a_n^2} Var\Big(\sum_{i=1}^{r_n} g(W_{n,i}) \Big) \to c^{(s)},
$$

where both convergences hold by assumption. For this last variance on the left hand side, we also obtain

$$
\begin{aligned}
&\frac{1}{r_n v_n s_n^2 a_n^2} Var\Big(\sum_{i=1}^{r_n} g(W_{n,i}) \Big) \\
&= \frac{1}{r_n v_n s_n^2 a_n^2} E\Big[\sum_{i=1}^{r_n} \sum_{j=1}^{r_n} g(W_{n,j}) g(W_{n,i}) \Big] - \frac{1}{r_n v_n s_n^2 a_n^2} r_n^2 E\Big[g(W_{n,0}) \Big]^2 \\
&= \frac{1}{v_n s_n^2 a_n^2} \sum_{k=-r_n+1}^{r_n-1} \Big(1 - \frac{|k|}{r_n} \Big) E\left[g(W_{n,0}) g(W_{n,k}) \right] - \frac{r_n}{v_n s_n^2 a_n^2} E\left[g(W_{n,0}) \right]^2 .
\end{aligned}
$$

The difference between the pre-asymptotic variances of the disjoint and sliding blocks statistics is

$$
\begin{aligned}
&\frac{1}{s_n v_n a_n^2} \sum_{k=-r_n+1}^{r_n-1} \mathbb{1}_{\{k \bmod s_n = 0\}} \Big(1 - \frac{|k|}{r_n} \Big) E\left[g(W_{n,k}) g(W_{n,0}) \right] - \frac{r_n}{s_n^2 v_n a_n^2} E[g(W_{n,0})]^2 \\
&\qquad - \frac{1}{v_n s_n^2 a_n^2} \sum_{k=-r_n+1}^{r_n-1} \Big(1 - \frac{|k|}{r_n} \Big) E\left[g(W_{n,0}) g(W_{n,k}) \right] + \frac{r_n}{v_n s_n^2 a_n^2} E\left[g(W_{n,0}) \right]^2 \\
&= \frac{1}{a_n^2} \frac{1}{s_n v_n} \left(\sum_{k=-r_n+1}^{r_n-1} \Big(1 - \frac{|k|}{r_n} \Big) \gamma_n(k) E\left[g(W_{n,0}) g(W_{n,k}) \right] \right),
\end{aligned}
$$

where, for $k \in \mathbb{Z}$,

$$
\gamma_n(k) = \begin{cases} 1 - \frac{1}{s_n}, & k \bmod s_n = 0, \\ -\frac{1}{s_n}, & k \bmod s_n \neq 0. \end{cases}
$$

Since $a_n^{-2}(s_n v_n)^{-1} \geq 0$, it is enough to show

$$\liminf_{n\to\infty} \sum_{k=-r_n+1}^{r_n-1} \Big(1 - \frac{|k|}{r_n}\Big)\gamma_n(k)E\left[g(W_{n,0})g(W_{n,k})\right] \geq 0.$$

From now on, we follow an idea by Zou et al. (2021), proof of Lemma A.10. Define U_n as a random variable uniformly distributed on $\{0,..,s_n-1\}$ which is independent of $(X_{n,i})_{1\leq i\leq n}$. Define

$$\phi_{n,k} = \begin{cases} \frac{s_n-1}{\sqrt{s_n}}, & k \bmod s_n = U_n, \\ -\frac{1}{\sqrt{s_n}}, & \text{else}, \end{cases}$$

for $k \in \mathbb{Z}$. If $(h \bmod s_n) = 0$, then

$$E[\phi_{n,k}\phi_{n,k+h}] = \frac{1}{s_n}\frac{(s_n-1)^2}{s_n} + \frac{s_n-1}{s_n}\frac{1}{s_n} = \frac{s_n-1}{s_n} = 1 - \frac{1}{s_n},$$

whereas, for $(h \bmod s_n) \neq 0$,

$$E[\phi_{n,k}\phi_{n,k+h}] = \frac{2}{s_n}\frac{s_n-1}{\sqrt{s_n}}\frac{-1}{\sqrt{s_n}} + \frac{s_n-2}{s_n}\frac{1}{s_n} = -\frac{1}{s_n}.$$

Thus, $E[\phi_{n,k}\phi_{n,k+h}] = \gamma_n(h)$ by construction. From this, it follows

$$\begin{aligned} E[\phi_{n,j}\phi_{n,i}g(W_{n,i})g(W_{n,j})] &= E[\phi_{n,j}\phi_{n,i}]E[g(W_{n,i})g(W_{n,j})] \\ &= \gamma_n(|i-j|)E[g(W_{n,0})g(W_{n,|i-j|})] \end{aligned}$$

for all $i,j \in \{1,...,r_n\}$. Here we applied the independence of U_n and $(X_{n,i})_{1\leq i\leq n}$ and the previous calculations.

Analogous arguments as above yield

$$\begin{aligned} 0 \leq \frac{1}{r_n}E\left[\Big(\sum_{j=1}^{r_n}\phi_{n,j}g(W_{n,j})\Big)^2\right] &= \frac{1}{r_n}\sum_{j=1}^{r_n}\sum_{i=1}^{r_n} E\left[\phi_{n,i}\phi_{n,j}g(W_{n,i})g(W_{n,j})\right] \\ &= \frac{1}{r_n}\sum_{j=1}^{r_n}\sum_{i=1}^{r_n}\gamma_n(|i-j|)E[g(W_{n,0})g(W_{n,|i-j|})] \\ &= \sum_{k=-r_n+1}^{r_n-1}\Big(1-\frac{|k|}{r_n}\Big)\gamma_n(|k|)E[g(W_{n,0})g(W_{n,k})]. \end{aligned}$$

Thus, $\liminf_{n\to\infty}\sum_{k=-r_n+1}^{r_n-1}(1-|k|/r_n)\gamma_n(k)E\left[g(W_{n,0})g(W_{n,k})|g(W_{n,0})\neq 0\right] \geq 0$. All in all, we have shown

$$c^{(d)} - c^{(s)} = \lim_{n\to\infty}\frac{1}{r_n v_n s_n a_n^2}E\left[\Big(\sum_{j=1}^{r_n}\phi_{n,j}g(W_{n,j})\Big)^2\right] \geq 0.$$

Therefore, $c^{(d)} \geq c^{(s)}$ which concludes the proof. $\square$

To derive the asymptotic normality of fractions, Lemma 3.3.5 can be employed, which is proven next by a continuous mapping argument.

Proof of Lemma 3.3.5. Direct calculations yield

$$\begin{aligned}
\sqrt{nv_n}\Big(\frac{Z_n^1}{Z_n^2} - \xi\Big) &= \sqrt{nv_n}\frac{(Z_n^1 - E[Z_n^1]) + E[Z_n^1] - \xi(Z_n^2 - E[Z_n^2]) - \xi E[Z_n^2]}{(Z_n^2 - E[Z_n^2]) + E[Z_n^2]} \\
&= \frac{\sqrt{nv_n}(Z_n^1 - E[Z_n^1]) - \sqrt{nv_n}\xi(Z_n^2 - E[Z_n^2]) + \sqrt{nv_n}(E[Z_n^1] - \xi E[Z_n^2])}{\sqrt{nv_n}^{-1}\sqrt{nv_n}(Z_n^2 - E[Z_n^2]) + E[Z_n^2]} \\
&= \frac{\sqrt{nv_n}(Z_n^1 - E[Z_n^1]) - \sqrt{nv_n}\xi(Z_n^2 - E[Z_n^2]) + \mathrm{o}(1)}{E[Z_n^2] + \mathrm{o}_P(1)} \\
&\xrightarrow{w} \frac{1}{b}(Z^1 - \xi Z^2).
\end{aligned}$$

In the second to last step we used $\sqrt{nv_n}^{-1} \to 0$ and the bias condition. In the last step $E[Z_n^2] \to b$ was used. Thus, $\sqrt{nv_n}\,(Z_n^1/Z_n^2 - \xi)$ is asymptotically normal. The asymptotic variance of the fraction is given by

$$Var(Z^1 - \xi Z^2) = \frac{1}{b^2}\Big(Var(Z^1) + \xi^2 Var(Z^2) - 2\xi Cov(Z^1, Z^2)\Big). \qquad \square$$

Applying the previous two results, we can now prove Theorem 3.3.6.

Proof of Theorem 3.3.6. W.l.o.g. we assume $g \geq 0$, which implies $\xi \geq 0$. For $g \leq 0$ the calculations remains the same, just change the sign.
Since $\tilde{c}^{(d)} = c^{(d)} + \xi^2 c^{(v)} - 2\xi c^{(d,v)}$ and $\tilde{c}^{(s)} = c^{(s)} + \xi^2 c^{(v)} - 2\xi c^{(s,v)}$ we have $\tilde{c}^{(d)} - \tilde{c}^{(s)} = c^{(d)} - c^{(s)} - 2\xi(c^{(d,v)} - c^{(s,v)})$. We already know from Lemma 3.3.4 that $c^{(s)} \leq c^{(d)}$. Therefore, it suffices to show that $c^{(d,v)} \leq c^{(s,v)}$.
Using the row-wise stationarity of the triangular scheme, the asymptotic covariance $c^{(s,v)}$ can be calculated as the limit of

$$\begin{aligned}
&\frac{1}{r_n v_n s_n a_n} Cov\Big(\sum_{j=1}^{r_n} g(W_{n,j}), \sum_{i=1}^{r_n} \mathbb{1}_{\{X_{n,i}\neq 0\}}\Big) \\
&= \frac{1}{r_n v_n s_n a_n} E\Big[\sum_{j=1}^{r_n} g(W_{n,j}) \sum_{i=1}^{r_n} \mathbb{1}_{\{X_{n,i}\neq 0\}}\Big] - \frac{1}{r_n v_n s_n a_n} E\left[\sum_{j=1}^{r_n} g(W_{n,j})\right] E\left[\sum_{i=1}^{r_n} \mathbb{1}_{\{X_{n,i}\neq 0\}}\right] \\
&= \frac{1}{r_n v_n s_n a_n} \sum_{j=1}^{r_n} E\Big[g(W_{n,j}) \sum_{i=1}^{r_n} \mathbb{1}_{\{X_{n,i}\neq 0\}}\Big] - \frac{1}{r_n v_n s_n a_n} \cdot r_n E[g(W_{n,1})] \cdot r_n v_n \\
&= \frac{1}{r_n v_n s_n a_n} \sum_{j=1}^{r_n} E\Big[g(W_{n,1}) \sum_{i=2-j}^{r_n-j+1} \mathbb{1}_{\{X_{n,i}\neq 0\}}\Big] - \frac{r_n E[g(W_{n,1})]}{s_n a_n}. \qquad (3.4.14)
\end{aligned}$$

Likewise, $c^{(d,v)}$ is the limit of

$$\begin{aligned}
&\frac{1}{r_n v_n a_n} Cov\Big(\sum_{k=1}^{r_n/s_n} g(W_{n,(k-1)s_n+1}), \sum_{i=1}^{r_n} \mathbb{1}_{\{X_{n,i}\neq 0\}}\Big)\\
&= \frac{1}{r_n v_n a_n} E\Big[\sum_{k=1}^{r_n/s_n} g(W_{n,(k-1)s_n+1}) \sum_{i=1}^{r_n} \mathbb{1}_{\{X_{n,i}\neq 0\}}\Big] - \frac{1}{r_n v_n a_n} E\Big[\sum_{j=1}^{r_n/s_n} g(W_{n,j})\Big] E\Big[\sum_{i=1}^{r_n} \mathbb{1}_{\{X_{n,i}\neq 0\}}\Big]\\
&= \frac{1}{r_n v_n a_n} \sum_{k=1}^{r_n/s_n} E\Big[g(W_{n,1}) \sum_{i=1-(k-1)s_n}^{r_n-(k-1)s_n} \mathbb{1}_{\{X_{n,i}\neq 0\}}\Big] - \frac{r_n E[g(W_{n,1})]}{s_n a_n}\\
&= \frac{1}{r_n v_n s_n a_n} \sum_{j=1}^{r_n} E\Big[g(W_{n,1}) \sum_{i=1-\lfloor (j-1)/s_n\rfloor s_n}^{r_n-\lfloor (j-1)/s_n\rfloor s_n} \mathbb{1}_{\{X_{n,i}\neq 0\}}\Big] - \frac{r_n E[g(W_{n,1})]}{s_n a_n}. \qquad (3.4.15)
\end{aligned}$$

In the last step, we added some summands in the first sum, such that in this sum each summand from the penultimate line occurs s_n times in the last line. This is why the normalization now contains an extra s_n.

It remains to show that the limes superior of the following difference between the right hand sides of (3.4.15) and (3.4.14) is not positive. To this end, note that

$$\begin{aligned}
&\frac{1}{r_n v_n s_n a_n} \sum_{j=1}^{r_n} E\Big[g(W_{n,1})\Big(\sum_{i=1-\lfloor (j-1)/s_n\rfloor s_n}^{r_n-\lfloor (j-1)/s_n\rfloor s_n} \mathbb{1}_{\{X_{n,i}\neq 0\}} - \sum_{i=2-j}^{r_n-j+1} \mathbb{1}_{\{X_{n,i}\neq 0\}}\Big)\Big]\\
&\quad\le \frac{1}{r_n v_n s_n a_n} \sum_{j=2}^{r_n} \sum_{i=r_n-j+2}^{r_n-\lfloor (j-1)/s_n\rfloor s_n} E\Big[g(W_{n,1})\mathbb{1}_{\{X_{n,i}\neq 0\}}\Big]\\
&\quad= \frac{1}{r_n v_n s_n a_n} \sum_{i=2}^{r_n} \sum_{j=r_n-i+2}^{(\lfloor (r_n-i)/s_n\rfloor+1)s_n} E\Big[g(W_{n,1})\mathbb{1}_{\{X_{n,i}\neq 0\}}\Big]\\
&\quad\le \frac{1}{r_n v_n a_n} \sum_{i=2}^{r_n} E\Big[g(W_{n,1})\mathbb{1}_{\{X_{n,i}\neq 0\}}\Big]. \qquad (3.4.16)
\end{aligned}$$

In the first step, we used that the sums in the difference are the same for $j = 1$, and for the other j both sums have the same length r_n and the second sum starts with a smaller first index than the first sum. Since all summands are positive, an upper bound for the difference is the sum over the summands of the first sums which do not occur in the second sum.

Note that $E[g(W_{n,1})] = O(s_n a_n v_n)$ by (3.3.3). Using

$$\begin{aligned}
E\Big[g(W_{n,1})\mathbb{1}_{\{X_{n,i}\neq 0\}}\Big] &\le E[g(W_{n,1})]P(X_{n,i}\neq 0) + 2\|g\|_\infty \beta^X_{n,i-s_n-1}\\
&= O(s_n a_n v_n^2) + 2\|g\|_\infty \beta^X_{n,i-s_n-1}
\end{aligned}$$

for $i > s_n + k_n$ (cf. Doukhan (1994), Section 1.2, Lemma 3 and Section 1.1, Prop. 1) and

$E\Big[g(W_{n,1})\mathbb{1}_{\{X_{n,i}\neq 0\}}\Big] \leq \|g\|_\infty v_n$ for $i \leq s_n + k_n$, we conclude that (3.4.16) is bounded by

$$\frac{s_n + k_n}{r_n a_n}\|g\|_\infty + \mathrm{O}(s_n v_n) + \frac{2\|g\|_\infty}{r_n v_n a_n} \sum_{l=k_n}^{r_n} \beta_{n,l}^X$$

which tends to 0 under the given conditions, i.e. $c^{(d,v)} - c^{(s,v)} \leq 0$. Thus, we have $c^{(s,v)} \geq c^{(d,v)}$. For the variances of the disjoint and sliding blocks estimator, it follows

$$\tilde{c}^{(s)} = c^{(s)} + \xi^2 c^{(v)} - 2\xi c^{(s,v)} \leq c^{(d)} + \xi^2 c^{(v)} - 2\xi c^{(d,v)} = \tilde{c}^{(d)}$$

which is the assertion. For the last equation, note that the sign of ξ is the same as the sign of g, since (3.3.4) holds. □

Similarly to the proof of the previous theorem, one can establish a lower bound on the difference between the pre-asymptotic covariances of $c^{(s,v)}$ and $c^{(d,v)}$.

Proof of Corollary 3.3.7. From the previous proof of Theorem 3.3.6 we already know $c^{(d,v)} - c^{(s,v)} \leq 0$. Here, we will show $c^{(d,v)} - c^{(s,v)} = 0$ which implies the assertion.
The limes superior of the following difference between the right hand sides of (3.4.14) and (3.4.15) is not positive. To this end, note that

$$\begin{aligned}
&\frac{1}{r_n v_n s_n a_n} \sum_{j=1}^{r_n} E\Bigg[g(W_{n,1})\Bigg(\sum_{i=2-j}^{r_n-j+1} \mathbb{1}_{\{X_{n,i}\neq 0\}} - \sum_{i=1-\lfloor (j-1)/s_n \rfloor s_n}^{r_n-\lfloor (j-1)/s_n \rfloor s_n} \mathbb{1}_{\{X_{n,i}\neq 0\}}\Bigg)\Bigg] \\
&\quad\leq \frac{1}{r_n v_n s_n a_n} \sum_{j=2}^{r_n} \sum_{i=2-j}^{-\lfloor (j-1)/s_n \rfloor s_n} E\Big[g(W_{n,1})\mathbb{1}_{\{X_{n,i}\neq 0\}}\Big] \\
&\quad= \frac{1}{r_n v_n s_n a_n} \sum_{i=2-r_n}^{0} \sum_{j=2-i}^{(\lfloor -i/s_n \rfloor+1)s_n} E\Big[g(W_{n,1})\mathbb{1}_{\{X_{n,i}\neq 0\}}\Big] \\
&\quad\leq \frac{1}{r_n v_n a_n} \sum_{i=2-r_n}^{0} E\Big[g(W_{n,1})\mathbb{1}_{\{X_{n,i}\neq 0\}}\Big]. \qquad (3.4.17)
\end{aligned}$$

In the first step, we used that the sums in the difference are the same for $j = 1$, both sums have the same length r_n and the second sum starts with a larger or equal index than the first sum. Since all summands are positive, an upper bound for the difference is the sum of the first summands of the first sums which do not occur in the second sum.
Using

$$\begin{aligned}
E\Big[g(W_{n,1})\mathbb{1}_{\{X_{n,i}\neq 0\}}\Big] &\leq E[g(W_{n,1})]P(X_{n,i} \neq 0) + 2\|g\|_\infty \beta_{n,|i|-s_n-1}^X \\
&= \mathrm{O}(s_n a_n v_n^2) + 2\|g\|_\infty \beta_{n,|i|-s_n-1}^X
\end{aligned}$$

for $i < -s_n - k_n$ (cf. Doukhan (1994), Section 1.2, Lemma 3 and Section 1.1, Prop. 1) and

$E\Big[g(W_{n,1})\mathbb{1}_{\{X_{n,i}\neq 0\}}\Big] \leq \|g\|_\infty v_n$ for $i \geq -s_n - k_n$, we conclude that (3.4.17) is bounded by

$$\frac{s_n + k_n}{r_n a_n}\|g\|_\infty + \mathrm{O}(s_n v_n) + \frac{2\|g\|_\infty}{r_n v_n a_n}\sum_{l=k_n}^{r_n} \beta_{n,l}^X$$

which tends to 0 under the given conditions, i.e. $c^{(s,v)} - c^{(d,v)} \leq 0$.
Thus, together with the proof of Theorem 3.3.6, it follows that $c^{(s,v)} = c^{(d,v)}$, which implies $\tilde{c}^{(d)} - \tilde{c}^{(s)} = c^{(d)} - c^{(s)} \geq 0$. □

For the proof of Corollary 3.3.8 about the Loewner order, we first introduce some notation: Since $|\mathcal{G}| < \infty$, we can enumerate the functions in $\mathcal{G}$ by $\mathcal{G} = \{g_1, ..., g_{|\mathcal{G}|}\}$. We define $\tilde{g}_w := \sum_{i=1}^{|\mathcal{G}|} w_i g_i$ with $w = (w_1, ..., w_{|\mathcal{G}|}) \in \mathbb{R}^{|\mathcal{G}|}$ and we define the set of all these $\tilde{g}_w$ as $\tilde{\mathcal{G}} := \{\tilde{g}_w : w \in \mathbb{R}^{|\mathcal{G}|}, g_j \in \mathcal{G}, j = 1, ..., |\mathcal{G}|\}$.

Proof of Corollary 3.3.8. The inequality $C^{(s)} \leq_L C^{(d)}$ is equivalent to

$$\sum_{i=1}^{|\mathcal{G}|}\sum_{j=1}^{|\mathcal{G}|} w_i w_j c^{(s)}(g_i, g_j) \leq \sum_{i=1}^{|\mathcal{G}|}\sum_{j=1}^{|\mathcal{G}|} w_i w_j c^{(d)}(g_i, g_j)$$

for all $(w_g)_{g\in\mathcal{G}} \in \mathbb{R}^{|\mathcal{G}|}$. It holds that

$$\begin{aligned}
\sum_{i=1}^{|\mathcal{G}|}\sum_{j=1}^{|\mathcal{G}|} w_i w_j c^{(s)}(g_i, g_j) &= \lim_{n\to\infty} \frac{1}{r_n v_n s_n^2 a_n^2}\sum_{i=1}^{|\mathcal{G}|}\sum_{j=1}^{|\mathcal{G}|} w_i w_j Cov\Big(\sum_{k=1}^{r_n} g_i(W_{n,k}), \sum_{k=1}^{r_n} g_j(W_{n,k})\Big)\\
&= \lim_{n\to\infty} \frac{1}{r_n v_n s_n^2 a_n^2} Cov\Big(\sum_{k=1}^{r_n}\sum_{i=1}^{|\mathcal{G}|} w_i g_i(W_{n,k}), \sum_{k=1}^{r_n}\sum_{j=1}^{|\mathcal{G}|} w_j g_j(W_{n,k})\Big)\\
&= \lim_{n\to\infty} \frac{1}{r_n v_n s_n^2 a_n^2} Var\Big(\sum_{k=1}^{r_n} \tilde{g}_w(W_{n,k})\Big) := c_w^{(s)}.
\end{aligned}$$

Therefore, the limit $c_w^{(s)}$ exists. Likewise, for $c^{(d)}$ we have

$$\sum_{i=1}^{|\mathcal{G}|}\sum_{j=1}^{|\mathcal{G}|} w_i w_j c^{(d)}(g_i, g_j) = \lim_{n\to\infty} \frac{1}{r_n v_n a_n^2} Var\Big(\sum_{k=1}^{r_n/s_n} \tilde{g}_w(W_{n,ks_n+1})\Big) =: c_w^{(d)}$$

and $c_w^{(d)}$ exists.
Since all conditions of Lemma 3.3.4 are satisfied for the single function $\tilde{g}_w$ by the linearity of the covariance, the lemma yields $c_w^{(s)} \leq c_w^{(d)}$. All in all, we have

$$\sum_{i=1}^{|\mathcal{G}|}\sum_{j=1}^{|\mathcal{G}|} w_i w_j c^{(s)}(g_i, g_j) = c_w^{(s)} \leq c_w^{(d)} = \sum_{i=1}^{|\mathcal{G}|}\sum_{j=1}^{|\mathcal{G}|} w_i w_j c^{(d)}(g_i, g_j)$$

which completes the proof. □

Chapter 4

Cluster index estimation

In Chapter 3 an abstract limit theorem and a more concrete limit theorem for sliding blocks statistics were developed. This chapter is dedicated to a class of indexes which can describe aspects of the extremal dependence structure of a stationary time series $(X_t)_{t\in\mathbb{Z}}$. The proofs in this chapter are mainly applications of the limit theorems from the previous chapter. Thus, this chapter gives relevant examples for the application of the theory developed so far.

4.1 Cluster indexes

We will consider cluster indexes for stationary time series $(X_t)_{t\in\mathbb{Z}}$ in general and the extremal index and stop-loss index specifically. Cluster indexes are values which describe some specific parts of the extremal behavior of a stationary time series. Often extreme events do not occur alone, but there are temporal dependencies and they occur in a collection, in so-called clusters. Cluster indexes can be used to describe facets of this clustering behavior. Thus, cluster indexes can be important variables for the understanding of the extremal behavior of a stationary time series. They were rigorously defined and motivated in Kulik and Soulier (2020), Section 6 and 10, one way to define these indexes is given in the next definition.

Denote $l_0 = \{(x_t)_{t\in\mathbb{Z}} \in (\mathbb{R}^d)^{\mathbb{Z}} : |x_t| \to 0 \text{ for } |t| \to \infty\}$ as the space of $\mathbb{R}^d$ valued double sided sequences converging to 0 as $|t| \to \infty$. In this section, we consider elements and convergences in l_0, equipped with the supremum norm $\|\cdot\|_\infty$. If some element $(x_s, .., x_t) \in (\mathbb{R}^d)^{t-s+1}$ occurs, we will interpret it as an element $(x_h)_{h\in\mathbb{Z}} \in l_0$ with $x_h = 0$ for $h \notin \{s, .., t\}$.

Definition 4.1.1 (Cluster index). *Let $(X_t)_{t\in\mathbb{Z}}$ be an $\mathbb{R}^d$-valued regular varying time series with index α and tail process $(Y_t)_{t\in\mathbb{Z}}$. Set $v_n = P(\|X_0\| > u_n)$. Let $r_n \in \mathbb{N}$ and u_n be sequences such that $nv_n \to \infty$ and $r_n v_n \to 0$.*

For a bounded or non-negative functional H on l_0, a general ***cluster index*** *$\nu^*(H)$ is* *defined as*

$$\nu^*(H) := \lim_{n\to\infty} \frac{1}{r_n v_n} E\left[H\Big(\frac{X_1}{u_n}, ..., \frac{X_{r_n}}{u_n}\Big)\right]$$

if the limit exists. We call H the ***cluster functional****.*

Depending on the cluster functional H, these cluster indexes can describe some specific behavior of the extremes of the stationary time series $(X_t)_{t\in\mathbb{Z}}$. One special example for a cluster index is the extremal index θ, which will be discussed in detail in Section 4.2. Some more well known cluster indexes are the deviation index (Mikosch and Wintenberger, 2013, 2014) and the cluster size distribution, where the index π_j is the probability that a cluster of extremes contains j extreme values. (Hsing, 1991; Drees and Rootzén, 2010).
Before we define estimators for $\nu^*(H)$, an important question is when and for which H the cluster index exists at all. We will consider the general cluster index only for non-negative, real-valued and stationary time series $(X_t)_{t\in\mathbb{Z}}$ which are regularly varying with index α. In this case the tail process $(Y_t)_{t\in\mathbb{Z}}$ exists (Theorem 2.1.7) and we will assume that the tail process satisfies the summability condition (SC), i.e. $\|Y\|_\alpha^\alpha < \infty$ a.s. By Lemma 2.2.3 this ensures that $(Y_t)_{t\in\mathbb{Z}}$ is a.s. an element of $l_\alpha \subset l_0$.
Denote $M_{i,j} = \max_{i\le k\le j} \|X_k\|$ and $U^*_{s,t} = \sup_{s\le i\le t} \|U_i\|$ for all $-\infty \le s < t \le \infty$ and a stochastic process $(U_t)_{t\in\mathbb{Z}}$.
The first proposition states conditions for the existence of the limit in Definition 4.1.1 and it also states an alternative representation. In preparation we define

$$\vartheta := P(Y^*_{-\infty,-1} \le 1) = P(Y^*_{1,\infty} \le 1)$$

as the candidate extremal index (Basrak and Segers (2009), Section 4). The process Q is defined by $Q = Z/Y^*_{-\infty,\infty}$ where $Z = (Z_t)_{t\in\mathbb{Z}}$ has the same distribution as $(Y_t)_{t\in\mathbb{Z}}$ conditioned on $Y^*_{-\infty,-1} \le 1$, i.e. $\mathcal{L}((Z_t)_{t\in\mathbb{Z}}) = \mathcal{L}((Y_t)_{t\in\mathbb{Z}} \mid Y^*_{-\infty,-1} \le 1)$ (Basrak and Segers (2009), Remark 4.6). The condition in the definition of Z means that the first value with norm larger than 1 occurs at time point 0, since $\|Y_0\| \ge 1$ a.s. This processes Z and Q are well known in the literature, see e.g. Planinić and Soulier (2018), Definition 3.5, or more abstract in Dombry et al. (2018).
Note that ϑ equals the extremal index θ under the conditions (θ1) and (θP) below, see Section 4.2 for details. For the existence of $\nu^*(H)$, recall the well known anti-clustering condition (AC) from page 12.

Proposition 4.1.2. *Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary and regular varying time series with tail process $(Y_t)_{t\in\mathbb{Z}}$. Let $nv_n \to \infty$ and $r_n v_n \to 0$. If (AC) holds, then for all bounded shift invariant functions H (i.e. $H((y_t)_{t\in\mathbb{Z}}) = H((y_{t+1})_{t\in\mathbb{Z}})$) with support separated from 0 the limit*

$$\nu^*(H) = \lim_{n\to\infty} \frac{1}{r_n v_n} E\left[H\Big(\frac{X_1}{u_n}, ..., \frac{X_{r_n}}{u_n}\Big)\right]$$

exists. Moreover,

$$\nu^*(H) = \vartheta \int_0^\infty E[H(rQ)]\alpha r^{-\alpha-1} dr. \tag{4.1.1}$$

The proof of the convergence can be found in Kulik and Soulier (2020), Theorem 6.2.5, the representation (4.1.1) is stated in equation (6.2.3) of that book. The alternative representation (4.1.1) could also be used to define the cluster index and it shows the connection to the tail measure, see Planinić and Soulier (2018).
Kulik and Soulier (2020) (Section 6, page 156) also show, that if $H \geq 0$ is a shift-invariant functional with $H(y) = 0$ if $\|y\|_\infty \leq \varepsilon$ for some $\varepsilon > 0$ (i.e. the support of H is separated from 0), then the representation (4.1.1) can be used to show

$$\nu^*(H) = \varepsilon^{-\alpha} E\left[H(\varepsilon Y)\mathbb{1}_{\{Y^*_{-\infty,-1}\leq 1\}}\right] = \varepsilon^{-\alpha} E\left[H(\varepsilon Y)\mathbb{1}_{\{Y^*_{1,\infty}\leq 1\}}\right]. \tag{4.1.2}$$

Now, where we have quite general conditions under which general cluster indexes exist, the statistical question arises how and how well one could estimate them. Therefore, we want to motivate three estimators for a general cluster index and analyze the asymptotic behavior. The first two estimators are classical disjoint and sliding blocks estimators, which fit in the setting of the previous chapter. Both estimators are block based extreme value statistics which are motivated by the interpretation that all large values in such a block form a cluster of extremes. Another interpretation of clusters of extremes is that all large values which are not separated in time by a certain number of smaller values form a cluster of extremes. This leads to our third estimator, a so-called runs estimators. The best-known example of a runs estimator is defined for the extremal index, cf. Section 4.2.2. Such runs estimators can be considered as a special type of sliding blocks estimators and can thus be analyzed with the techniques developed in Section 3.2. This interpretation broadens the possible field of application of the sliding blocks limit theorem developed in Section 3.2. Moreover, the theory developed there offers a first general framework to derive asymptotic results for general runs estimators.
We start with the motivation of the disjoint and sliding blocks (pseudo) estimators, which can be obtained as empirical counterparts of the expected value in the definition of $\nu^*(H)$. Based on observations $X_1, ..., X_n$, they are given by

$$\hat{\nu}^*(H)_n^d := \frac{1}{nv_n} \sum_{i=1}^{\lfloor n/s_n \rfloor} H\left(\frac{X_{(i-1)s_n+1}}{u_n}, ..., \frac{X_{is_n}}{u_n}\right) \text{ and} \tag{4.1.3}$$
$$\hat{\nu}^*(H)_n^s := \frac{1}{s_n n v_n} \sum_{i=1}^{n-s_n+1} H\left(\frac{X_i}{u_n}, ..., \frac{X_{i+s_n-1}}{u_n}\right)$$

for some block length $s_n = o(n)$. The expectations of the estimators are given by

$$E[\hat{\nu}^*(H)_n^d] = \left\lfloor \frac{n}{s_n} \right\rfloor \frac{1}{nv_n} E\left[H\left(\frac{X_1}{u_n}, ..., \frac{X_{s_n}}{u_n}\right)\right],$$
$$E[\hat{\nu}^*(H)_n^s] = \frac{n - s_n + 1}{n} \frac{1}{s_n v_n} E\left[H\left(\frac{X_1}{u_n}, ..., \frac{X_{s_n}}{u_n}\right)\right]$$

for the disjoint and sliding blocks case, respectively. Those expectations both converge to $\nu^*(H)$ under suitable conditions as stated in Proposition 4.1.2. Since v_n is usually unknown itself, this has to be replaced by a reasonable estimator, for example by

$$\hat{v}_n := \frac{1}{n}\sum_{i=1}^{n-s_n+1} \mathbb{1}_{\{\|X_i\|>u_n\}}.$$

One could replace the upper end of the sum by n, however for simplicity of the proofs we will use this estimator for v_n. We will see that both versions are asymptotic equivalent, meaning that the difference converges to 0 in probability, i.e.

$$\left|\frac{1}{n}\sum_{i=1}^{n-s_n+1} \mathbb{1}_{\{\|X_i\|>u_n\}} - \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{\|X_i\|>u_n\}}\right| \overset{d}{=} \frac{1}{n}\sum_{i=1}^{s_n-1} \mathbb{1}_{\{\|X_i\|>u_n\}} = \mathrm{o}_P\left(\sqrt{\frac{v_n}{n}}\right)$$

under our conditions used used below, in particular since $s_n = \mathrm{o}(\sqrt{nv_n})$.
As third estimator we will propose the runs estimator for cluster indexes. This can be motivated using the following proposition.

Proposition 4.1.3. *Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary and regular varying time series with tail process $(Y_t)_{t\in\mathbb{Z}}$. Let $u_n \to \infty$ be a scaling sequence and r_n be an intermediary sequence ($r_n \to \infty$, $r_n = o(n)$). If (AC) holds, then for all $x > 0$ the weak convergence*

$$\mathcal{L}\left(\frac{X_{-r_n}}{xu_n}, ..., \frac{X_{r_n}}{xu_n} \,\middle|\, \|X_0\| > xu_n\right) \to \mathcal{L}((Y_t)_{t\in\mathbb{Z}})$$

holds in l_0.

This proposition is a simple corollary of Lemma 2.1.9, where a more general statement is given. An alternative proof can be found in Kulik and Soulier (2020), Theorem 6.1.1. With this proposition and (4.1.2) we motivate the runs estimator. Note that due to the restriction on H for the application of (4.1.2), this runs estimator is only well motivated for bounded functions H with $H(y) = 0$ if $\|y\|_\infty < \varepsilon$ for some $\varepsilon > 0$. To ease the following notation, we now consider $\varepsilon = 1$. The runs (pseudo-)estimator of $\nu^*(H)$ based on observations $X_1, ..., X_n$ can be defined by

$$\hat{\nu}^*(H)^r_n := \frac{1}{(n-2s_n)v_n}\sum_{j=s_n+1}^{n-s_n} H\left(\frac{X_{j-s_n}}{u_n}, ..., \frac{X_{j+s_n}}{u_n}\right)\mathbb{1}_{\{\|X_j\|>u_n\}}\mathbb{1}_{\{M_{j+1,j+s_n}\le u_n\}}. \quad (4.1.4)$$

This estimator is called runs estimator, because each nonzero summand j with $\|X_j\| > u_n$ has at least s_n subsequent summands which are 0, i.e. the summands different from 0 are separated by runs of length s_n. This separation is used such that each cluster of extremes has impact only on the value of one summand of the estimator. This is the version with a run after an extreme observation. Likewise one can define a runs estimator with runs before the extreme observations:

$$\hat{\nu}^*(H)_n^{r'} := \frac{1}{(n-2s_n)v_n} \sum_{j=s_n+1}^{n-s_n} H\left(\frac{X_{j-s_n}}{u_n}, ..., \frac{X_{j+s_n}}{u_n}\right) \mathbb{1}_{\{\|X_j\|>u_n\}} \mathbb{1}_{\{M_{j-s_n,j-1} \le u_n\}}.$$

The expected values of this estimators are

$$E[\hat{\nu}^*(H)_n^r] = \frac{1}{v_n} E\left[H\left(\frac{X_{-s_n}}{u_n}, ..., \frac{X_{s_n}}{u_n}\right) \mathbb{1}_{\{\|X_0\|>u_n\}} \mathbb{1}_{\{M_{1,s_n} \le u_n\}}\right],$$
$$E[\hat{\nu}^*(H)_n^{r'}] = \frac{1}{v_n} E\left[H\left(\frac{X_{-s_n}}{u_n}, ..., \frac{X_{s_n}}{u_n}\right) \mathbb{1}_{\{\|X_0\|>u_n\}} \mathbb{1}_{\{M_{s_n,-1} \le u_n\}}\right].$$

Under the conditions of Proposition 4.1.3, these expectations converge to the cluster indexes for bounded cluster functionals H with $H(y) = 0$ if $\|y\|_\infty < 1$:

$$E[\hat{\nu}^*(H)_n^r] \to E\left[H(Y) \mathbb{1}_{\{Y^*_{1,\infty} \le 1\}}\right] = \nu^*(H)$$

where the last equation holds because of (4.1.2) with $\varepsilon = 1$. Note that we replaced r_n by s_n in Proposition 4.1.3. For $s_n = o(r_n)$ the conditions of the proposition are met, in particular (AC) also holds for $s_n \le r_n$. Similar $E[\hat{\nu}^*(H)_n^{r'}] \to E\left[H(Y) \mathbb{1}_{\{Y^*_{-\infty,-1} \le 1\}}\right] = \nu^*(H)$. Hence, the runs (pseudo-)estimators are asymptotically unbiased. Again, since v_n is usually unknown it has to be replaced by $\hat{v}_n$.

For the sliding blocks estimator and the runs estimator the asymptotic distributions can be derived under the corresponding conditions from Theorem 3.2.1. This result also allows the derivation of the joint convergences of a larger family of cluster indexes. The asymptotics of the disjoint blocks estimator can be derived with the more abstract Theorem 3.1.4. The application of the theory from Chapter 3 to achieve asymptotic results for all three estimators underlines one big advantage of the abstract setting: it suffices to consider this one setting to analyze the three considered estimators for one index and to get actual comparable conditions. Therefore, the motto of this chapter could be: *three types of estimator, one method of proof - unifying the settings.*

Since the cluster functionals H are quite general, the conditions of Theorem 3.2.1 would remain unchanged for a limit theorem for general cluster index estimators. The only difference to the conditions in Section 3.2 would be that the function $g((X_{j+t}/u_n)_{|t| \le s_n})$ is replaced by the more specific, but still quite general function

$$H\left(\frac{X_{j-s_n}}{u_n}, ..., \frac{X_{j+s_n}}{u_n}\right) \mathbb{1}_{\{\|X_j\|>u_n\}} \mathbb{1}_{\{M_{j+1,j+s_n} \le u_n\}}$$

for runs estimators (for sliding blocks estimators similar). Further simplifications are not possible without additional assumptions on H or the time series, but one could replace some conditions by stronger ones. However, for some special family of cluster indexes with a concrete function H one could derive a more specific limit theorem under simpler conditions. In the following sections we show by the example of the extremal index and stop-loss indexes how one can derive this more specific limit theorems with simpler sets of conditions.

Recently, Cissokho and Kulik (2021) considered disjoint and sliding blocks estimators for cluster indexes and developed an alternative limit theorem for general estimators and some special examples. They used estimators based on order statistics and derived a limit theorem for them. For this they used some other kind of conditions than we used in Chapter 3. In particular, they applied a stronger β-mixing condition, another condition to control extremal dependence and some ANSJB condition which controls the occurrence of small jumps in the time series. In their setting they proved asymptotic normality of the sliding and disjoint blocks estimator. Moreover, they proved that the asymptotic variance is equal for both estimators under their assumptions.

If the asymptotic normality of $\hat{\nu}^*(H)_n^s$ and $\hat{\nu}^*(H)_n^d$ is derived with the theory of Chapter 3, then Theorem 3.3.6 shows that the asymptotic variance for the disjoint blocks estimator is greater or equal to the asymptotic variance of the sliding blocks counterpart. Thus, Cissokho and Kulik (2021) got a stronger result under more specific conditions than we used in the much more general setting of Section 3.3.

In the following sections we derive specific conditions for asymptotic normality of the disjoint and sliding blocks estimators of the extremal index (Section 4.2) and the stop-loss index (Section 4.3). Under these conditions we also derive the same asymptotic variance for the disjoint and sliding blocks estimator. In addition to the mentioned paper we also consider the runs estimator, which in the case of the extremal index has also the same variance as the other two estimators.

4.2 Extremal index

The main results from this section have already been published in advance in Section 3 of Drees and Neblung (2021).

In this section, we consider only $\mathbb{R}^+$-valued time series $(X_t)_{t\in\mathbb{Z}}$, in particular $d = 1$. Using the cluster functional $H((x_t)_{t\in\mathbb{Z}}) = \mathbb{1}_{\{\max_{t\in\mathbb{Z}} x_t > 1\}}$ one obtains the extremal index as special cluster index, if one considers only non-negative time series. The extremal index is analyzed in the literature since Leadbetter (1983), including disjoint blocks estimators and runs estimators. In order to built a bridge to the literature, we introduce the extremal index as a new variable and not only as a special case of the generalizing concept of cluster indexes. In particular, the extremal index can be defined without the assumption of regular variation of $(X_t)_{t\in\mathbb{Z}}$ as used for the general cluster index above.

Definition 4.2.1 (Extremal index (Leadbetter, 1983))**.** *Let $(X_t)_{t\in\mathbb{Z}}$ be a $\mathbb{R}$-valued stationary time series. $(X_t)_{t\in\mathbb{Z}}$ possesses the* ***extremal index*** $\theta \in [0, 1]$ *if for some thresholds*

$u_n(\tau)$ with $nP(X_0 > u_n(\tau)) \to \tau > 0$ for some $\tau > 0$

$$\lim_{n\to\infty} P\left(\max_{1\le i\le n} X_i \le u_n(\tau)\right) = e^{-\theta\tau}$$

holds.

This definition implies that for large n it is $P(\max_{1\le i\le n} X_i \le u_n(\tau)) \sim F^{n\theta}(u_n(\tau))$, where F is the distribution function of X_1. If $(X_t)_{t\in\mathbb{Z}}$ is an iid. sequence, then $\theta = 1$. For $0 < \theta < 1$ there have to be some dependences in the time series, in case of dependences $\theta = 1$ is still possible. In general, θ is defined for values in $[0,1]$, but in what follows we assume $\theta > 0$ and exclude the degenerate case $\theta = 0$. In this degenerate $\theta = 0$ case $\max_{1\le i\le n} X_i$ would not exceed the threshold $u_n(\tau)$ for n large enough with probability 1, which would imply that asymptotically extreme events occur only with lower order than implied by the thresholds $u_n(\tau)$, i.e. another normalization would be needed.
There are several other characterizations of the extremal index. Under suitable additional conditions, the extremal index is the reciprocal of the expected length of a cluster, where a cluster is a block of X_i's which exceeds a high thresholds, see (4.2.1) and Smith and Weissman (1994). Due to this interpretation, the extremal index is an important parameter for measuring the degree of extremal dependence of $(X_t)_{t\in\mathbb{Z}}$. Therefore, the estimation of θ can be an important step for the analysis of the extremal dependence of a time series. There is much literature that deals with this estimation problem, see e.g. Smith and Weissman (1994), Ferro and Segers (2003), Süveges (2007), Robert et al. (2009), Berghaus and Bücher (2018), Bücher and Jennessen (2020*a*) among others.
Besides a popular disjoint blocks estimator for θ, we discuss the corresponding sliding blocks estimator and we will consider a well known runs estimator.
Let $(X_t)_{t\in\mathbb{Z}}$ be a real-valued stationary time series and recall $M_{i,j} = \max(X_i, ..., X_j)$. If the extremal index exists then, under weak additional conditions,

$$\frac{P(M_{1,k_n} > u_n)}{k_n P(X_1 > u_n)} \to \theta, \tag{4.2.1}$$

for sequences $k_n \to \infty$ and u_n such that $k_n P(X_1 > u_n) \to 0$. In particular, this holds if $\beta^X_{n,l_n}/(k_n v_n) \to 0$ for some $l_n = \mathrm{o}(k_n)$ (with $X_{n,i} = X_i/u_n \mathbb{1}_{\{X_i > u_n\}}$ for the β-mixing coefficients, cf. Leadbetter (1983), Theorem 3.4). One other possible and weaker condition is the $\mathrm{D}(u_n)$ condition introduced by Leadbetter (1983), this condition is in particular satisfied under the β-mixing condition mentioned before.
This representation (4.2.1) together with Proposition 4.1.2 shows, that the cluster functional $H((x_t)_{t\in\mathbb{Z}}) = \mathbb{1}_{\{\max_{t\in\mathbb{Z}} |x_t| > 1\}}$ actually defines the extremal index as cluster index.
The first estimator we consider is the disjoint blocks estimator

$$\hat{\theta}^d_n := \frac{\sum_{i=1}^{\lfloor n/s_n \rfloor} \mathbb{1}_{\{M_{(i-1)s_n+1, is_n} > u_n\}}}{\sum_{i=1}^{n-s_n+1} \mathbb{1}_{\{X_i > u_n\}}}.$$

This estimator was first proposed by Hsing (1991) and is an empirical counterpart of (4.2.1) using disjoint blocks to estimate the numerator for $k_n = s_n$. Here, s_n is the length of the considered disjoints blocks.
Hsing (1991) already stated an asymptotic result for $\hat{\theta}_n^d$ under some tailor-made conditions. We will show the asymptotics under different assumptions and in the same setting as for the runs and sliding blocks estimator.
The sliding blocks estimator is given by

$$\hat{\theta}_n^s := \frac{s_n^{-1} \sum_{i=1}^{n-s_n+1} \mathbb{1}_{\{M_{i,i+s_n-1} > u_n\}}}{\sum_{i=1}^{n-s_n+1} \mathbb{1}_{\{X_i > u_n\}}}.$$

This estimator is motivated like the disjoint blocks estimator by (4.2.1), the difference is that one uses sliding blocks for the empirical counterpart of the numerator. Due to the larger number of summands in the numerator, the normalization must be adapted, therefore, the factor s_n^{-1} appears. This factor assures convergence to a non-trivial (non 0) limit of the covariance (cf. the proofs below). The sliding blocks have an overlap which results in a higher dependency between the summands in the numerator, which must be considered in the asymptotic analysis. With Section 3.2 the asymptotics can still be treated. With $H((x_t)_{t\in\mathbb{Z}}) = \mathbb{1}_{\{\max_{t\in\mathbb{Z}} x_t > 1\}}$ this estimators corresponds to the block estimators in (4.1.3) for cluster indexes.
A (slightly modified) sliding blocks estimator for the extreme index was also given in Beirlant et al. (2004), Section 10.3.4. It was stated there, that this sliding version should be more efficient than the disjoint blocks estimator. The asymptotics given below is the first systematic investigation of the behavior of this sliding blocks estimator. In particular, we will see that the sliding blocks estimator has the same asymptotic variance as $\hat{\theta}_n^d$, which contradicts the suggestion of Beirlant et al. (2004).
The so-called runs estimator of θ is based on the following characterization of the extremal index:

$$P(M_{2,k_n} \le u_n \mid X_1 > u_n) \to \theta, \tag{4.2.2}$$

which was first proven by O'Brien (1987) under suitable conditions. With Theorem 1 from Segers (2003) it follows that (4.2.1) and (4.2.2) are equivalent conditions under some suitable conditions. One of these suitable conditions is the anti-clustering condition (AC) with r_n replaced by k_n. The equivalence of (4.2.1) and (4.2.2) ensures that the asymptotic distribution of all three estimators can be established under comparable conditions.
The runs estimator is defined as

$$\hat{\theta}_n^r := \frac{\sum_{i=1}^{n-s_n+1} \mathbb{1}_{\{X_i > u_n, M_{i+1,i+s_n-1} \le u_n\}}}{\sum_{i=1}^{n-s_n+1} \mathbb{1}_{\{X_i > u_n\}}},$$

where $s_n - 1$ is the length of the run. This estimator is the empirical counterpart of (4.2.2) with $k_n = s_n$ and was first suggested by Hsing (1993). The idea of this estimator

is the same as for the runs estimator for cluster indexes defined in (4.1.4), now in the special case of the extremal index. However, unlike in Section 4.1 we do not need regular variation here.
Its asymptotic normality was first established in Weissman and Novak (1998) who also proved the asymptotic normality of $\hat{\theta}_n^d$ under somewhat simpler conditions than Hsing (1991) including $\theta < 1$ and ϕ-mixing. For a very specific model, Weissman and Novak (1998) showed that the asymptotic variances of both estimators are the same, but they did not realize that this indeed holds true under quite general structural assumptions, as we will show below. Our analysis here uses a different approach and therefore also sets other conditions. Recently, Cai (2019) derived an asymptotic result for the runs estimator with random thresholds. She too, uses different dependency conditions which are not directly comparable with the conditions for asymptotic normality of the blocks estimators.
The asymptotic results in this section will be the first asymptotic results, which allow a direct comparison of the asymptotic variance of both blocks estimators and the runs estimator. It turns out that under mild conditions all three estimators of the extremal index have the same asymptotic variance. While the asymptotic normality of the disjoint blocks estimator and the runs estimator has already been proved by Weissman and Novak (1998), the equality of their asymptotic variances has been overlooked, because the variances were expressed differently. In addition, we establish the asymptotic normality of the sliding blocks analogously to the disjoint blocks estimator for the first time. This example demonstrates that, by analyzing different estimators of the same parameter in a unifying framework as developed in Chapter 3, one may gain new insights.
In the following we will state individual conditions for all three estimators, which are similar but not identical for the different estimators. However, one can formulate a set of not too strict uniform conditions under which the asymptotic normality of all three estimators holds.

(θ1) $(X_t)_{t\in\mathbb{Z}}$ is a $\mathbb{R}^+$-valued stationary time series. For $v_n := P(X_1 > u_n) \to 0$, one has $nv_n \to \infty$ and $s_n \to \infty$. In addition, there exists a sequence $(r_n)_{n\in\mathbb{N}}$ such that $s_n = \mathrm{o}(r_n)$, $r_nv_n \to 0$, $r_n = \mathrm{o}(\sqrt{nv_n})$ and $(n/r_n)\beta^X_{n,s_n-1} \to 0$.

(θ2) The limit

$$c := \lim_{n\to\infty} \frac{1}{r_nv_n} E\Big[\Big(\sum_{j=1}^{r_n} \mathbb{1}_{\{X_j>u_n\}}\Big)^2\Big]$$

exists in $[0,\infty)$.

(θP) For all $n \in \mathbb{N}$ and $k \in \mathbb{N}$ there exists $e_n(k)$ such that

$$e_n(k) \geq P(X_k > u_n \mid X_0 > u_n)$$

and $\lim_{n\to\infty} \sum_{k=1}^{r_n} e_n(k) = \sum_{k=1}^{\infty} \lim_{n\to\infty} e_n(k) < \infty$.

By Pratt's lemma (Pratt (1960)), condition (θP) enables us to exchange sums and limits in the calculation of covariances, it was also used e.g. by Drees et al. (2015).
Under (θ1) and (θP), both (4.2.1) and (4.2.2) hold for all $k_n \leq r_n$ such that $k_n \to \infty$. This follows from Theorem 1 and Corollary 2 of Segers (2003) in combination with the aforementioned result on convergence (4.2.1). In particular, under the above assumption Segers (2003) showed

$$P(M_{2,k_n} \leq u_n \mid X_1 > u_n) \sim \frac{P(M_{1,k_n} > u_n)}{k_n P(X_1 > u_n)},$$

i.e. (4.2.2) holds if and only if (4.2.1) holds.
Note that if (θ1) and (θP) hold for some sequence r_n, then the former is obviously fulfilled by $r_n^* := \lfloor r_n/s_n \rfloor s_n$, too, and ($\theta$P) remains true with r_n^* instead of r_n because of

$$\sum_{k=r_n^*+1}^{r_n} P(X_k > u_n \mid X_0 > u_n) \leq \frac{s_n}{v_n}(v_n^2 + \beta^X_{n,r_n^*}) \leq r_n v_n + \frac{n}{r_n}\beta^X_{n,s_n}\frac{r_n^2}{n v_n} \to 0.$$

Moreover, the arguments given in the proof of Lemma 3.3.3 show that the limit c in (θ2) does not change if we replace r_n with r_n^*. Thus, w.l.o.g. we may assume that r_n/s_n is a natural number (tending to ∞) for all $n \in \mathbb{N}$. This is used to ease the notation.
The limit c is the asymptotic variance of the estimator $\hat{v}_n$ for v_n in the denominator of each estimator. If (θP) holds and the positive part $(X_t^+)_{t\in\mathbb{Z}}$ of the time series is regularly varying, then c can be represented in terms of its tail process $(Y_t)_{t\in\mathbb{Z}}$, i.e. (θ2) holds with

$$\begin{aligned} c &= \lim_{n\to\infty} \frac{1}{r_n v_n} E\Big[\Big(\sum_{j=1}^{r_n} \mathbb{1}_{\{X_j > u_n\}}\Big)^2\Big] = \lim_{n\to\infty} \sum_{k=-r_n+1}^{r_n-1} \Big(1 - \frac{|k|}{r_n}\Big) P(X_k > u_n \mid X_0 > u_n) \\ &= \sum_{k=-\infty}^{\infty} \lim_{n\to\infty} \Big(1 - \frac{|k|}{r_n}\Big) P(X_k > u_n \mid X_0 > u_n) = \sum_{k\in\mathbb{Z}} P(Y_k > 1). \end{aligned} \tag{4.2.3}$$

In the third step we applied Pratt's Lemma and in the penultimate step we used the definition of the tail process. Alternatively, one may use the representation

$$\begin{aligned} c &= 1 + \lim_{n\to\infty} \sum_{k=1}^{r_n-1} \Big(1 - \frac{k}{r_n}\Big)\Big(P(X_k > u_n \mid X_0 > u_n) + P(X_0 > u_n \mid X_{-k} > u_n)\Big) \\ &= 1 + 2\sum_{k=1}^{\infty} P(Y_k > 1). \end{aligned}$$

In addition to the previous assumptions, we have to assume that the convergence (4.2.1) for $k_n = s_n$ and the convergence (4.2.2), respectively, is sufficiently fast to ensure that the bias of the block based estimators or runs estimators, respectively, is asymptotically negligible:

(B$_b$) $\sqrt{n v_n}\Big(\frac{P(M_{1,s_n} > u_n)}{s_n v_n} - \theta\Big) \to 0.$

(B$_r$) $\sqrt{nv_n}(P(M_{2,s_n} \le u_n \mid X_1 > u_n) - \theta) \to 0.$

Since (4.2.1) and (4.2.2) hold under (θ1) and (θP), these bias conditions impose only conditions on the rate of theses convergences. These bias conditions ensure that the bias converges faster to 0 than the stochastic error.
In the following Sections 4.2.1 and 4.2.2 we will show that under our unified conditions all three estimator have the same limit distribution.

Theorem 4.2.2. *If the conditions (θ1), (θ2) and (θP) are satisfied, then*

$$\sqrt{nv_n}(\hat{\theta}_n^{\sharp} - \theta) \xrightarrow{w} \mathcal{N}(0, \theta(\theta c - 1)),$$

provided (B$_b$) holds when $\sharp$ stands for 'd' or 's', and (B$_r$) holds when $\sharp$ stands for or 'r'.

This theorem directly follows from Theorems 4.2.4, 4.2.7 and 4.2.10, which will be established for each of the three estimators separately below. Note that (θ1) implies the more specific conditions (θ1S) (Lemma 4.2.5) and (θ1R) below.
In practice, usually the threshold u_n is replaced with some data driven choice $\hat{u}_n$, like an intermediate order statistic of the observed time series. By the techniques developed in Drees and Knezevic (2020), one may prove that these versions of the estimators of the extremal index asymptotically behave the same, provided $\hat{u}_n/u_n \xrightarrow{P} 1$ and the time series $(X_t^+)_{t\in\mathbb{Z}}$ is regularly varying. To this end, the results about the convergence of the fidis are not sufficient anymore, but the full process convergence is needed. The precise results for the sliding blocks estimator is discussed in Section 4.2.3.

4.2.1 Extremal index - disjoint and sliding blocks estimators

We start with the asymptotic analysis of the disjoint blocks estimator, and first discuss how one can embed this setting in the framework of Chapter 3. For all three estimators, first the numerator and denominator are examined individually and the bivariate asymptotic normality will be established using the theory of Chapter 3. Then the asymptotics of the estimator as a whole are derived using Lemma 3.3.5. In this concrete application only four different functions are considered, so that from the abstract setting only the results about the fidi convergence are needed.
For the application of Theorem 3.1.4 for the disjoint blocks estimator one possible choice for the treatment of the numerator would be $V_{n,i}^d := m_n^{-1/2}\mathbb{1}_{\{M_{(i-1)s_n+1,is_n} > u_n\}}$ in the setting of Section 3.1. However, we will choose $V_{n,i}$ differently to increase the comparability to the asymptotic results for the sliding blocks estimator and the runs estimator. For this we introduce artificial big blocks of length r_n which summarize some blocks of length s_n. Therefore, let

$$V_{n,i}^d := \frac{1}{\sqrt{m_n}} \sum_{j=1}^{r_n/s_n} \mathbb{1}_{\{M_{(i-1)r_n+(j-1)s_n+1,(i-1)r_n+js_n} > u_n\}},$$

$$\tilde{V}^d_{n,i} := \frac{1}{\sqrt{m_n}} \sum_{j=1}^{r_n/s_n-1} \mathbb{1}_{\{M_{(i-1)r_n+(j-1)s_n+1,(i-1)r_n+js_n} > u_n\}},$$
$$V^c_{n,i} := \frac{1}{\sqrt{m_n}} \sum_{j=1}^{r_n} \mathbb{1}_{\{X_{(i-1)r_n+j} > u_n\}}, \tag{4.2.4}$$
$$\tilde{V}^c_{n,i} := \frac{1}{\sqrt{m_n}} \sum_{j=1}^{r_n-s_n} \mathbb{1}_{\{X_{(i-1)r_n+j} > u_n\}},$$

for $i \in \{1, \dots, m_n\}$ and with $m_n = \lfloor (n - s_n + 1)/r_n \rfloor$. Here, $V^d_{n,i}$ is defined to deal with the numerator of $\hat{\theta}^d_n$ and $V^c_{n,i}$ is defined to treat the denominator. By the stationarity of $(X_t)_{t\in\mathbb{Z}}$ the stationarity of $(V^d_{n,i}, \tilde{V}^d_{n,i})_{1\le i\le m_n}$ follows. Let $p_n = P(M_{1,r_n} > u_n)$ and $v_n = P(X_1 > u_n)$.
For the asymptotic result for the disjoint blocks estimator we directly use the unified conditions (θ1), (θ2) and (θP). The next proposition states the bivariate asymptotic normality of the numerator and denominator of $\hat{\theta}^d_n$. This proposition follows directly from Theorem 3.1.4.

Proposition 4.2.3. *Suppose that the conditions (θ1), (θ2) and (θP) are satisfied. Then the weak convergence*

$$\begin{pmatrix} Z^d_n \\ Z^c_n \end{pmatrix} := \begin{pmatrix} p_n^{-1/2} \sum_{i=1}^{m_n} \left(V^d_{n,i} - E[V^d_{n,i}]\right) \\ p_n^{-1/2} \sum_{i=1}^{m_n} \left(V^c_{n,i} - E[V^c_{n,i}]\right) \end{pmatrix} \to \begin{pmatrix} Z^d \\ Z^c \end{pmatrix} \sim \mathcal{N}_2 \left(0, \begin{pmatrix} 1 & 1/\theta \\ 1/\theta & c/\theta \end{pmatrix}\right).$$

holds.

Here, $\mathcal{N}_2$ denotes the two-dimensional normal distribution. Recall that we consider only the extremal index for $\theta > 0$. Note, that we standardized the numerator with $\sqrt{p_n}$ and not with $\sqrt{r_n v_n}$. This is due to the same normalization for both processes and for Z^d_n the used normalization is natural for the calculations. But this standardization has an impact on the variance of Z^c, since $p_n/(r_n v_n) \to \theta$ under given conditions.
In the proof of this proposition, among others, one has to check condition (L) for the denominator $V^c_{n,i}$. This condition is comparable with condition (b) in Theorem 4.5 of Hsing (1991). However, this condition is implied by (θ1), so that we do not have to assume it. This is an example, where our conditions are a bit stronger compared to the conditions of asymptotic results in the literature, but therefore they are the same for all three estimators.
For the asymptotics of the estimation error of the disjoint blocks estimator we will in addition assume the bias condition (B_b).

Theorem 4.2.4. *Suppose the conditions (θ1), (θ2) and (θP) are satisfied and (B_b) holds. Then*

$$\sqrt{n v_n}(\hat{\theta}^d_n - \theta) \xrightarrow{w} \mathcal{N}(0, \theta(\theta c - 1)).$$

The resulting covariance is comparable with the covariance calculated by Hsing (1991).

For $\theta = 1$ asymptotic normality for $\hat{\theta}_n^d - \theta = \hat{\theta}_n^d - 1 \leq 0$ is not possible because of $\hat{\theta}_n^d \leq 1$. Our result includes this case $\theta = 1$, but the asymptotic variance has to be 0, i.e. $c = 1$. For instance, if $(X_t)_{t\in\mathbb{Z}}$ is regularly varying with tail process $(Y_t)_{t\in\mathbb{Z}}$, then $c = \sum_{k\in\mathbb{Z}} P(Y_k > 1)$, see (4.2.3). For $\theta = 1$ one has $Y_k = 0$ a.s. for $k \neq 0$, so in the case of regular variation it follows $c = 1$ as proposed, since $Y_0 > 1$ a.s. Thus, there is no contradiction for the case $\theta = 1$.
Next we turn to the sliding blocks counterpart $\hat{\theta}_n^s$.
For the analysis we rewrite the estimator as

$$\hat{\theta}_n^s := \frac{s_n^{-1}\sum_{i=1}^{n-s_n+1} \mathbb{1}_{\{M_{i,i+s_n-1}>u_n\}}}{\sum_{i=1}^{n-s_n+1} \mathbb{1}_{\{X_i>u_n\}}} = \frac{(\sqrt{nv_n}s_n)^{-1}\sum_{i=1}^{n-s_n+1} \mathbb{1}_{\{M_{i,i+s_n-1}>u_n\}}}{\sqrt{nv_n}^{-1}\sum_{i=1}^{n-s_n+1} \mathbb{1}_{\{X_i>u_n\}}}$$

Here, as before, the numerator and denominator are treated separately and the joint asymptotic normality is established. The behavior of the denominator has already been discussed in the previous part for the disjoint blocks estimator.
The analysis of the sliding blocks estimator can be done with Theorem 3.2.1 and with the process $\bar{Z}_n$ defined in (3.2.3). Here one has $X_{n,i} := X_i/u_n$ and we use the following bounded functions:

$$g(x_1, \dots, x_s) := \mathbb{1}_{\{\max_{1\leq i\leq s} x_i > 1\}},$$
$$h(x_1, \dots, x_s) := \mathbb{1}_{\{x_1 > 1\}}.$$

Obviously, $0 \leq g, h \leq 1$. For the sliding blocks estimator the normalization sequences are chosen as $b_n(g) = b_n(h) = (nv_n/p_n^s)^{1/2} s_n$ with $p_n^s = P(\sum_{i=1}^{r_n} \mathbb{1}_{\{M_{i,s_n+i-1}>u_n\}} > 0)$.
Again, we assume X as β-mixing and the usual restrictions on the order of the sequences s_n, l_n, r_n. However, in contrast to condition (θ1) we assume weaker β-mixing conditions but in addition another condition which restricts the extremal dependence of the time series $(X_t)_{t\in\mathbb{Z}}$.

(θ1S) $(X_t)_{t\in\mathbb{Z}}$ is a $\mathbb{R}^+$-valued stationary time series. For $v_n := P(X_1 > u_n) \to 0$, one has $nv_n \to \infty$ and $s_n \to \infty$. In addition, there exist sequences $(l_n)_{n\in\mathbb{N}}$ and $(r_n)_{n\in\mathbb{N}}$ such that $2s_n \leq l_n = o(r_n)$, $r_n v_n \to 0$, $r_n = o(\sqrt{nv_n})$ and $(n/r_n)\beta^X_{n,l_n-s_n} \to 0$.

Moreover, it is

$$\frac{1}{s_n}\sum_{l=s_n}^{r_n} P(M_{l,s_n+l} > u_n \mid X_1 > u_n) \to 0. \tag{4.2.5}$$

Lemma 4.2.5. *(i) Condition (θ1) implies (θ1S).*

(ii) (4.2.5) is implied by $(r_n/s_n)P(M_{s_n,r_n+s_n} > u_n|X_1 > u_n) \to 0$, provided ($\theta$P) holds.

The condition $(r_n/s_n)P(M_{s_n,r_n+s_n} > u_n|X_1 > u_n) \to 0$ in part (ii) of the lemma is stronger than (θP), since it imposes an additional rate on the convergence of the probability. Instead of (4.2.5) one could use this condition for (θ1S). Alternatively to condition (4.2.5)

one could also use a condition on maximal correlation coefficients (to bound II in (4.4.7), cf. the proofs) as e.g. Robert et al. (2009) uses them. However, we use (4.2.5) which is implied by ($\theta 1$) such that we have uniform conditions for all three estimators.
Under the previous conditions one can derive joint asymptotic normality of numerator and denominator of $\hat{\theta}_n^s$ by Theorem 3.2.1, part (a).

Proposition 4.2.6. *Suppose the conditions ($\theta 1S$), ($\theta 2$) and (θP) hold. Then,*

$$\begin{pmatrix} Z_n^S \\ Z_n^N \end{pmatrix} := \begin{pmatrix} (\sqrt{nv_n} s_n)^{-1} \sum_{i=1}^{n-s_n+1} \left(\mathbb{1}_{\{M_{i,i+s_n-1} > u_n\}} - P(M_{1,s_n} > u_n) \right) \\ \sqrt{nv_n}^{-1} \sum_{i=1}^{n-s_n+1} \left(\mathbb{1}_{\{X_i > u_n\}} - P(X_1 > u_n) \right) \end{pmatrix}$$
$$\xrightarrow{w} \begin{pmatrix} Z(g) \\ Z(h) \end{pmatrix} \sim \mathcal{N}_2 \left(0, \begin{pmatrix} \theta & 1 \\ 1 & c \end{pmatrix} \right).$$

With this proposition and the bias condition (B_b) we can show the asymptotics for $\hat{\theta}_n^s$ with methods as in Lemma 3.3.5.

Theorem 4.2.7. *Suppose the conditions ($\theta 1S$), ($\theta 2$) and (θP) are satisfied. In addition assume that (B_b) holds. Then*

$$\sqrt{nv_n}(\hat{\theta}_n^s - \theta) \xrightarrow{w} \mathcal{N}(0, \theta(c\theta - 1)).$$

By part (i) of Lemma 4.2.5, this theorem implies Theorem 4.2.2 for the sliding blocks estimator.

4.2.2 Extremal index - runs estimator

Next we continue with the runs estimator. The principle for the runs estimator $\hat{\theta}_n^r$ was first motivated by Leadbetter et al. (1989). The concrete estimator was given by Hsing (1993) and discussed in e.g. Smith and Weissman (1994) and Embrechts et al. (2013). A first asymptotic analysis was given by Weissman and Novak (1998). We will establish the asymptotic distribution of the estimator in the same setting as the asymptotics of the disjoint and sliding blocks estimators. We will examine the numerator and the denominator of $\hat{\theta}_n^r$ each with the sliding blocks result in Theorem 3.2.1 and with the process $\bar{Z}_n$ defined in (3.2.3). The denominator alone was already analyzed for $\hat{\theta}_n^s$.
The indicator in the numerator of $\hat{\theta}_n^r$ can be interpreted as a function of the sliding blocks $(X_i, ..., X_{i+s_n-1})$ with block length s_n. The runs estimator is a sliding blocks estimator with the special block function

$$f(x_1, ..., x_s) = \mathbb{1}_{\{x_1 > 1, \max_{2 \le i \le s} x_i \le 1\}}$$

for the numerator and for the denominator

$$h(x_1, \ldots, x_s) = \mathbb{1}_{\{x_1 > 1\}}$$

as in Section 4.2.1. In the notation of the abstract setting we consider $X_{n,i} = X_i/u_n$ such that $f(X_{n,1}, ..., X_{n,s_n}) = \mathbb{1}_{\{X_{n,1}>1, \max_{2\le j\le s_n} X_{n,j}\le 1\}}$, and the function f does not depend on n. Again, we have $m_n := \lfloor (n - s_n + 1)/r_n \rfloor$ and $p_n = P(M_{1,r_n} > u_n)$. As normalization we choose $b_n(f) = b_n(h) = (nv_n/p_n)^{1/2}$.
As before we assume that X is β-mixing, but we use a slightly weaker mixing condition than in (θ1). In addition the usual conditions on the rates of s_n, l_n and r_n are needed:

(θ1R) $(X_t)_{t\in\mathbb{Z}}$ is a $\mathbb{R}^+$-valued stationary time series. For $v_n := P(X_1 > u_n) \to 0$, one has $nv_n \to \infty$ and $s_n \to \infty$. In addition, there exists sequences $(l_n)_{n\in\mathbb{N}}$, $(r_n)_{n\in\mathbb{N}}$ such that $2s_n \le l_n = \mathrm{o}(r_n)$, $r_nv_n \to 0$, $r_n = \mathrm{o}(\sqrt{nv_n})$ and $(n/r_n)\beta^X_{n,l_n-s_n} \to 0$ and $(s_n/v_n)\beta^X_{n,s_n-1} \to 0$.

Note that s_n is still the block length, but in contrast to (θ1) we have an additional sequence $l_n \ge 2s_n$ which is used for the weaker β-mixing assumption. The β-mixing assumption $(s_n/v_n)\beta^X_{n,s_n-1} \to 0$ in (θ1R) is implied by the stronger β-mixing assumption $(n/r_n)\beta^X_{n,s_n-1} \to 0$ in (θ1). Indeed,

$$\frac{s_n}{v_n}\beta^X_{n,s_n-1} = \frac{r_ns_n}{nv_n}\frac{n}{r_n}\beta^X_{n,s_n-1} \le \frac{r_n^2}{nv_n}\frac{n}{r_n}\beta^X_{n,s_n-1} \to 0,$$

since $r_n = \mathrm{o}(\sqrt{nv_n})$. Therefore, ($\theta$1) implies ($\theta$1R).
The estimator $\hat{\theta}^r_n$ is motivated by the convergence (4.2.2) so that our condition should ensure that this convergence holds. Recall that this is equivalent to (4.2.1) which in particular holds under (θ1R) and (θP).
The result of the asymptotic normality of the numerator and denominator is given in the following proposition.

Proposition 4.2.8. *Suppose the conditions (θ1R), (θ2) and (θP) are satisfied. Then the weak convergence*

$$\begin{pmatrix} \bar{Z}_n(f) \\ \bar{Z}_n(h) \end{pmatrix} := \begin{pmatrix} (nv_n)^{-1/2}\sum_{i=1}^{n-s_n+1}(\mathbb{1}_{\{X_i>u_n, M_{2,s_n}\le u_n\}} - P(X_1 > u_n, M_{2,s_n} \le u_n)) \\ (nv_n)^{-1/2}\sum_{i=1}^{n-s_n+1}\left(\mathbb{1}_{\{X_i>u_n\}} - P(X_1 > u_n)\right) \end{pmatrix}$$
$$\xrightarrow{w} \begin{pmatrix} Z(f) \\ Z(h) \end{pmatrix} \sim \mathcal{N}_2\left(0, \begin{pmatrix} \theta & 1 \\ 1 & c \end{pmatrix}\right).$$

holds.

As stated in the next lemma, we can modify the mixing conditions of the previous proposition.

Lemma 4.2.9. *For Proposition 4.2.8 one could use the condition*

$$s_nP(M_{s_n,2s_n} > u_n \mid X_1 > u_n) \to 0 \tag{4.2.6}$$

instead of $(s_n/v_n)\beta^X_{n,s_n-1} \to 0$.

By (θP) one has $P(M_{s_n,r_n} > u_n \mid X_1 > u_n) \to 0$. Thus, (4.2.6) is an additional assumption on the rate of convergence. If $s_n^2 v_n \to 0$, then $(s_n/v_n)\beta^X_{n,s_n-1} \to 0$ implies (4.2.6), since

$$\begin{aligned}
&s_n P(M_{s_n,2s_n} > u_n \mid X_1 > u_n) = \frac{s_n}{v_n} P(M_{s_n,2s_n} > u_n, X_1 > u_n)\\
&\le \frac{s_n}{v_n}(P(M_{s_n,2s_n} > u_n)P(X_1 > u_n) + \beta^X_{n,s_n-1}) = \frac{s_n}{v_n}\mathrm{O}(s_n v_n^2) + \frac{s_n}{v_n}\beta^X_{n,s_n-1} \to 0.
\end{aligned}$$

The β-condition is a restriction for the whole distribution of the time series $(X_t)_{t\in\mathbb{Z}}$ (since we defined $X_{n,i} = X_i/u_n \mathbb{1}_{\{X_i > u_n\}}$), whereas (4.2.6) restricts only the dependence in the extreme parts of the time series. This is an advantage of the alternative condition, however, $(s_n/v_n)\beta^X_{n,s_n-1} \to 0$, without additional assumption $s_n^2 v_n \to 0$, is directly implied by the uniform condition (θ1), which is why we use this condition here. In particular, (θ1) implies (θ1R).

With the previous proposition and the bias condition (B_r) we can establish the asymptotic normality of $\hat\theta^r_n$.

Theorem 4.2.10. *Suppose the conditions (θ1R), (θ2) and (θP) are satisfied and (B_r) holds. Then*

$$\sqrt{n v_n}(\hat\theta^r_n - \theta) \xrightarrow{w} \mathcal{N}(0, \theta(c\theta - 1)).$$

Without additional assumptions there is no way to express the emerging constants c more explicitly. This expression is the same as for the disjoint blocks estimator and the sliding blocks estimator and has another form than the asymptotic variance in Weissman and Novak (1998), Theorem 3.

Concluding remarks on the three estimators for the extremal index

The non-trivial limit of the standardized estimator $\hat\theta^r_n$ shows that the normalization with $(n v_n)^{-1}$ of the numerator was chosen correctly. This is an example where a sliding blocks statistic (in the special case of runs) can have the same normalization as disjoint blocks statistic (which is not the direct counterpart). $\hat\theta^s_n$ requires the normalization $(n v_n s_n)^{-1}$ for a non-trivial limit (but the convergence rate still is $\sqrt{n v_n}$). This example shows that sliding blocks do not always require the same standardization and that the general normalization $b_n(g)$ in Section 3.2 is an important feature.

Theorem 4.2.2 proves that $\hat\theta^d_n$, $\hat\theta^s_n$ and $\hat\theta^r_n$ are equally efficient in terms of asymptotic variances. However, for the asymptotic result we used some bias condition. Without such bias condition Smith and Weissman (1994) showed under some special conditions, that in terms of the asymptotic bias $\hat\theta^r_n$ should be clearly preferred. Since $\hat\theta^d_n$ and $\hat\theta^s_n$ posses the same expected value, both estimators have the same bias. Thus, following the arguments of Smith and Weissman (1994) one would prefer the runs estimator.

Our result for the asymptotic variance does not confirm the hypothesis stated by Beirlant et al. (2004) that the sliding blocks estimator is more efficient. We can only confirm that

$\hat{\theta}_n^s$ is not worse than $\hat{\theta}_n^d$ in terms of the asymptotic variance. The result of our asymptotic variances fits to the result in Section 3.3. In fact we even have equality of the asymptotic variances of the disjoint and corresponding sliding blocks estimator.
Since all three estimators have the same asymptotic variance, one question is, whether a convex combination of two of these estimators $\lambda\hat{\theta}_n^{\sharp_1}+(1-\lambda)\hat{\theta}_n^{\sharp_2}$, with $\lambda \in (0,1)$ and $\sharp_1, \sharp_2 \in \{s, d, n\}$, could lead to a mixed estimator with smaller asymptotic variance. This would be possible, if the correlation between this estimators is smaller than 1, or here equivalently if the asymptotic covariance between the estimators is smaller than the asymptotic variance of one estimator. The asymptotic covariance for the standardized estimation error of the sliding blocks estimator and the runs estimator is given by

$$\begin{aligned}&Cov(Z(g)-\theta Z(h), Z(f)-\theta Z(h))\\&= Cov(Z(g), Z(h)) + \theta^2 c - \theta - \theta = \theta\Big(\theta c - 2 + \frac{Cov(Z(g), Z(h))}{\theta}\Big).\end{aligned}$$

Here we used the known covariance from Proposition 4.2.6 and 4.2.8. Hence, the covariance is smaller than the variance $\theta(\theta c - 1)$ if and only if $Cov(Z(g), Z(h)) < \theta$. This asymptotic covariance can be calculated like the asymptotic variances as

$$\lim_{n\to\infty} \frac{1}{r_n v_n s_n} Cov\Big(\sum_{i=1}^{r_n} \mathbb{1}_{\{X_i>u_n, M_{i+1,i+s_n-1}\le u_n\}}, \sum_{i=1}^{r_n} \mathbb{1}_{\{M_{i,i+s_n-1}>u_n\}}\Big) = \theta.$$

This limit can be calculated with exactly the same arguments as for the covariances of the runs and sliding blocks estimators in Lemma 4.4.1 and Lemma 4.4.2. Thus, it is $Cov(Z(g), Z(h)) = \theta$ and, therefore, the correlation between $\hat{\theta}_n^r$ and $\hat{\theta}_n^s$ equals 1.
With similar arguments one could also calculate the covariance between $\hat{\theta}_n^d$ and $\hat{\theta}_n^s$ or $\hat{\theta}_n^r$ and always achieve the covariance $\theta(\theta c - 1)$. Thus, the correlation between all three estimators is 1. Therefore, a convex combination of these estimators would not reduce the asymptotic variance.
Intuitively, this is also the expected result, since all three estimators do essentially the same thing: they count the number of clusters and divide it by the number of extreme observations. Although the way the estimators detect a cluster is defined differently for the runs estimator, the clusters are asymptotically the same for all three estimators (see above the equivalence of (4.2.1) and (4.2.2)). Therefore, the asymptotic correlation should intuitively be 1, as it is.

4.2.3 Extremal index - sliding blocks estimator with random threshold

In the previous sections, three extremal index estimators were considered, all based on observations exceeding a deterministic threshold u_n. In practice, when estimating extreme value parameters this threshold u_n is often replaced by a random threshold $\hat{u}_n$, which

depends on the observed time series, e.g. as order statistics. This practice does not fit to the mathematical limit results that were derived. Therefore, it is interesting to consider modified estimators where u_n is replaced by an estimator $\hat{u}_n$. For the asymptotic analysis of the estimator with random thresholds we can apply the theory of Chapter 3, but we need to apply the more complex theory about process convergence. Thus, the proofs of this section, which use methods already introduced by Drees and Knezevic (2020), are an example for the application of the process convergence.
The disjoint blocks estimator $\hat{\theta}_n^d$ was already introduced and considered by Hsing (1991) as an estimator with order statistics as thresholds. In the paper, however, no asymptotic behavior was considered due to the technical challenge and complex empirical process theory required. In this section, we consider exemplary the sliding blocks estimator $\hat{\theta}_n^s$ with a random threshold. For $\hat{\theta}_n^d$ and $\hat{\theta}_n^r$ one could derive analog results following the same arguments as described here for the sliding blocks estimator.
The modified sliding blocks estimator with random thresholds is defined as

$$\hat{\theta}_{n,\hat{u}_n}^s := \frac{(\sqrt{nv_n}s_n)^{-1}\sum_{i=1}^{n-s_n+1}\mathbb{1}_{\{M_{i,i+s_n-1}>\hat{u}_n\}}}{(\sqrt{nv_n})^{-1}\sum_{i=1}^{n-s_n+1}\mathbb{1}_{\{X_i>\hat{u}_n\}}}.$$

We will consider a random threshold $\hat{u}_n$ which is a consistent estimator for u_n in the sense that

$$D_n := \frac{\hat{u}_n}{u_n} \xrightarrow{P} 1.$$

Under some conditions, Drees and Knezevic (2020), Lemma 2.2, showed, that the k_n-th largest order statistics $X_{n-k_n+1:n}$ of the sample $(X_1, ..., X_n)$ satisfies $X_{n-k_n+1:n}/u_n \to 1$ in probability if $k_n = \lceil nv_n \rceil$. Thus, such order statistics fulfill our consistency condition and one can replace $\hat{u}_n$ by the k_n-th largest order statistics.
The asymptotic analysis of this estimator is more or less the same as in Theorem 4.2.7, i.e. we will consider numerator and denominator separately, represent them as empirical processes and derive their asymptotics as the asymptotics of a sliding blocks statistic. The basic idea of the asymptotic analysis of $\hat{\theta}_{n,\hat{u}_n}^s$ is to amend the empirical process $(\bar{Z}_n(g), \bar{Z}_n(h))$ used in the proof of Theorem 4.2.7 by an additional parameter $d \in [1-\varepsilon, 1+\varepsilon]$ (for some $\varepsilon > 0$) that later on is replaced with D_n. This extension makes the use of process convergence necessary, the fidis convergence is no longer sufficient. The parameter d is multiplied by the deterministic threshold u_n. By inserting D_n for d the estimator with random thresholds is obtained.
This new parameter d requires some strengthened conditions. The condition (θ1) will be used as before (note that we use (θ1) instead of (θ1S)). The other conditions are similar to the conditions for the deterministic threshold, just the threshold u_n is replaced by $(1-\varepsilon)u_n$ for some $\varepsilon > 0$.

(θPR) There exist $\varepsilon > 0$ and, for all $n \in \mathbb{N}$ and $k \in \mathbb{N}$ there exist $e_n(k)$ such that

$$e_n(k) \geq P(X_k > (1-\varepsilon)u_n \mid X_0 > (1-\varepsilon)u_n)$$

and $\lim_{n\to\infty} \sum_{k=1}^{r_n} e_n(k) = \sum_{k=1}^{\infty} \lim_{n\to\infty} e_n(k) < \infty$.

($\mathbf{B_b R}$) For all sequences $d_n \to 1$,

$$\sqrt{nv_n}\left(\frac{P(M_{1,s_n} > d_n u_n)}{s_n P(X_0 > d_n u_n)} - \theta\right) \to 0.$$

In addition, we assume that the positive part $(X_t^+)_{t\in\mathbb{Z}} := (X_t \mathbb{1}_{\{X_t \geq 0\}})_{t\in\mathbb{Z}}$ of the time series is regularly varying with index α. In particular, this implies the existence of the tail process $(Y_t)_{t\in\mathbb{Z}}$ for $(X_t^+)_{t\in\mathbb{Z}}$ (Theorem 2.1.7). This assumption greatly simplifies the calculations for some conditions, in particular for the covariances and the condition (D1). We thus assume

(R) $(X_t^+)_{t\in\mathbb{Z}}$ is regularly varying with tail process $(Y_t)_{t\in\mathbb{Z}}$, spectral process $(\Theta_t)_{t\in\mathbb{Z}}$ and index α.

Observe that if (θPR) and (R) are satisfied then the following generalization of (θPR) holds as well: one has for all $c, d \in [1-\varepsilon, 1+\varepsilon]$

$$\begin{aligned}
&P(X_k > cu_n \mid X_0 > du_n) = P(X_k > cu_n, X_0 > du_n)\frac{1}{P(X_0 > du_n)} \\
&\leq P(X_k > (1-\varepsilon)u_n, X_0 > (1-\varepsilon)u_n) \cdot \frac{P(X_0 > (1-\varepsilon)u_n)}{P(X_0 > (1-\varepsilon)u_n)P(X_0 > (1+\varepsilon)u_n)} \\
&\leq P(X_k > (1-\varepsilon)u_n \mid X_0 > (1-\varepsilon)u_n) \cdot \frac{P(X_0 > (1-\varepsilon)u_n)}{P(X_0 > (1+\varepsilon)u_n)} \\
&\leq 2\left(\frac{1-\varepsilon}{1+\varepsilon}\right)^{-\alpha} e_n(k) =: \tilde{e}_n(k) \qquad (4.2.7)
\end{aligned}$$

with $\lim_{n\to\infty} \sum_{k=1}^{r_n} \tilde{e}_n(k) = \sum_{k=1}^{\infty} \lim_{n\to\infty} \tilde{e}_n(k) < \infty$. This holds for n large enough, since $P(X_0 > (1-\varepsilon)u_n)/P(X_0 > (1+\varepsilon)u_n) \leq 2((1-\varepsilon)/(1+\varepsilon))^{-\alpha}$ for n large enough, due to the Potter bounds.

Due to (θP) and (R) the constant c from (θ2) can be expressed with the tail process as $c = \sum_{k\in\mathbb{Z}} P(Y_k > 1)$, see also (4.2.3) (under regular variation one can also replace r_n there by s_n.).

To apply Theorem 3.2.1 we consider the sliding blocks functionals g_d with

$$g_d(x_1, ..., x_s) = \mathbb{1}_{\{\max_{1\leq i\leq s} x_i > d\}}$$
$$h_d(x_1, ..., x_s) = \mathbb{1}_{\{x_1 > d\}}$$

for $d \in [1-\varepsilon, 1+\varepsilon]$ for some $\varepsilon > 0$, where g_d is used for the numerator and h_d for

the denominator of $\hat{\theta}^s_{n,\hat{u}_n}$. Thus, the functional class which we analyze in the following is $\mathcal{G} = \{g_d, h_d : d \in [1-\varepsilon, 1+\varepsilon]\}$.
Define

$$Z_n(g_d) := \frac{1}{\sqrt{nv_n}s_n} \sum_{i=1}^{n-s_n+1} (g_d(X_{n,i}, ..., X_{n,i+s_n-1}) - E[g_d(X_{n,i}, ..., X_{n,i+s_n-1})]),$$
$$Z_n(h_d) := \frac{1}{\sqrt{nv_n}} \sum_{i=1}^{n-s_n+1} (h_d(X_{n,i}, ..., X_{n,i+s_n-1}) - E[h_d(X_{n,i}, ..., X_{ni+s_n-1})]).$$

Recall that $X_{n,i} = X_i/u_n$ and we use the notation $M_{i,j} = \max(X_i, ..., X_j)$ for $-\infty < i \leq j < \infty$. As normalization we choose $b_n(g_d) = (nv_n/p^s_n)^{1/2}s_n$ and $b_n(h_d) = (nv_n/p^s_n)^{1/2}$ with $p^s_n = P(M_{1,r_n+s_n-1} > (1-\varepsilon)u_n)$.
Similar to the deterministic thresholds in Section 4.2.1, we prove the asymptotic normality of $(Z_n(f))_{f\in\mathcal{G}}$ in a first step.

Proposition 4.2.11. *Suppose the conditions (θ1), (θPR) and (R) are satisfied. Then the weak convergence*

$$(Z_n(f))_{f\in\mathcal{G}} \xrightarrow{w} (Z(f))_{f\in\mathcal{G}}$$

holds, where Z *is a centered Gaussian process with existing covariance function. In particular,* $Var(Z(g_1)) = \theta$, $Var(Z(h_1)) = c$ *and* $Cov(Z(g_1), Z(h_1)) = 1$ *with* $c = \sum_{k\in\mathbb{Z}} P(Y_k > 1)$.

The sliding blocks estimator $\hat{\theta}^s_{n,\hat{u}_n}$ is based on the exceedances over the random threshold $\hat{u}_n = D_n u_n$. The following result shows that $\hat{\theta}^s_{n,\hat{u}_n}$ has the same limit distribution as the estimators with deterministic thresholds.

Theorem 4.2.12. *Suppose the conditions (θ1), (θPR), (B_bR), and (R) are satisfied and* $D_n \to 1$ *in probability. Then*

$$\sqrt{nv_n}(\hat{\theta}^s_{n,\hat{u}_n} - \theta) \xrightarrow{w} \mathcal{N}(0, \theta(c\theta - 1)).$$

At this point we end the consideration of estimators for the extremal index. In the next section we consider another cluster index, the so-called stop-loss index. Again the asymptotic behavior will be established by an application of the theory from Chapter 3.

4.3 Stop-loss index

As second example for a cluster index we consider the stop-loss index. This index is defined as cluster index with the cluster functional $H_S(y) := \mathbb{1}_{\{\sum_{j\in\mathbb{Z}}(y_j-1)^+ > S\}}$ for some $S > 0$. For a non-negative, stationary and regularly varying time series $(X_t)_{t\in\mathbb{Z}}$ the index is given by

$$\theta_{sl}(S) := \lim_{n\to\infty} \frac{P\left(\sum_{j=1}^{s_n}(X_j - u_n)^+ > Su_n\right)}{s_n P(X_0 > u_n)} = P\bigg(\sum_{j=0}^{\infty}(Y_j - 1)^+ > S, Y^*_{-\infty,-1} \le 1\bigg). \tag{4.3.1}$$

The convergence holds under the condition of Proposition 4.1.2 and by (4.1.2). (Note that summation over $j \in \mathbb{N}$ and $j \in \mathbb{Z}$ in the last probability is the same due to the second argument.)
The stop-loss index can be interpreted as the probability that the total extreme losses are at least S, given that today the first extreme loss occurred. This parameter could be used e.g. for risk management to control total losses from extreme risks. By such an application not necessarily only one stop loss index is relevant, but the entire stop loss function $(\theta_{sl}(S))_{S>S_0}$, for some $S_0 > 0$. We will call $(\theta_{sl}(S))_{S>S_0}$ the stop-loss distribution. In the following we will estimate this function point-wise by the runs estimator and the disjoint and sliding blocks estimators. For the analysis we will use the theory from Chapter 3. Recently, this index was also considered as example in Cissokho and Kulik (2021) under different conditions and without the runs estimator.

4.3.1 Stop-loss index - runs estimator

Here we start with the runs estimator. The (pseudo)-runs estimator motivated above (more precisely the second version) for the stop-loss index is given by

$$\begin{aligned}
\tilde{\theta}^r_{sl,n}(S) &= \frac{1}{(n-2s_n)v_n} \sum_{j=s_n+1}^{n-s_n} \mathbb{1}_{\{\sum_{i=j-s_n}^{j+s_n}(X_i-u_n)^+ > Su_n\}} \mathbb{1}_{\{X_j > u_n\}} \mathbb{1}_{\{M_{j-s_n,j-1} \le u_n\}} \\
&= \frac{1}{(n-2s_n)v_n} \sum_{j=s_n+1}^{n-s_n} \mathbb{1}_{\{\sum_{i=j}^{j+s_n}(X_i-u_n)^+ > Su_n\}} \mathbb{1}_{\{X_j > u_n\}} \mathbb{1}_{\{M_{j-s_n,j-1} \le u_n\}}.
\end{aligned}$$

Note that an analogous analysis to the one that now follows can also be performed for the first version of the above runs estimator for cluster indexes.
Check that H_S is bounded, shift-invariant and $H_S(y) = 0$ if $\|y\|_\infty < 1$, which is why the motivation of the general runs estimator can be used for this specific cluster index. By stationarity and Proposition 4.1.3, the expected value of this estimator is

$$\begin{aligned}
E[\tilde{\theta}^r_{sl,n}(S)] &= P\bigg(\sum_{i=0}^{s_n}(X_i - u_n)^+ > Su_n, M_{-s_n,-1} \le u_n \,\bigg|\, X_0 > u_n\bigg) \\
&\to P\bigg(\sum_{j=0}^{\infty}(Y_j - 1)^+ > S, Y^*_{-\infty,-1} \le 1\bigg),
\end{aligned}$$

i.e. the estimator is asymptotically unbiased.
Note that this pseudo estimator still depends on $v_n = P(X_0 > u_n)$, which is in general

unknown. One has to replace v_n e.g. by the empirical version $\hat{v}_n = \sum_{i=s_n+1}^{n-s_n} \mathbb{1}_{\{X_i>u_n\}}/n$. The estimator then has basically the same denominator as the estimators for the extremal index: $\hat{\theta}_n^d$, $\hat{\theta}_n^r$ and $\hat{\theta}_n^s$. The full runs estimator for the stop loss index is

$$\hat{\theta}_{sl,n}^r(S) = \frac{\sum_{j=s_n+1}^{n-s_n} \mathbb{1}_{\{\sum_{i=j}^{j+s_n}(X_i-u_n)^+>Su_n\}} \mathbb{1}_{\{X_j>u_n\}} \mathbb{1}_{\{M_{j-s_n,j-1}\le u_n\}}}{\sum_{j=s_n+1}^{n-s_n} \mathbb{1}_{\{X_j>u_n\}}}.$$

Here the run $\mathbb{1}_{\{M_{-s_n,-1}\le u_n, X_0>u_n\}}$ occurs before the required extreme observation occurs. The minimum number of s_n non-extreme observations in a run serve to separate clusters of extreme events and to use only one observation from each cluster. For the runs estimator $\hat{\theta}_n^r$ for the extremal index the last s_n observations of a cluster of extremes is used for the statistic, while in contrast the estimator considered here uses the first s_n observations of a cluster of extremes. Nevertheless, the treatment in the sliding blocks setting is completely analogous.
For the analysis of $(\hat{\theta}_{sl,n}^r(S))_{S>0}$ we will use the theory of Section 3.2. To this end, we consider the functions

$$g_S(x_{-s},...,x_s) = \mathbb{1}_{\{\sum_{i=0}^{s}(x_i-1)^+>S\}} \mathbb{1}_{\{x_0>1\}} \mathbb{1}_{\{\max_{-s\le i\le -1} x_i\le 1\}}$$

for some $S>0$. For the denominator of the estimator we consider the function

$$h(x_{-s},...,x_s) = \mathbb{1}_{\{x_0>1\}}.$$

Note that this is not exactly the setting of Section 3.2, since here we consider $g(x_{-s},...,x_s)$ instead of $g(x_1,...,x_s)$, i.e. we use shifted blocks here. This is mainly due to a convenient representation here. Formally, we have to use the observations $X'_{n,t} = X_{n,t-s_n}$ and $s'_n = 2s_n+1$ for the framework of Section 3.2 (see also the beginning of the proof of Proposition 5.2.5). We define $X_{n,i} = X_i/u_n$ as before. As normalization we choose $b_n(g_S) = b_n(h) = (nv_n/p_n)^{1/2}$ with $p_n = P(M_{1,r_n} > u_n)$.
The following conditions are needed, such that the asymptotic normality of the estimator $(\hat{\theta}_{sl,n}^r(S))_{S>0}$ can be established by Theorem 3.2.1.

(S1) Let $(X_t)_{t\in\mathbb{Z}}$ be a $\mathbb{R}^+$-valued stationary, regularly varying time series with index α. For $v_n := P(X_0 > u_n) \to 0$, one has $nv_n \to \infty$ and $s_n \to \infty$. In addition, there exists a sequence $(r_n)_{n\in\mathbb{N}}$ such that $s_n = \mathrm{o}(r_n)$, $r_n v_n \to 0$, $r_n = \mathrm{o}(\sqrt{nv_n})$ and $(n/r_n)\beta_{n,s_n-1}^X \to 0$.

(SP) For all $n\in\mathbb{N}$ and $k\in\mathbb{N}$ there exists $e_n(k)$ such that

$$e_n(k) \ge P(X_k > u_n \mid X_1 > u_n)$$

and $\lim_{n\to\infty}\sum_{k=1}^{r_n} e_n(k) = \sum_{k=1}^{\infty}\lim_{n\to\infty} e_n(k) < \infty$.

(SB$_r$) There exists an $S_0 > 0$ such that

$$\lim_{n\to\infty} \sup_{S\in[S_0,\infty)} \sqrt{nv_n} \left| \frac{1}{v_n} E\left[g_S(X_{-s_n}, ..., X_{s_n})\right] - \theta_{sl}(S) \right| = 0.$$

The first condition is the usual condition on the rates of the sequences s_n, l_n, r_n, u_n and on the dependence structure of $(X_t)_{t\in\mathbb{Z}}$. Condition (SP) is the same Pratt's condition as (θP) in Section 4.2 which restricts the extremal dependence structure, we restated it here for completeness. Condition (SB$_r$) is a bias condition. For the expectation one has

$$\begin{aligned}
\frac{1}{v_n} E[g_S(X_{-s_n}, ..., X_{s_n})] &= \frac{1}{v_n} P\Big(\sum_{i=0}^{s_n} (X_i - u_n)^+ > Su_n, X_0 > u_n, \max_{-s_n\le i\le -1} X_i \le u_n \Big) \\
&= P\Big(\sum_{i=0}^{s_n} (X_i - u_n)^+ > Su_n, \max_{-s_n\le i\le -1} X_i \le u_n \,\Big|\, X_0 > u_n \Big) \\
&\to P\Big(\sum_{j=0}^{\infty} (Y_j - 1)^+ > S, Y^*_{-\infty,1} \le 1 \Big)
\end{aligned}$$

by Proposition 4.1.3 if condition (AC) is satisfied. Thus, this condition (SB$_r$) is only a restriction for the rate of the convergence, or to be more precise a restriction for u_n, since nv_n is not allowed to increase too fast. More precisely, this condition ensures that the bias converges to 0 faster than the stochastic error.
With these conditions we can state the following asymptotic statement for the runs estimator of the stop loss index. Recall $c = \sum_{k\in\mathbb{Z}} P(Y_k > 1)$ from (4.2.3), which will be again the asymptotic variance of $\hat{v}_n$.

Theorem 4.3.1. *Suppose the conditions (S1), (SP), (SB$_r$) and (AC) are satisfied. Then the weak convergence*

$$\sqrt{nv_n} \left(\hat{\theta}^r_{sl,n}(S) - \theta_{sl}(S) \right)_{S\in[S_0,\infty)} \xrightarrow{w} (Z_S)_{S\in[S_0,\infty)}$$

holds, where Z_S is a centered Gaussian process with covariance

$$\begin{aligned}
Cov(Z_s, Z_t) =& \theta_{sl}(\max(s,t)) + \theta_{sl}(s)\theta_{sl}(t)c \\
&- \theta_{sl}(s) \sum_{k=0}^{\infty} P\Big(\sum_{j=0}^{\infty} (Y_j - 1)^+ > t, Y^*_{-\infty,-1} \le 1, Y_k > 1 \Big) \\
&- \theta_{sl}(t) \sum_{k=0}^{\infty} P\Big(\sum_{j=0}^{\infty} (Y_j - 1)^+ > s, Y^*_{-\infty,-1} \le 1, Y_k > 1 \Big),
\end{aligned}$$

with c defined in (4.2.3).

Instead of considering $[S_0, \infty)$ one could also consider (S_0, ∞). The statement remains the same, only the index set changes.

4.3.2 Stop-loss index - blocks estimators

After the runs estimator, we now want to consider the disjoint blocks estimator and the sliding blocks counterpart, starting with the disjoint blocks version. The disjoint blocks estimator for this specific cluster index is given by

$$\hat{\theta}^d_{sl,n}(S) = \frac{\sum_{j=1}^{\lfloor n/s_n \rfloor} \mathbb{1}_{\{\sum_{i=(j-1)s_n+1}^{js_n}(X_i-u_n)^+ > Su_n\}}}{\sum_{j=1}^{n-s_n} \mathbb{1}_{\{X_j>u_n\}}}.$$

This estimator is directly motivated as empirical counterpart of the limit in (4.3.1) and it is defined as the general cluster index estimator in (4.1.3) for the special function H_S. The corresponding sliding blocks estimator is given by

$$\hat{\theta}^s_{sl,n}(S) = \frac{s_n^{-1}\sum_{j=1}^{n-s_n+1} \mathbb{1}_{\{\sum_{i=j}^{j+s_n-1}(X_i-u_n)^+ > Su_n\}}}{\sum_{j=1}^{n-s_n+1} \mathbb{1}_{\{X_j>u_n\}}}.$$

Note that the role of s_n here slightly differs from the role of s_n for the runs estimator. Here s_n is the length of each considered block, while for the runs estimator the effective block length was $2s_n+1$. In particular, we are directly in the setting of Chapter 3, without redefinition of $X_{n,i} = X_i/u_n$ and the block lengths.
This disjoint blocks estimator cannot be treated in the setting of Section 3.2 but directly in the setting of Section 3.1.1. For the application of Theorem 3.1.10 one possible choice for $V_{n,i}$ would be $V^d_{n,j}(S) = m_n^{-1/2}\mathbb{1}_{\{\sum_{i=(j-1)s_n+1}^{js_n}(X_i-u_n)^+>Su_n\}}$ with $m_n = \lfloor n/s_n \rfloor$ and $p_n = P(M_{1,s_n} > u_n)$. However, similar as for the disjoint blocks estimator of the extremal index in Section 4.2.1, we consider

$$V^d_{n,j}(S) = \frac{1}{\sqrt{m_n}} \sum_{k=1}^{r_n/s_n} \mathbb{1}_{\{\sum_{i=(j-1)r_n+(k-1)s_n+1}^{(j-1)r_n+ks_n}(X_i-u_n)^+>Su_n\}} \tag{4.3.2}$$

$$\tilde{V}^d_{n,j}(S) = \frac{1}{\sqrt{m_n}} \sum_{k=1}^{r_n/s_n-1} \mathbb{1}_{\{\sum_{i=(j-1)r_n+(k-1)s_n+1}^{(j-1)r_n+ks_n}(X_i-u_n)^+>Su_n\}}$$

for $j \in \{1, \dots, m_n\}$, $m_n = \lfloor (n-s_n+1)/r_n \rfloor$ and $V^c_{n,i}$ and $\tilde{V}^c_{n,i}$ defined in (4.2.4) for the disjoint blocks estimator of θ. In this case one has $p_n = P(M_{1,r_n} > u_n)$. This choice will increase the comparability with the result of the sliding blocks estimator.
Due to the choice of $V^d_{n,i}(S)$, the same set of conditions as for the runs estimator is sufficient to prove the asymptotic normality of $\hat{\theta}^d_{sl,n}(S)$. Just the bias condition has to be replaced by the following bias condition:

(SB$_b$) There exists an $S_0 > 0$ such that

$$\lim_{n\to\infty} \sup_{S\in[S_0,\infty)} \sqrt{nv_n}\left|\frac{1}{s_nv_n}P\Big(\sum_{i=1}^{s_n}(X_i-u_n)^+ > Su_n\Big) - \theta_{sl}(S)\right| = 0.$$

By Proposition 4.1.2 the convergence

$$\frac{1}{s_n v_n} P\Big(\sum_{i=1}^{s_n} (X_i - u_n)^+ > S u_n \Big) - \theta_{sl}(S) \to 0$$

holds. Note that we replaced r_n by s_n in Proposition 4.1.2. For s_n from (S1) the conditions of the proposition are met, in particular (AC) holds also for $s_n \le r_n$. Thus, the condition (SB$_b$) is a condition on the rate and uniformity of the convergence. With this condition we can establish asymptotic normality of $\hat{\theta}^d_{sl,n}(S)$.

Theorem 4.3.2. *Suppose the conditions (S1), (SP), (SB$_b$) and (AC) holds. Then the weak convergence*

$$\sqrt{n v_n} \left(\hat{\theta}^d_{sl,n}(S) - \theta_{sl}(S) \right)_{S \in [S_0, \infty)} \xrightarrow{w} (Z_S)_{S \in [S_0, \infty)}$$

holds, where $(Z_S)_{S \in [S_0,\infty)}$ is a centered Gaussian process with covariance function

$$\begin{aligned} Cov(Z_s, Z_t) =& \theta_{sl}(\max(s,t)) + \theta_{sl}(s)\theta_{sl}(t)c \\ & - \theta_{sl}(s) P\Big(\sum_{j \in \mathbb{Z}} (Y_j - 1)^+ > t \Big) - \theta_{sl}(t) P\Big(\sum_{j \in \mathbb{Z}} (Y_j - 1)^+ > s \Big). \end{aligned}$$

Finally, the last estimator for a cluster index we will analyze is the sliding blocks estimator $\hat{\theta}^s_{sl,n}(S)$ for the stop loss index. Analog to the runs estimator we will deal with this estimator in the setting of Section 3.2. For this application we still have $X_{n,i} = X_i/u_n$ and we consider the functions $h(x_1, ..., x_m) = \mathbb{1}_{\{x_1 > 1\}}$ and $f_S(x_1, ..., x_m) := \mathbb{1}_{\{\sum_{i=1}^m (x_j - 1)^+ > S\}}$. Obviously it is $0 \le h, f_S \le 1$ for all $S \ge S_0$. We consider the normalization $b_n(f_S) = (n v_n / p^s_n)^{1/2} s_n$ for all $S \ge S_0$ and $b_n(h) = (n v_n / p^s_n)^{1/2}$ with $p^s_n = P(M_{1,r_n} > u_n \text{ or } M_{r_n+1, r_n+s_n-1} > (1+S_0) u_n) = r_n v_n \theta (1 + o(1))$ by (4.2.1).
Under the same conditions as for the disjoint blocks estimator, the uniform asymptotic normality of the stop-loss index estimator for all $S \in [S_0, \infty)$ is shown in the next theorem.

Theorem 4.3.3. *Suppose that the conditions (S1), (SP), (SB$_b$) and (AC) hold. Then the weak convergence*

$$\sqrt{n v_n} \left(\hat{\theta}^s_{sl,n}(S) - \theta_{sl}(S) \right)_{S \in [S_0, \infty)} \xrightarrow{w} (Z_S)_{S \in [S_0, \infty)}$$

holds, where $(Z_S)_{S \in [S_0,\infty)}$ is a centered Gaussian process with covariance function

$$\begin{aligned} Cov(Z_s, Z_t) =& \theta_{sl}(\max(s,t)) + \theta_{sl}(s)\theta_{sl}(t)c \\ & - \theta_{sl}(s) P\Big(\sum_{j \in \mathbb{Z}} (Y_j - 1)^+ > t \Big) - \theta_{sl}(t) P\Big(\sum_{j \in \mathbb{Z}} (Y_j - 1)^+ > s \Big). \end{aligned}$$

The asymptotic variance for the disjoint and sliding blocks estimators are the same. As for the extremal index, asymptotically both estimators are equally efficient, in accordance

with the theory of Section 3.3. In general it is not obvious which variance is smaller, that of the runs estimator $\hat{\theta}^r_{sl,n}(S)$ or that of the blocks estimators. However, some calculations, using Problem 5.29 from Kulik and Soulier (2020), show that the asymptotic variance of the runs estimator $\hat{\theta}^r_{sl,n}(S)$ and that of the block estimators are the same. (This was shown by Anja Janßen.) Thus, again all three estimators have the same limiting covariances.
For the sliding and disjoint blocks estimators for some fixed S (and not for the whole process) the asymptotic normality could also be shown under some different conditions as stated in Cissokho and Kulik (2021). The resulting variance calculated there coincides with the variance obtained here. Indeed, they also show for a larger class, that the sliding blocks and disjoint blocks estimator for cluster indexes have the same asymptotic variance.
This brings us to the end of our consideration of the cluster indexes and the analysis of sliding blocks and runs estimators for them. In the next Chapter 5 we will turn to the next big topic of this thesis: the estimation of the whole extreme dependence structure of a time series in form of the distribution of the spectral tail process. There again we will see examples for the application of the limit theory from Chapter 3.

4.4 Proofs

In this section, all proofs for theorems, lemmas and propositions in this chapter are given.

4.4.1 Proofs for Section 4.2.1

We start with the proofs for the disjoint blocks estimator $\hat{\theta}^d_n$. Recall, that we have $X_{n,i} = X_i/u_n$ and that under the conditions (θ1) and (θP) equation (4.2.1) holds for all $k_n \to \infty$, $k_n \le r_n$. This in particular yields

$$p_n = r_n v_n(\theta + \mathrm{o}(1)), \qquad P(M_{1,s_n} > u_n) = s_n v_n(\theta + \mathrm{o}(1)). \tag{4.4.1}$$

Proof of Proposition 4.2.3. We will apply the abstract theory from Theorem 3.1.4. The condition (A1) is directly given by (θ1). By the definition of $V^d_{n,i}$, $\tilde{V}^d_{n,i}$, $V^c_{n,i}$ and $\tilde{V}^c_{n,i}$ the condition (V) is directly implied by the stationarity of $(X_t)_{t\in\mathbb{Z}}$. Condition (D0) is obvious since we consider only finitely many functions. (M$\tilde{V}$) and (MX$_2$) follow readily from the β-mixing assumption in (θ1). The latter conditions follow since $r_n - s_n > s_n - 2$ for sufficiently large n implies $\beta^X_{n,r_n-s_n} \le \beta^X_{n,s_n-1}$.
Thus, it suffices to verify the conditions (3.1.4) (or (Δ)), (L) and (C), in order to conclude the assertion from Theorem 3.1.4. Note that (3.1.4) and (L) can be checked separately for $V^d_{n,i}$ and $V^c_{n,i}$. We start with $V^d_{n,i}$. Check that

$$\Delta^d_n := V^d_{n,1} - \tilde{V}^d_{n,1} = \frac{1}{\sqrt{m_n}} \mathbb{1}_{\{M_{r_n-s_n+1,r_n} > u_n\}} \overset{d}{=} \frac{1}{\sqrt{m_n}} \mathbb{1}_{\{M_{1,s_n} > u_n\}}.$$

Now (4.2.1) and $s_n = \mathrm{o}(r_n)$ imply

$$\frac{m_n}{p_n} E[(\Delta_n^d)^2] = \frac{1}{p_n} E\left[\mathbb{1}_{\{M_{1,s_n} > u_n\}}\right] = \frac{P(M_{1,s_n} > u_n)}{P(M_{1,r_n} > u_n)}$$
$$= \frac{P(M_{1,s_n} > u_n)}{s_n P(X_1 > u_n)} \frac{r_n P(X_1 > u_n)}{P(M_{1,r_n} > u_n)} \frac{s_n}{r_n} \to \theta \cdot \frac{1}{\theta} \cdot 0 = 0.$$

Here, the convergence of the first two terms follows from (4.2.1) for $k_n = s_n$ and $k_n = r_n$. Thus, (3.1.4) holds for $V_{n,i}^d$.
Condition (L) for $V_{n,i}^d$ follows immediately from

$$V_{n,i}^d \leq m_n^{-1/2} r_n / s_n = \mathrm{O}\Big(\frac{r_n}{s_n \sqrt{n v_n}} \sqrt{r_n v_n}\Big) = \mathrm{o}(\sqrt{p_n}),$$

because of (4.4.1) and (θ1). Thus, $V_n^d \leq \varepsilon \sqrt{p_n}$ for sufficiently large n. This implies that the indicator in the expectation in (L) equals 0 for n large enough, i.e the left hand side is 0 for n large enough and in particular condition (L) is satisfied.
Now we check the condition (3.1.4) and (L) for the denominator $V_{n,1}^c$. Since $V_{n,1}^c$ is a sliding blocks statistic with $X_{n,i} = X_i / u_n$, bounded function $0 \leq h(x_1, \ldots, x_s) = \mathbb{1}_{\{x_1 > 1\}} \leq 1$ and $b_n(h) = \sqrt{m_n}$, the proof of Theorem 3.2.1 shows that (3.1.4) and (L) hold if $r_n = \mathrm{o}(\sqrt{p_n} b_n) = \mathrm{o}(\sqrt{r_n v_n m_n}) = \mathrm{o}(\sqrt{n v_n})$ and condition (3.2.4) is satisfied, i.e.

$$E\Big[\Big(\sum_{j=1}^{r_n} \mathbb{1}_{\{X_j > u_n\}}\Big)^2\Big] = \mathrm{O}\left(v_n r_n\right) = \mathrm{O}(p_n).$$

Furthermore, $r_n = \mathrm{o}(\sqrt{n v_n})$ is an immediate consequence of assumptions (θ1). Moreover, by stationarity, (θ2) and $r_n v_n / p_n = \mathrm{O}(1)$ it holds that

$$\frac{1}{p_n} E\Big[\Big(\sum_{j=1}^{r_n} \mathbb{1}_{\{X_j > u_n\}}\Big)^2\Big] = \frac{1}{p_n} E\Big[\sum_{j=1}^{r_n} \sum_{i=1}^{r_n} \mathbb{1}_{\{X_j > u_n\}} \mathbb{1}_{\{X_i > u_n\}}\Big]$$
$$= \frac{r_n v_n}{p_n} + \frac{2}{p_n} \sum_{k=1}^{r_n - 1} (r_n - k) P(X_k > u_n, X_0 > u_n)$$
$$= \frac{r_n v_n}{p_n} \Big(1 + 2 \sum_{k=1}^{r_n - 1} \Big(1 - \frac{k}{r_n}\Big) P(X_k > u_n \mid X_0 > u_n)\Big)$$
$$\leq \frac{r_n v_n}{p_n} \Big(1 + 2 \sum_{k=1}^{s_n - 1} e_n(k)\Big) = \mathrm{O}(1).$$

Thus, (3.1.4) and (L) hold for $V_{n,1}^c$.
It remains to show convergence (C) of the covariance matrix. To this end, first note that by stationarity one has uniformly for all $1 \leq \ell \leq r_n - s_n$

$$\sum_{j=\ell+s_n+1}^{r_n} P\Big(M_{\ell+1,\ell+s_n} > u_n, X_j > u_n\Big)$$

$$\begin{aligned}
&\le \sum_{j=s_n+1}^{r_n} P(M_{1,s_n} > u_n, X_j > u_n)\\
&\le \sum_{i=1}^{s_n} \sum_{j=s_n+1}^{r_n} P(X_i > u_n, X_j > u_n)\\
&= s_n v_n \sum_{k=1}^{r_n} \min\left(1, \frac{k}{s_n}, \frac{r_n - k}{s_n}\right) P(X_k > u_n \mid X_0 > u_n)\\
&= \mathrm{o}(s_n v_n). \qquad (4.4.2)
\end{aligned}$$

In the last step we have used Pratt's lemma (Pratt, 1960) according to which, under condition (θP), the limit of the last sum can be calculated as the infinite sum of the limit of each summand, which all equal 0, because $k/s_n \to 0$. Likewise,

$$\begin{aligned}
&\sum_{j=1}^{\ell} P\Big(M_{\ell+1,\ell+s_n} > u_n, X_j > u_n\Big)\\
&\quad\le \sum_{j=-r_n+1}^{0} P\Big(M_{1,s_n} > u_n, X_j > u_n\Big)\\
&\quad\le \sum_{i=1}^{s_n} \sum_{j=-r_n+1}^{0} P(X_i > u_n, X_j > u_n)\\
&\quad= s_n v_n \sum_{k=1}^{r_n} \min\left(1, \frac{k}{s_n}, \frac{r_n + s_n - k}{s_n}\right) P(X_k > u_n \mid X_0 > u_n)\\
&\qquad + s_n v_n \sum_{k=r_n+1}^{r_n+s_n} \min\left(1, \frac{k}{s_n}, \frac{r_n + s_n - k}{s_n}\right) P(X_k > u_n \mid X_0 > u_n)\\
&\quad\le \mathrm{o}(s_n v_n) + s_n v_n \Big(s_n v_n + \frac{s_n}{v_n} \beta^X_{n,r_n}\Big) = \mathrm{o}(s_n v_n) \qquad (4.4.3)
\end{aligned}$$

uniformly for $1 \le \ell \le r_n$, here we also used the mixing condition from (θ1) and $s_n/v_n = \mathrm{o}(n/r_n)$. By stationarity and (4.4.1),

$$\begin{aligned}
\frac{m_n}{p_n} Var(V_n^d) &= \frac{m_n}{p_n} \frac{1}{m_n} \frac{r_n}{s_n} Var\Big(\mathbb{1}_{\{M_{1,s_n} > u_n\}}\Big)\\
&\qquad + \frac{2m_n}{p_n} \frac{1}{m_n} \sum_{1 \le i < j \le r_n/s_n} Cov\Big(\mathbb{1}_{\{M_{(i-1)s_n+1,is_n} > u_n\}}, \mathbb{1}_{\{M_{(j-1)s_n+1,js_n} > u_n\}}\Big)\\
&= \frac{r_n}{s_n p_n} P(M_{1,s_n} > u_n)(1 - P(M_{1,s_n} > u_n))\\
&\qquad + \frac{2}{p_n} \sum_{1 \le i < j \le r_n/s_n} Cov\Big(\mathbb{1}_{\{M_{(i-1)s_n+1,is_n} > u_n\}}, \mathbb{1}_{\{M_{(j-1)s_n+1,js_n} > u_n\}}\Big)\\
&= (1 + \mathrm{o}(1)) + \frac{2}{p_n} \sum_{1 \le i < j \le r_n/s_n} P\Big(M_{(i-1)s_n+1,is_n} > u_n, M_{(j-1)s_n+1,js_n} > u_n\Big)\\
&\qquad - \frac{2}{p_n} \sum_{1 \le i < j \le r_n/s_n} P\Big(M_{(i-1)s_n+1,is_n} > u_n\Big) P\Big(M_{(j-1)s_n+1,js_n} > u_n\Big)\\
&= (1 + \mathrm{o}(1)) + \frac{2}{p_n} \sum_{1 \le i < j \le r_n/s_n} P\Big(M_{(i-1)s_n+1,is_n} > u_n, M_{(j-1)s_n+1,js_n} > u_n\Big)
\end{aligned}$$

$$+ \mathrm{O}\bigg(\frac{1}{p_n}\Big(\frac{r_n}{s_n}\Big)^2 (s_n v_n)^2\bigg).$$

In view of (4.4.2), the second term can be bounded as follows:

$$\begin{aligned}
&\frac{2}{p_n} \sum_{1\le i<j\le r_n/s_n} P\Big(M_{(i-1)s_n+1,is_n} > u_n, M_{(j-1)s_n+1,js_n} > u_n\Big)\\
&= \frac{2}{p_n} \sum_{i=1}^{r_n/s_n-1} \sum_{j=i+1}^{r_n/s_n} P\Big(M_{(i-1)s_n+1,is_n} > u_n, M_{(j-1)s_n+1,js_n} > u_n\Big)\\
&= \frac{2}{p_n} \sum_{i=1}^{r_n/s_n-1} \sum_{j=i+1}^{r_n/s_n} \sum_{k=(j-1)s_n+1}^{js_n} P\Big(M_{(i-1)s_n+1,is_n} > u_n, X_k > u_n, M_{k,js_n} \le u_n\Big)\\
&\le \frac{2}{p_n} \sum_{i=1}^{r_n/s_n-1} \sum_{j=i+1}^{r_n/s_n} \sum_{k=(j-1)s_n+1}^{js_n} P\Big(M_{(i-1)s_n+1,is_n} > u_n, X_k > u_n\Big)\\
&= \frac{2}{p_n} \sum_{i=1}^{r_n/s_n-1} \sum_{k=is_n+1}^{r_n} P\Big(M_{(i-1)s_n+1,is_n} > u_n, X_k > u_n\Big)\\
&= \mathrm{o}\Big(\frac{r_n v_n}{p_n}\Big) = \mathrm{o}(1). \qquad (4.4.4)
\end{aligned}$$

Since $(r_n/s_n)^2(s_n v_n)^2/p_n = \mathrm{O}(r_n v_n) \to 0$ by (4.4.1) and (θ1), we conclude

$$\frac{m_n}{p_n} Var(V_n^d) \to 1.$$

Next check that, by (4.4.1) and (θ2),

$$\begin{aligned}
\frac{m_n}{p_n} Var(V_{n,1}^c) &= \frac{1}{p_n} Var\bigg(\sum_{j=1}^{r_n} \mathbb{1}_{\{X_j>u_n\}}\bigg)\\
&= \frac{r_n v_n}{p_n} \cdot \frac{1}{r_n v_n} E\bigg[\Big(\sum_{j=1}^{r_n} \mathbb{1}_{\{X_j>u_n\}}\Big)^2\bigg] - \frac{1}{p_n} E\Big[\sum_{j=1}^{r_n} \mathbb{1}_{\{X_j>u_n\}}\Big]^2\\
&= \frac{r_n v_n}{p_n} \cdot \frac{1}{r_n v_n} E\bigg[\bigg(\sum_{j=1}^{r_n} \mathbb{1}_{\{X_j>u_n\}}\bigg)^2\bigg] - \frac{1}{p_n}(r_n v_n)^2\\
&= \Big(\frac{1}{\theta} + \mathrm{o}(1)\Big)(c + \mathrm{o}(1)) + \mathrm{O}(r_n v_n)\\
&\to \frac{c}{\theta}. \qquad (4.4.5)
\end{aligned}$$

Finally, again by (4.4.1), (4.4.2) and (4.4.3),

$$\begin{aligned}
&\frac{m_n}{p_n} Cov\Big(V_{n,1}^d, V_{n,1}^c\Big)\\
&= \frac{1}{p_n}\bigg(\sum_{i=1}^{r_n/s_n} \sum_{j=1}^{r_n} \Big(E\Big[\mathbb{1}_{\{M_{(i-1)s_n+1,is_n}>u_n, X_j>u_n\}}\Big] - E\Big[\mathbb{1}_{\{M_{(i-1)s_n+1,is_n}>u_n\}}\Big] E\Big[\mathbb{1}_{\{X_j>u_n\}}\Big]\Big)\bigg)\\
&= \frac{1}{p_n}\bigg(\sum_{i=1}^{r_n/s_n} \sum_{j=1}^{r_n} P(M_{(i-1)s_n+1,is_n} > u_n, X_j > u_n) - \frac{r_n}{s_n} P(M_{1,s_n} > u_n) r_n v_n\bigg)
\end{aligned}$$

$$= \frac{1}{p_n} \sum_{i=1}^{r_n/s_n} \Bigg(s_n v_n + \sum_{j=1}^{(i-1)s_n} P(M_{(i-1)s_n+1, is_n} > u_n, X_j > u_n)$$
$$+ \sum_{j=is_n+1}^{r_n} P(M_{(i-1)s_n+1, is_n} > u_n, X_j > u_n) \Bigg) + \mathrm{O}(r_n v_n)$$
$$= \frac{1}{p_n} \sum_{i=1}^{r_n/s_n} \big(s_n v_n + \mathrm{o}(s_n v_n) \big) + \mathrm{O}(r_n v_n)$$
$$= \frac{r_n v_n}{p_n} + \mathrm{o}\Big(\frac{r_n v_n}{p_n} \Big) + \mathrm{O}(1) \to 1/\theta.$$

Thus, condition (C) is satisfied and the assertion follows from Theorem 3.1.4. □

The proof of Theorem 4.2.4 is based on Lemma 3.3.5.

Proof of Theorem 4.2.4. We have $E[V^c_{n,1}] = r_n v_n / \sqrt{m_n}$ and $E[V^d_{n,1}] = r_n p_n / (s_n \sqrt{m_n})$. One could apply Lemma 3.3.5 with

$$Z^1_n = \frac{1}{\sqrt{n v_n}} \frac{1}{\sqrt{p_n}} \sum_{i=1}^{m_n} V^d_{n,i} \qquad \text{and} \qquad Z^2_n = \frac{1}{\sqrt{n v_n}} \frac{1}{\sqrt{p_n}} \sum_{i=1}^{m_n} V^c_{n,i}.$$

We will follow the proof of Lemma 3.3.5 step by step, instead of applying it directly: Note that due to (4.4.1) and (θ1) it is $p_n^{1/2} m_n^{-1/2} (r_n v_n)^{-1} = (\theta/(n v_n))^{1/2}(1 + \mathrm{o}(1)) = \mathrm{o}(1)$. Direct calculations similar to the proof of Lemma 3.3.5 show that

$$\sqrt{n v_n}(\hat\theta^d_n - \theta) = \sqrt{n v_n} \left(\frac{\sum_{i=1}^{m_n} V^d_{n,i}}{\sum_{i=1}^{m_n} V^c_{n,i}} - \theta \right)$$
$$= \sqrt{n v_n} \cdot \frac{\sqrt{p_n}(Z^d_n - \theta Z^c_n) + m_n(E[V^d_n] - \theta E[V^c_n])}{m_n E[V^c_n] + \sqrt{p_n} Z^c_n}$$
$$= \sqrt{\frac{n v_n p_n}{m_n (r_n v_n)^2}} \cdot \frac{Z^d_n - \theta Z^c_n + \sqrt{m_n/p_n}\, r_n v_n \Big(P\{M_{1,s_n} > u_n\}/(s_n v_n) - \theta \Big)}{1 + \sqrt{p_n/m_n}(r_n v_n)^{-1} Z^c_n}$$
$$= \sqrt{\theta}(1 + \mathrm{o}(1)) \frac{Z^d_n - \theta Z^c_n + \mathrm{O}(\sqrt{n v_n}) \Big(P\{M_{1,s_n} > u_n\}/(s_n v_n) - \theta \Big)}{1 + \mathrm{o}_P(1)}$$
$$\to \sqrt{\theta}(Z^d - \theta Z^c),$$

where in the last step we have used Proposition 4.2.3 and the bias condition (B_b). Therefore, the asymptotic centered normal distribution of $\sqrt{n v_n}(\hat\theta^d_n - \theta)$ follows. The variance can be calculated as

$$Var(\sqrt{\theta}(Z^d - \theta Z^c)) = \theta \Big(Var(Z^d) + \theta^2 Var(Z^c) - 2\theta Cov(Z^d, Z^c) \Big)$$
$$= \theta \left(1 + \theta^2 \frac{c}{\theta} - 2\theta \frac{1}{\theta} \right) = \theta(\theta c - 1).$$

Thus, we have $\sqrt{n v_n}(\hat\theta^d_n - \theta) \xrightarrow{w} \sqrt{\theta}(Z^d - \theta Z^c) \sim \mathcal{N}(0, \theta(\theta c - 1))$. □

The next proof deals with condition (θ1S) for the sliding blocks estimator $\hat\theta^s_n$.

Proof of Lemma 4.2.5. We start with (i). The condition (4.2.5) in (θ1S) is implied by the stronger mixing condition $(n/r_n)\beta^X_{n,s_n-1} \to 0$ in (θ1), if $r_n = \mathrm{o}(\sqrt{nv_n})$. Indeed, this can be concluded by direct calculations:

$$\begin{aligned}
&\frac{1}{s_n}\sum_{l=s_n}^{r_n} P(M_{l,s_n+l} > u_n \mid X_1 > u_n) \\
&\le \frac{1}{s_n v_n}\sum_{l=s_n}^{r_n} P(M_{l,s_n+l} > u_n)P(X_1 > u_n) + \frac{r_n}{s_n v_n}\beta^X_{n,s_n-1} \\
&= \mathrm{o}\Big(\frac{r_n}{s_n v_n}(s_n v_n)v_n\Big) + \frac{r_n^2}{nv_n s_n}\frac{n}{r_n}\beta^X_{n,s_n-1} \\
&= \mathrm{o}(s_n v_n) + \mathrm{o}(1)\frac{n}{r_n}\beta^X_{n,s_n-1} \to 0. \qquad (4.4.6)
\end{aligned}$$

Because of $2s_n \le l_n$, $\beta^X_{n,s_n-1} \to 0$ implies $\beta^X_{n,l_n-s_n} \to 0$. Thus, (4.2.5) follows from $(n/r_n)\beta^X_{n,s_n-1} \to 0$. Since all other conditions of (θ1S) are directly given in (θ1), this proves the assertion.
Next we turn to part (ii). $(r_n/s_n)P(M_{s_n,r_n+s_n} > u_n|X_1 > u_n) \to 0$ implies (4.2.5), since it directly holds

$$\frac{1}{s_n}\sum_{l=s_n}^{r_n} P(M_{l,s_n+l} > u_n \mid X_1 > u_n) \le \frac{r_n}{s_n}P(M_{s_n,r_n+s_n} > u_n \mid X_1 > u_n). \qquad \square$$

Now we turn to the proofs for the asymptotic normality of the sliding blocks estimator $\hat{\theta}^s_n$. Here one has $v_n = P(X_1 > u_n)$ and now $p^s_n = P(\sum_{i=1}^{r_n}\mathbb{1}_{\{M_{1,s_n}>u_n\}} > 0) = P(M_{1,r_n+s_n-1} > u_n) = r_n v_n\theta(1+\mathrm{o}(1))$ by (4.2.1). A crucial part for the proof of Proposition 4.2.6 is the verification of the convergence of covariances in condition (C) of Theorem 3.2.1. This convergence of the variance for the function g and the covariance for h and g is shown in the following lemma.

Lemma 4.4.1. *Suppose condition (θ1S) and (θP) are satisfied. Then,*

(i) $\displaystyle\lim_{n\to\infty}\frac{1}{r_n s_n^2 v_n}Var\Big(\sum_{i=1}^{r_n}\mathbb{1}_{\{M_{i,i+s_n-1}>u_n\}}\Big) = \theta,$

(ii) $\displaystyle\lim_{n\to\infty}\frac{1}{r_n s_n v_n}Cov\Big(\sum_{i=1}^{r_n}\mathbb{1}_{\{M_{i,i+s_n-1}>u_n\}}, \sum_{j=1}^{r_n}\mathbb{1}_{\{X_j>u_n\}}\Big) = 1.$

Proof of Lemma 4.4.1. We start with the proof of (i). By the stationarity of $(X_t)_{t\in\mathbb{Z}}$ and $P(M_{1,s_n} > u_n) = \mathrm{O}(s_n v_n)$

$$\begin{aligned}
&\frac{1}{r_n s_n^2 v_n}Var\Big(\sum_{i=1}^{r_n}\mathbb{1}_{\{M_{i,i+s_n-1}>u_n\}}\Big) \\
&= \frac{1}{r_n s_n^2 v_n}E\Big[\sum_{i=1}^{r_n}\sum_{j=1}^{r_n}\mathbb{1}_{\{M_{j,j+s_n-1}>u_n\}}\mathbb{1}_{\{M_{i,i+s_n-1}>u_n\}}\Big] - \frac{r_n^2}{r_n s_n^2 v_n}P(M_{1,s_n} > u_n)^2 \\
&= \frac{r_n}{r_n s_n^2 v_n}P(M_{1,s_n} > u_n) + 2\frac{r_n}{r_n s_n^2 v_n}\sum_{k=2}^{r_n}\Big(1 - \frac{k}{r_n}\Big)P(M_{1,s_n} > u_n, M_{k,k+s_n-1} > u_n)
\end{aligned}$$

$$
\begin{aligned}
&\quad + \mathrm{O}\Big(\frac{r_n^2 s_n^2 v_n^2}{r_n s_n^2 v_n}\Big) \\
&= \mathrm{O}\Big(\frac{1}{s_n}\Big) + \frac{2}{s_n^2 v_n}\sum_{k=2}^{r_n}\Big(1-\frac{k}{r_n}\Big)P(M_{1,s_n} > u_n, M_{k,k+s_n-1} > u_n) + \mathrm{O}(r_n v_n) \\
&= 2\frac{1}{s_n^2 v_n}\sum_{k=2}^{s_n}\Big(1-\frac{k}{r_n}\Big)P(M_{1,s_n} > u_n, M_{k,k+s_n-1} > u_n) \\
&\quad + 2\frac{1}{s_n^2 v_n}\sum_{k=s_n+1}^{r_n}\Big(1-\frac{k}{r_n}\Big)P(M_{1,s_n} > u_n, M_{k,k+s_n-1} > u_n) + \mathrm{o}(1) \\
&=: 2(I + II) + \mathrm{o}(1). \qquad (4.4.7)
\end{aligned}
$$

We treat the terms I and II separately, starting with term I. In the next calculation we decompose the maximum which exceeds a threshold according to the last excedent. Then,

$$
\begin{aligned}
I &= \frac{1}{s_n^2 v_n}\sum_{k=2}^{s_n}\Big(1-\frac{k}{r_n}\Big)P(M_{1,s_n} > u_n, M_{k,k+s_n-1} > u_n) \\
&= \frac{1}{s_n^2 v_n}\sum_{k=2}^{s_n}\Big(1-\frac{k}{r_n}\Big)\sum_{l=1}^{s_n}P(M_{k,k+s_n-1} > u_n, X_l > u_n, M_{l+1,s_n} \le u_n) \\
&= \frac{1}{s_n^2}\sum_{k=2}^{s_n}\Big(1-\frac{k}{r_n}\Big)\sum_{l=1}^{s_n}P(M_{k,k+s_n-1} > u_n, M_{l+1,s_n} \le u_n \mid X_l > u_n) \\
&= \frac{1}{s_n^2}\sum_{k=2}^{s_n}\Big(1-\frac{k}{r_n}\Big)\sum_{l=k}^{s_n}P(M_{l+1,s_n} \le u_n \mid X_l > u_n) \\
&\quad + \frac{1}{s_n^2}\sum_{k=2}^{s_n}\Big(1-\frac{k}{r_n}\Big)\sum_{l=1}^{k-1}P(M_{s_n+1,k+s_n-1} > u_n, M_{l+1,s_n} \le u_n \mid X_l > u_n) \\
&=: I_1 + I_2. \qquad (4.4.8)
\end{aligned}
$$

In the last step we used

$$
\begin{aligned}
&P(M_{k,k+s_n-1} > u_n, M_{l+1,s_n} \le u_n \mid X_l > u_n) \\
&\qquad = \begin{cases} P(M_{l+1,s_n} \le u_n \mid X_l > u_n) & k \le l, \\ P(M_{s_n+1,k+s_n-1} > u_n, M_{l+1,s_n} \le u_n \mid X_l > u_n) & k > l. \end{cases}
\end{aligned}
$$

Now again we deal with both sums separately, starting with the first one. We begin with some index shift, rearranging the sums and the use of the stationarity. In the following calculation we will use an intermediate sequence t_n which fulfills $t_n \to \infty$ and $t_n = \mathrm{o}(s_n)$.

$$
\begin{aligned}
I_1 &= \frac{1}{s_n^2}\sum_{k=2}^{s_n}\Big(1-\frac{k}{r_n}\Big)\sum_{l=k}^{s_n}P(M_{l+1,s_n} \le u_n \mid X_l > u_n) \\
&= \frac{1}{s_n^2}\sum_{k=2}^{s_n}\Big(1-\frac{k}{r_n}\Big)\sum_{l=k}^{s_n}P(M_{2,s_n+1-l} \le u_n \mid X_1 > u_n)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{s_n^2}\sum_{l=2}^{s_n}\sum_{k=2}^{l}\Big(1-\frac{k}{r_n}\Big)P(M_{2,s_n+1-l}\le u_n \mid X_1>u_n)\\
&= \frac{1}{s_n^2}\sum_{l=2}^{s_n}\Big(l-\frac{l(l+1)}{2r_n}+\frac{1}{r_n}-1\Big)P(M_{2,s_n+1-l}\le u_n \mid X_1>u_n)\\
&= \frac{1}{s_n^2}\sum_{l=2}^{s_n-t_n}\Big(l-\frac{l(l+1)}{2r_n}+\frac{1}{r_n}-1\Big)P(M_{2,s_n+1-l}\le u_n \mid X_1>u_n)\\
&\qquad + \frac{1}{s_n^2}\sum_{l=s_n-t_n+1}^{s_n}\Big(l-\frac{l(l+1)}{2r_n}+\frac{1}{r_n}-1\Big)P(M_{2,s_n+1-l}\le u_n \mid X_1>u_n) \qquad (4.4.9)\\
&= \frac{1}{s_n^2}\sum_{l=2}^{s_n-t_n} lP(M_{2,s_n+1-l}\le u_n \mid X_1>u_n)\\
&\qquad - \frac{1}{s_n^2}\sum_{l=2}^{s_n-t_n}\Big(\frac{l(l+1)}{2r_n}-\frac{1}{r_n}+1\Big)P(M_{2,s_n+1-l}\le u_n \mid X_1>u_n)+\mathrm{o}(1) \qquad (4.4.10)\\
&= \frac{1}{s_n^2}\sum_{l=2}^{s_n-t_n} lP(M_{2,s_n+1-l}\le u_n \mid X_1>u_n)+\mathrm{o}(1).
\end{aligned}
$$

For this calculation check that the sum in (4.4.9) is bounded by $s_n^{-2}t_n s_n \cdot 1 = t_n/s_n \to 0$ and the sum in (4.4.10) is bounded by $s_n^{-2}(s_n-t_n)s_n(s_n+1)/(2r_n)\cdot 1 = \mathrm{O}(s_n/r_n)=\mathrm{o}(1)$. The last remaining sum has the upper bound

$$
\begin{aligned}
&\frac{1}{s_n^2}\sum_{l=2}^{s_n-t_n} lP(M_{2,s_n+1-l}\le u_n \mid X_1>u_n)\\
&\le \frac{1}{s_n^2}\sum_{l=2}^{s_n-t_n} lP(M_{2,t_n+1}\le u_n \mid X_1>u_n)\\
&= \frac{(s_n-t_n)(s_n-t_n+1)-2}{2s_n^2}P(M_{2,t_n+1}\le u_n \mid X_1>u_n)\\
&\to \frac{1}{2}\theta,
\end{aligned}
$$

since $P(M_{2,t_n+1}\le u_n \mid X_1>u_n)\to\theta$ holds by (4.2.2), which, in turn, holds due to Segers (2003), Theorem 1 and Corollary 2, under condition (θ1S) and (θP). Likewise, the same sum I_1 can be bounded from below by

$$
\begin{aligned}
&\frac{1}{s_n^2}\sum_{l=2}^{s_n-t_n} lP(M_{2,s_n+1-l}\le u_n \mid X_1>u_n)\\
&\ge \frac{1}{s_n^2}\sum_{l=2}^{s_n-t_n} lP(M_{2,s_n+1}\le u_n \mid X_1>u_n)\\
&= \frac{(s_n-t_n)(s_n-t_n+1)-2}{2s_n^2}P(M_{2,s_n+1}\le u_n \mid X_1>u_n)\\
&\to \frac{1}{2}\theta.
\end{aligned}
$$

Therefore,

$$I_1 = \frac{1}{s_n^2}\sum_{k=2}^{s_n}\left(1-\frac{k}{r_n}\right)\sum_{l=k}^{s_n} P(M_{l+1,s_n} \le u_n \mid X_l > u_n) \to \frac{1}{2}\theta.$$

Now we will continue with the second sum I_2 in (4.4.8). Here one has

$$\begin{aligned}
I_2 &= \frac{1}{s_n^2}\sum_{k=2}^{s_n}\left(1-\frac{k}{r_n}\right)\sum_{l=1}^{k-1} P(M_{s_n+1,k+s_n-1} > u_n, M_{l+1,s_n} \le u_n \mid X_l > u_n)\\
&\le \frac{1}{s_n^2}\sum_{k=2}^{s_n}\sum_{l=1}^{k-1} P(M_{s_n+1,k+s_n-1} > u_n \mid X_l > u_n)\\
&= \frac{1}{s_n^2}\sum_{k=2}^{s_n}\sum_{l=1}^{k-1} P(M_{s_n+1-l+1,k-l+1+s_n-1} > u_n \mid X_1 > u_n)\\
&\le \frac{1}{s_n^2}\sum_{k=2}^{s_n}\sum_{l=1}^{k-1} P(M_{s_n+2-l,2s_n} > u_n \mid X_1 > u_n)\\
&= \frac{1}{s_n^2}\sum_{l=1}^{s_n-1}\sum_{k=l+1}^{s_n} P(M_{s_n+2-l,2s_n} > u_n \mid X_1 > u_n)\\
&= \frac{1}{s_n^2}\sum_{l=1}^{s_n-t_n} (s_n-l)P(M_{s_n+2-l,2s_n} > u_n \mid X_1 > u_n)\\
&\qquad + \frac{1}{s_n^2}\sum_{l=s_n-t_n+1}^{s_n-1} (s_n-l)P(M_{s_n+3-l,2s_n} > u_n \mid X_1 > u_n)\\
&\le \frac{(s_n(s_n-t_n))}{s_n^2}P(M_{t_n+2,2s_n} > u_n \mid X_1 > u_n) + \frac{s_n t_n}{s_n^2}\cdot 1\\
&\le P(M_{t_n+2,2s_n} > u_n \mid X_1 > u_n) + \frac{t_n}{s_n} \to 0.
\end{aligned}$$

The last convergence holds, since $P(M_{t_n+2,2s_n} > u_n \mid X_1 > u_n) \le \sum_{k=t_n+2}^{2s_n} P(X_k > u_n \mid X_1 > u_n) \to 0$ is implied by (θP). Thus, for the first sum in (4.4.7) we obtain the convergence $2I \to 2\cdot\theta/2 = \theta$. Next we consider the second sum II in (4.4.7). Again we decompose the event of a maximum exceeding the threshold according to the last observation which exceeds the threshold.

$$\begin{aligned}
II &= \frac{1}{s_n^2 v_n}\sum_{k=s_n+1}^{r_n}\left(1-\frac{k}{r_n}\right)P(M_{1,s_n} > u_n, M_{k,k+s_n-1} > u_n)\\
&= \frac{1}{s_n^2}\sum_{k=s_n+1}^{r_n}\left(1-\frac{k}{r_n}\right)\sum_{l=1}^{s_n} P(M_{k,k+s_n-1} > u_n, M_{l+1,s_n} \le u_n \mid X_l > u_n)\\
&\le \frac{1}{s_n^2}\sum_{k=s_n+1}^{r_n}\sum_{l=1}^{s_n} P(M_{k,k+s_n-1} > u_n \mid X_l > u_n)\\
&= \frac{1}{s_n^2}\sum_{k=s_n+1}^{r_n}\sum_{l=1}^{s_n} P(M_{k-l+1,k+s_n-1-l+1} > u_n \mid X_1 > u_n)\\
&= \frac{1}{s_n^2}\sum_{k=s_n+1}^{r_n}\sum_{l=k-s_n+1}^{k} P(M_{l,s_n+l-1} > u_n \mid X_1 > u_n)
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{s_n^2}\sum_{l=1}^{r_n}\sum_{k=(s_n+1)\vee l}^{(l+s_n)\wedge r_n} P(M_{l,s_n+l} > u_n \mid X_1 > u_n) \\
&\le \frac{1}{s_n^2}\sum_{l=t_n}^{s_n-1}(s_n+1)P(M_{l,s_n+l} > u_n \mid X_1 > u_n) \\
&\qquad + \frac{1}{s_n^2}\sum_{l=s_n}^{r_n}(s_n+1)P(M_{l,s_n+l} > u_n \mid X_1 > u_n) \\
&\qquad + \frac{1}{s_n^2}\sum_{l=1}^{t_n-1}(s_n+1)P(M_{l,s_n+l} > u_n \mid X_1 > u_n) \\
&\le \frac{(s_n-t_n)(s_n+1)}{s_n^2}P(M_{t_n,2s_n} > u_n \mid X_1 > u_n) \\
&\qquad + \frac{s_n+1}{s_n}\frac{1}{s_n}\sum_{l=s_n}^{r_n} P(M_{l,s_n+l} > u_n \mid X_1 > u_n) + \frac{(t_n-1)(s_n+1)}{s_n^2} \\
&= \frac{1+\mathrm{o}(1)}{s_n}\sum_{l=s_n}^{r_n} P(M_{l,s_n+l} > u_n \mid X_1 > u_n) + \mathrm{o}(1) \to 0. \qquad (4.4.11)
\end{aligned}$$

In the penultimate step we used $P(M_{t_n,2s_n} > u_n \mid X_1 > u_n) \to 0$, which holds as before, and $t_n = \mathrm{o}(s_n)$, $t_n \to \infty$. The last convergence in (4.4.11) holds due to the assumption in (θ1S). All in all, we have shown

$$\frac{1}{r_n s_n^2 v_n} Var\Big(\sum_{i=1}^{r_n} \mathbb{1}_{\{M_{i,i+s_n-1} > u_n\}}\Big) \to 2\frac{\theta}{2} = \theta$$

which is the assertion (i).

Now we turn to the joint covariance of numerator and denominator as stated in (ii). Here stationarity yields

$$\begin{aligned}
&\frac{1}{r_n s_n v_n} Cov\Big(\sum_{i=1}^{r_n} \mathbb{1}_{\{M_{i,i+s_n-1} > u_n\}}, \sum_{i=1}^{r_n} \mathbb{1}_{\{X_i > u_n\}}\Big) \\
&= \frac{1}{r_n v_n s_n} E\Big[\sum_{i=1}^{r_n}\sum_{j=1}^{r_n} \mathbb{1}_{\{X_j > u_n\}}\mathbb{1}_{\{M_{i,i+s_n-1} > u_n\}}\Big] - \frac{1}{r_n v_n s_n} r_n r_n P(M_{1,s_n} > u_n) P(X_1 > u_n) \\
&= \frac{1}{r_n s_n v_n}\sum_{i=1}^{r_n}\sum_{j=1}^{r_n} P(M_{i,i+s_n-1} > u_n, X_j > u_n) - \frac{P(M_{1,s_n} > u_n)}{v_n s_n} r_n v_n \\
&= \frac{1}{r_n s_n v_n} r_n s_n P(X_1 > u_n) + \frac{1}{r_n s_n v_n}\sum_{i=1}^{r_n}\sum_{j=1}^{i-1} P(M_{i,i+s_n-1} > u_n, X_j > u_n) \\
&\qquad + \frac{1}{r_n s_n v_n}\sum_{i=1}^{r_n}\sum_{j=i+s_n}^{r_n} P(M_{i,i+s_n-1} > u_n, X_j > u_n) + \mathrm{O}(r_n v_n) \\
&= 1 + \frac{1}{r_n s_n v_n}\sum_{i=1}^{r_n}\sum_{j=1}^{i-1} P(M_{i,i+s_n-1} > u_n, X_j > u_n) \\
&\qquad + \frac{1}{r_n s_n v_n}\sum_{i=1}^{r_n}\sum_{j=i+s_n}^{r_n} P(M_{i,i+s_n-1} > u_n, X_j > u_n) + \mathrm{o}(1) \\
&\to 1 + c^S,
\end{aligned}$$

where c^S is defined by

$$c^S := \lim_{n\to\infty} \frac{1}{r_n s_n v_n} \sum_{i=1}^{r_n} \sum_{j=1}^{i-1} P(M_{i,i+s_n-1} > u_n, X_j > u_n) + \lim_{n\to\infty} \frac{1}{r_n s_n v_n} \sum_{i=1}^{r_n} \sum_{j=i+s_n+1}^{r_n} P(M_{i,i+s_n-1} > u_n, X_j > u_n) \geq 0. \tag{4.4.12}$$

The limit exists due to the following discussion. Equations (4.4.2) and (4.4.3) hold by stationarity and condition (θP). This yields

$$\sum_{j=1}^{i-1} P(M_{i,i+s_n-1} > u_n, X_j > u_n) = \mathrm{o}(s_n v_n),$$
$$\sum_{j=i+s_n+1}^{r_n} P(M_{i,i+s_n-1} > u_n, X_j > u_n) = \mathrm{o}(s_n v_n)$$

for all $i = 1, ..., r_n$. (As usual we interpret $\sum_{j=a}^{b} = 0$ for $a > b$.) Therefore,

$$\frac{1}{r_n s_n v_n} \sum_{i=1}^{r_n} \sum_{j=1}^{i-1} P(M_{i,i+s_n-1} > u_n, X_j > u_n) + \frac{1}{r_n s_n v_n} \sum_{i=1}^{r_n} \sum_{j=i+s_n+1}^{r_n} P(M_{i,i+s_n-1} > u_n, X_j > u_n)$$
$$= \frac{2}{r_n s_n v_n} r_n \mathrm{o}(s_n v_n) = \mathrm{o}(1) \to 0$$

and thus $c^S = 0$. This completes the proof. □

Proof of Proposition 4.2.6. We are going to apply part (a) of Theorem 3.2.1. Conditions (A), (A2) and (MX) are an immediate consequence of (θ1S). (Note that this remains true if one uses (θ1) instead of (θ1S), then with $l_n = 2s_n - 1$.) To this end, note that $b_n(g) = (nv_n/p_n^s)^{1/2} s_n$, $b_n(h) = (nv_n/p_n^s)^{1/2}$ and the conditions on convergence rates follow directly from the rates in (θ1S) and (4.4.1), which hold under the given conditions. Condition (D0) is obvious since we consider only finitely many functions.
Since $0 \leq g, h \leq 1$ we can apply Theorem 3.2.1. (3.2.4) for h follows directly from (θ2), see the proof of Proposition 4.2.3 for the calculation. To check it for g, we employ Lemma 3.2.4. First note that $p_n^s b_n(g)^2/n = s_n^2 v_n$. By stationarity of the time series

$$\frac{1}{s_n^2 v_n} \sum_{k=1}^{r_n} P(M_{1,s_n} > u_n, M_{k,k+s_n-1} > u_n)$$
$$\leq \frac{1}{s_n^2 v_n} \sum_{k=1}^{r_n} \sum_{i=1}^{s_n} \sum_{j=k}^{k+s_n-1} P(X_i > u_n, X_j > u_n)$$
$$\leq \frac{1}{s_n v_n} \sum_{i=1}^{s_n} \sum_{j=1}^{r_n+s_n-1} P(X_i > u_n, X_j > u_n)$$

$$
\begin{aligned}
&\leq \frac{1}{s_n v_n} \sum_{i=1}^{s_n} \Bigg(\sum_{j=1}^{s_n} P(X_i > u_n, X_j > u_n) + \sum_{j=s_n+1}^{r_n+s_n-1} P(X_i > u_n, X_j > u_n) \Bigg) \\
&\leq 1 + \frac{2}{v_n} \sum_{k=1}^{s_n-1} \Big(1 - \frac{k}{s_n}\Big) P(X_k > u_n, X_0 > u_n) \\
&\qquad\qquad + \frac{1}{v_n} \sum_{k=1}^{r_n+s_n-2} P(X_k > u_n, X_0 > u_n). \\
&\leq 1 + 2 \sum_{k=1}^{s_n-1} P(X_k > u_n \mid X_0 > u_n) + \sum_{k=1}^{r_n+s_n-2} P(X_k > u_n \mid X_0 > u_n) \\
&\leq 1 + 3 \sum_{k=1}^{r_n+s_n-2} P(X_k > u_n \mid X_0 > u_n).
\end{aligned}
$$

For the second step observe that for the sum over k each summand $P(X_i > u_n, X_j > u_n)$ can occur at most s_n times, since k shifts the index j and the sum over j has length s_n. Moreover,

$$
\sum_{k=r_n+1}^{r_n+s_n-2} P(X_k > u_n \mid X_0 > u_n) \leq \frac{s_n}{v_n}\Big(v_n^2 + \beta_{n,r_n}^X\Big) = \mathrm{O}\Big(s_n v_n + \frac{n}{r_n}\beta_{n,r_n}^X\Big) \to 0.
$$

Therefore, condition (S) follows from (θP) and the previous calculations. Then, condition (3.2.4) for g follows from Lemma 3.2.4.
It remains to prove convergence (C) of the standardized covariance matrix. For the variance pertaining to g and the covariance, this is done in Lemma 4.4.1. For this note, that $m_n/(p_n^s b_n(g)^2) \sim n/(r_n n v_n s_n^2) = 1/(r_n s_n^2 v_n)$ and $m_n/(p_n^s b_n(g) b_n(h)) \sim n/(r_n n v_n s_n) = 1/(r_n s_n v_n)$.
The covariance convergence

$$
\frac{m_n}{p_n^s} Var(V_n(h)) = \frac{1 + \mathrm{o}(1)}{r_n v_n} Var\Big(\sum_{j=1}^{r_n} \mathbb{1}_{\{X_j > u_n\}}\Big) \to c
$$

for the function h has been shown in (4.4.5).
Thus, all conditions are verified and part (a) of Theorem 3.2.1 provides the asserted asymptotic normality. □

The next proof of Theorem 4.2.7 combines the results of Proposition 4.2.6 and Lemma 3.3.5 to show the asymptotically normality of $\hat{\theta}_n^s$.

Proof of Theorem 4.2.7. With

$$
Z_n^1 = \frac{1}{n v_n s_n} \sum_{i=1}^{n-s_n+1} \mathbb{1}_{\{M_{i,i+s_n-1} > u_n\}} \quad \text{and} \quad Z_n^2 = \frac{1}{n v_n} \sum_{i=1}^{n-s_n+1} \mathbb{1}_{\{X_i > u_n\}}
$$

one has $E[Z_n^2] = (n v_n)^{-1}(n - s_n + 1) v_n = (n - s_n + 1)/n \to 1$ and $E[Z_n^1] = (n v_n s_n)^{-1}(n - s_n + 1) P(M_{1,s_n} > u_n) = ((n - s_n + 1)/n) P(M_{1,s_n} > u_n)/(s_n v_n)$.

Due to the bias condition (B_b) and $(n-s_n+1)/n \to 1$ it follows

$$\sqrt{nv_n}(E[Z_n^1] - \theta E[Z_n^2]) = \sqrt{nv_n}\left(\frac{P(M_{1,s_n} > u_n)}{s_n v_n} - \theta\right)\frac{n-s_n+1}{n} \to 0.$$

Thus, by Proposition 4.2.6 all conditions of Lemma 3.3.5 are met, and therefore

$$\begin{aligned}\sqrt{nv_n}(\hat{\theta}_n^s - \theta) &= \sqrt{nv_n}\Bigg(\frac{s_n^{-1}\sum_{i=1}^{n-s_n+1}\mathbb{1}_{\{M_{i,i+s_n-1}>u_n\}}}{\sum_{i=1}^{n-s_n+1}\mathbb{1}_{\{X_i>u_n\}}} - \theta\Bigg)\\ &= \sqrt{nv_n}\Bigg(\frac{Z_n^1}{Z_n^2} - \theta\Bigg) \to (Z(g) - \theta Z(h)).\end{aligned}$$

The asymptotic variance of the estimator is given by

$$\begin{aligned}Var(Z(g) - \theta Z(h)) &= Var(Z(g)) + \theta^2 Var(Z(h)) - 2\theta Cov(Z(g), Z(h))\\ &= \theta + c\theta^2 - 2\theta = \theta(c\theta - 1).\end{aligned}$$

□

4.4.2 Proofs for Section 4.2.2

In preparation for the proof of Proposition 4.2.8, the following lemma shows the convergence of the covariance needed for condition (C) from Chapter 3.

Lemma 4.4.2. *If the conditions (θ1R), (θ2) and (θP) are met, then*

(i) $$\lim_{n\to\infty}\frac{1}{r_n v_n} Var\Bigg(\sum_{i=1}^{r_n}\mathbb{1}_{\{X_i>u_n, M_{i+1,i+s_n-1}\le u_n\}}\Bigg) = \theta,$$

(ii) $$\lim_{n\to\infty}\frac{1}{r_n v_n} Cov\Bigg(\sum_{i=1}^{r_n}\mathbb{1}_{\{X_i>u_n\}}, \sum_{j=1}^{r_n}\mathbb{1}_{\{X_j>u_n, M_{j+1,j+s_n-1}\le u_n\}}\Bigg) = 1.$$

Proof of Lemma 4.4.2. For the variance of the numerator in (i) it follows by stationarity that

$$\begin{aligned}&\frac{1}{r_n v_n} Var\Bigg(\sum_{i=1}^{r_n}\mathbb{1}_{\{X_i>u_n, M_{i+1,i+s_n-1}\le u_n\}}\Bigg)\\ &= \frac{1}{r_n v_n} E\Bigg[\sum_{i=1}^{r_n}\sum_{j=1}^{r_n}\mathbb{1}_{\{X_j>u_n, M_{j+1,j+s_n-1}\le u_n\}}\mathbb{1}_{\{X_i>u_n, M_{i+1,i+s_n-1}\le u_n\}}\Bigg]\\ &\qquad - \frac{r_n^2}{r_n v_n} P(X_1 > u_n, M_{2,s_n} \le u_n)^2\\ &= \frac{r_n}{r_n v_n} P(X_1 > u_n, M_{2,s_n} \le u_n)\\ &\quad + \frac{2}{r_n v_n}\sum_{i=1}^{r_n-s_n}\sum_{j=i+s_n}^{r_n} P(X_i > u_n, M_{i+1,s_n+i-1} \le u_n, X_j > u_n, M_{j+1,s_n+j-1} \le u_n)\\ &\quad + \frac{2}{r_n v_n}\sum_{i=r_n-s_n+1}^{r_n}\sum_{j=i+1}^{r_n} P(X_i > u_n, M_{i+1,s_n+i-1} \le u_n, X_j > u_n, M_{j+1,s_n+j-1} \le u_n)\\ &\quad + \frac{2}{r_n v_n}\sum_{i=1}^{r_n-s_n}\sum_{j=i+1}^{i+s_n-1} P(X_i > u_n, M_{i+1,s_n+i-1} \le u_n, X_j > u_n, M_{j+1,s_n+j-1} \le u_n)\end{aligned}$$

$$
\begin{aligned}
&- \frac{r_n^2 v_n^2}{r_n v_n} P(M_{2,s_n} \le u_n \mid X_1 > u_n)^2 \\
&= P(M_{2,s_n} \le u_n \mid X_1 > u_n) - r_n v_n P(M_{2,s_n} \le u_n \mid X_1 > u_n)^2 \\
&\qquad + \frac{2}{r_n v_n} \sum_{i=1}^{r_n - s_n} \sum_{j=i+s_n}^{r_n} P(X_i > u_n, M_{i+1,s_n+i-1} \le u_n, X_j > u_n, M_{j+1,s_n+j-1} \le u_n) \\
&= \theta(1+\mathrm{o}(1)) + \mathrm{o}(1) \\
&\qquad + \frac{2}{r_n v_n} \sum_{i=1}^{r_n - s_n} \sum_{j=i+s_n}^{r_n} P(X_i > u_n, M_{i+1,s_n+i-1} \le u_n, X_j > u_n, M_{j+1,s_n+j-1} \le u_n)
\end{aligned} \tag{4.4.13}
$$

since $P(M_{2,s_n} \le u_n \mid X_1 > u_n) \to \theta$ by (4.2.2) and $r_n v_n \to 0$. In the second to last step we applied that obviously, for each $1 \le |i-j| < s_n$ it is

$$
P(X_i > u_n, M_{i+1,i+s_n-1} \le u_n, X_j > u_n, M_{j+1,j+s_n-1} \le u_n) = 0.
$$

The last sum in (4.4.13) can be bounded with

$$
\begin{aligned}
&\frac{2}{r_n v_n} \sum_{i=1}^{r_n - s_n} \sum_{j=i+s_n}^{r_n} P(X_i > u_n, M_{i+1,s_n+i-1} \le u_n, X_j > u_n, M_{j+1,s_n+j-1} \le u_n) \\
&\le \frac{2}{r_n v_n} \sum_{i=1}^{r_n - s_n} \sum_{j=i+s_n}^{r_n} P(X_i > u_n, X_j > u_n) \\
&\le \frac{2}{r_n v_n} \sum_{k=s_n-1}^{r_n} r_n P(X_k > u_n, X_0 > u_n) \\
&= 2 \sum_{k=s_n-1}^{r_n} P(X_k > u_n \mid X_0 > u_n) \to 0.
\end{aligned} \tag{4.4.14}
$$

This last term tends to 0 due to condition (θP). Indeed, $\lim_{n\to\infty} \sum_{k=1}^{r_n} P(X_k > u_n \mid X_0 > u_n) = \sum_{k=1}^{\infty} \lim_{n\to\infty} P(X_k > u_n \mid X_0 > u_n) < \infty$ and, therefore, $\lim_{n\to\infty} \sum_{k=1}^{s_n-2} P(X_k > u_n \mid X_0 > u_n) = \sum_{k=1}^{\infty} \lim_{n\to\infty} P(X_k > u_n \mid X_0 > u_n)$, which implies for the difference of this limits $\lim_{n\to\infty} \sum_{k=s_n-1}^{r_n} P(X_k > u_n \mid X_0 > u_n) = 0$.
Thus, we conclude from (4.4.13) that

$$
\frac{1}{r_n v_n} Var\Big(\sum_{i=1}^{r_n} \mathbb{1}_{\{X_i > u_n, M_{i+1,i+s_n-1} \le u_n\}} \Big) \to \theta.
$$

Next we turn to (ii). By stationarity

$$
\begin{aligned}
&\frac{1}{r_n v_n} Cov\Big(\sum_{i=1}^{r_n} \mathbb{1}_{\{X_i > u_n, M_{i+1,i+s_n-1} \le u_n\}}, \sum_{i=1}^{r_n} \mathbb{1}_{\{X_i > u_n\}} \Big) \\
&= \frac{1}{r_n v_n} E\Big[\sum_{i=1}^{r_n} \sum_{j=1}^{r_n} \mathbb{1}_{\{X_j > u_n\}} \mathbb{1}_{\{X_i > u_n, M_{i+1,i+s_n-1} \le u_n\}} \Big] + \frac{r_n r_n v_n}{r_n v_n} P(X_1 > u_n, M_{2,s_n} \le u_n) \\
&= \frac{1}{r_n v_n} r_n P(X_1 > u_n, M_{2,s_n} \le u_n)
\end{aligned}
$$

$$
\begin{aligned}
&\quad + \frac{r_n}{r_n v_n} \sum_{k=1}^{r_n-1} \Big(1 - \frac{k}{r_n}\Big) P(X_0 > u_n, X_k > u_n, M_{k+1,k+s_n-1} \leq u_n) \\
&\quad + \frac{r_n}{r_n v_n} \sum_{k=1}^{r_n-1} \Big(1 - \frac{k}{r_n}\Big) P(X_0 > u_n, X_k > u_n, M_{1,s_n-1} \leq u_n) + \mathrm{o}(1) \\
&= P(M_{2,s_n} \leq u_n \mid X_1 > u_n) \\
&\quad + \sum_{k=1}^{r_n-1} \Big(1 - \frac{k}{r_n}\Big) P(X_k > u_n, M_{k+1,k+s_n-1} \leq u_n \mid X_0 > u_n) \\
&\quad + \sum_{k=1}^{r_n-1} \Big(1 - \frac{k}{r_n}\Big) P(X_0 > u_n, M_{1,s_n-1} \leq u_n \mid X_k > u_n) + \mathrm{o}(1) \\
&=: P(M_{2,s_n} \leq u_n \mid X_1 > u_n) + I + II + \mathrm{o}(1).
\end{aligned}
$$

Equation (4.2.2) shows $P(M_{2,s_n} \leq u_n \mid X_1 > u_n) \to \theta$. Next we want to find the limit of I and II. We start with II:

$$
\begin{aligned}
II &= \sum_{k=1}^{r_n-1} \Big(1 - \frac{k}{r_n}\Big) P(X_0 > u_n, M_{1,s_n-1} \leq u_n \mid X_k > u_n) \\
&= \sum_{k=1}^{r_n-1} \Big(1 - \frac{k}{r_n}\Big) P(X_k > u_n, M_{1,s_n-1} \leq u_n \mid X_0 > u_n) \frac{P(X_k > u_n)}{P(X_0 > u_n)} \\
&\leq \sum_{k=2}^{r_n-1} e_n(k).
\end{aligned}
$$

Thus, with condition (θP) and Pratt's Lemma

$$
\begin{aligned}
&\lim_{n\to\infty} \sum_{k=1}^{r_n-1} \Big(1 - \frac{k}{r_n}\Big) P(X_0 > u_n, M_{1,s_n-1} \leq u_n \mid X_k > u_n) \\
&\quad = \sum_{k=1}^{\infty} \lim_{n\to\infty} \Big(1 - \frac{k}{r_n}\Big) P(X_0 > u_n, M_{1,s_n-1} \leq u_n \mid X_k > u_n) = \sum_{k=1}^{\infty} 0 = 0.
\end{aligned}
$$

The penultimate equation holds, since obviously $P(X_0 > u_n, M_{1,s_n-1} \leq u_n \mid X_k > u_n) = 0$ for all $1 \leq k \leq s_n - 1$. Thus, $II \to 0$.
For the remaining term I we obtain

$$
\begin{aligned}
I &= \sum_{k=1}^{r_n-1} \Big(1 - \frac{k}{r_n}\Big) P(X_k > u_n, M_{k+1,k+s_n-1} \leq u_n \mid X_0 > u_n) \\
&= \sum_{k=1}^{2s_n} \Big(1 - \frac{k}{r_n}\Big) P(X_k > u_n, M_{k+1,k+s_n-1} \leq u_n \mid X_0 > u_n) \\
&\qquad + \sum_{k=2s_n+1}^{r_n-1} \Big(1 - \frac{k}{r_n}\Big) P(X_k > u_n, M_{k+1,k+s_n-1} \leq u_n \mid X_0 > u_n) \\
&= \sum_{k=1}^{2s_n} \Big(1 - \frac{k}{r_n}\Big) P(X_k > u_n, M_{k+1,k+s_n-1} \leq u_n, M_{k+s_n,2s_n} \leq u_n \mid X_0 > u_n) \\
&\qquad + \sum_{k=1}^{2s_n} \Big(1 - \frac{k}{r_n}\Big) P(X_k > u_n, M_{k+1,k+s_n-1} \leq u_n, M_{k+s_n,2s_n} > u_n \mid X_0 > u_n)
\end{aligned}
$$

$$+\sum_{k=2s_n+1}^{r_n-1}\Big(1-\frac{k}{r_n}\Big)P(X_k>u_n,M_{k+1,k+s_n-1}\le u_n\mid X_0>u_n)$$
$$=\sum_{k=1}^{2s_n}P(X_k>u_n,M_{k+1,2s_n}\le u_n\mid X_0>u_n)$$
$$-\sum_{k=1}^{2s_n}\frac{k}{r_n}P(X_k>u_n,M_{k+1,2s_n}\le u_n\mid X_0>u_n)$$
$$+\sum_{k=1}^{2s_n}\Big(1-\frac{k}{r_n}\Big)P(X_k>u_n,M_{k+1,k+s_n-1}\le u_n,M_{k+s_n,2s_n}>u_n\mid X_0>u_n)$$
$$+\sum_{k=2s_n+1}^{r_n-1}\Big(1-\frac{k}{r_n}\Big)P(X_k>u_n,M_{k+1,k+s_n-1}\le u_n\mid X_0>u_n)$$
$$=: I_1+I_2+I_3+I_4.$$

In the second step we used for each $k=2,..,s_n$

$$\{X_k>u_n,M_{k+1,k+s_n-1}\le u_n\}=\{X_k>u_n,M_{k+1,k+s_n-1}\le u_n,M_{k+s_n,2s_n}\le u_n\}$$
$$\dot\cup\{X_k>u_n,M_{k+1,k+s_n-1}\le u_n,M_{k+s_n,2s_n}>u_n\}.$$

These last four sums will be considered individually. One has

$$I_1=\sum_{k=1}^{2s_n}P(X_k>u_n,M_{k+1,2s_n}\le u_n\mid X_0>u_n)$$
$$=P(M_{1,2s_n}>u_n\mid X_0>u_n)=1-P(M_{2,2s_n}\le u_n\mid X_0>u_n)\to 1-\theta.$$

In the first step we used that the sum is the decomposition of the event $\{M_{2,2s_n}>u_n\}$ by the last X_k which exceeds the threshold. The convergence holds due to (4.2.2). For the second sum we obtain

$$|I_2|=\sum_{k=1}^{2s_n}\frac{k}{r_n}P(X_k>u_n,M_{k+1,2s_n}\le u_n\mid X_0>u_n)$$
$$\le\frac{2s_n}{r_n}\sum_{k=1}^{2s_n}P(X_k>u_n\mid X_0>u_n)\le\frac{2s_n}{r_n}\sum_{k=1}^{2s_n}e_n(k)\to 0,$$

since $s_n=\mathrm{o}(r_n)$ and $\lim_{n\to\infty}\sum_{k=1}^{2s_n}e_n(k)<\infty$ by condition (θP). The third sum converges to 0, since $(s_n/v_n)\beta^X_{n,s_n-1}\to 0$ and

$$I_3=\sum_{k=1}^{2s_n}\Big(1-\frac{k}{r_n}\Big)P(X_k>u_n,M_{k+1,k+s_n-1}\le u_n,M_{k+s_n,2s_n}>u_n\mid X_0>u_n)\quad(4.4.15)$$
$$\le\frac{1}{v_n}\sum_{k=1}^{2s_n}P(X_k>u_n,M_{k+s_n,2s_n}>u_n,X_0>u_n)$$
$$\le\frac{1}{v_n}\sum_{k=1}^{2s_n}(P(X_k>u_n,X_0>u_n)P(M_{k+s_n,2s_n}>u_n)+\beta^X_{n,s_n-1})$$

$$
\begin{aligned}
&= \sum_{k=1}^{2s_n} P(X_k > u_n \mid X_0 > u_n) P(M_{k+s_n,2s_n} > u_n) + \frac{2s_n}{v_n} \beta^X_{n,s_n-1} \\
&\le \sum_{k=1}^{\infty} e_n(k) P(M_{1,s_n} > u_n) + \frac{2s_n}{v_n} \beta^X_{n,s_n-1} \to 0,
\end{aligned}
$$

where the first term converges to 0 because the sum is bounded and $P(M_{1,s_n} > u_n) = \mathrm{O}(s_n v_n) \to 0$.
The fourth and last sum converges to 0, since

$$
\begin{aligned}
I_4 &= \sum_{k=2s_n+1}^{r_n-1} \Big(1 - \frac{k}{r_n}\Big) P(X_k > u_n, M_{k+1,k+s_n-1} \le u_n \mid X_0 > u_n) \\
&\le \sum_{k=2s_n+1}^{r_n} P(X_k > u_n \mid X_0 > u_n) \le \sum_{k=2s_n+1}^{r_n} e_n(k) \to 0.
\end{aligned}
$$

The convergence holds due to $\lim_{n\to\infty} \sum_{k=1}^{r_n} e_n(k) < \infty$ which is implied by condition (θP). Putting things together, we have shown $I \to 1 - \theta$ and thereby

$$
\begin{aligned}
&\frac{1}{r_n v_n} Cov\Big(\sum_{i=1}^{r_n} \mathbb{1}_{\{X_i > u_n, M_{i+1,i+s_n-1} \le u_n\}}, \sum_{i=1}^{r_n} \mathbb{1}_{\{X_i > u_n\}} \Big) \\
&\qquad = P(M_{2,s_n} \le u_n \mid X_1 > u_n) + I + II + \mathrm{o}(1) \to \theta + 1 - \theta + 0 = 1. \qquad \square
\end{aligned}
$$

Using this preparation we can apply Theorem 3.2.1 to establish the joint convergence of numerator and denominator of $\hat{\theta}^r_n$. Note that the two functions f and h are bounded with $0 \le f, h \le 1$, which is why Theorem 3.2.1 is applicable.

Proof of Proposition 4.2.8. Since we consider only two functionals, we only need to prove fidi convergences. To this end, mainly the conditions (3.2.4) and (C) must be checked. The conditions (A), (A2) and (MX) are direct consequences from condition (θ1R). To see this, note that $b_n(f) = b_n(h) = (nv_n/p_n)^{1/2}$, $p_n = r_n v_n \theta(1 - \mathrm{o}(1))$ by (4.4.1) and therefore $m_n p_n \to \infty$, $b_n(f)^2 p_n \to \infty$, $p_n \to 0$, $q_{\mathcal{G},n} = P(X_1 > u_n) \to 0$ and $r_n = \mathrm{o}(\sqrt{p_n} b_n(f))$ directly by (θ1R). The mixing condition (MX) is directly given in (θ1R). Condition (D0) is obvious since we consider only finitely many functions.
Condition (3.2.4) for the function h was already shown in the proof of Proposition 4.2.6. Since $f(x) \neq 0$ implies $h(x) \neq 0$, the condition (3.2.4) for f follows from the same condition for h and

$$
\lim_{n\to\infty} \frac{1}{r_n v_n} E\Bigg[\Big(\sum_{i=1}^{r_n} \mathbb{1}_{\{f(X_{n,i},\dots,X_{n,i+s_n-1}) \neq 0\}} \Big)^2 \Bigg] \le \lim_{n\to\infty} \frac{1}{r_n v_n} E\Bigg[\Big(\sum_{i=1}^{r_n} \mathbb{1}_{\{X_i > u_n\}} \Big)^2 \Bigg] < \infty.
$$

The condition (C) for the function h has been verified in the proof of Proposition 4.2.6, where

$$
\frac{1}{r_n v_n} Var\Big(\sum_{i=1}^{r_n} \mathbb{1}_{\{X_i > u_n\}} \Big) = \frac{1}{r_n v_n} E\Big[\Big(\sum_{j=1}^{r_n} \mathbb{1}_{\{X_j > u_n\}} \Big)^2 \Big] - \frac{1}{r_n v_n} E\Big[\sum_{j=1}^{r_n} \mathbb{1}_{\{X_j > u_n\}} \Big]^2 \to c
$$

was shown. The remaining convergence of the covariances in condition (C) is established in Lemma 4.4.2. Thus, the joint convergence of $\bar{Z}_n(f)$ and $\bar{Z}_n(h)$ follows from part (a) of Theorem 3.2.1, which proves the assertion. □

The next short proof establishes an alternative condition for the mixing property.

Proof of Lemma 4.2.9. The condition $(s_n/v_n)\beta^X_{n,s_n-1} \to 0$ is only used to bound the sum (4.4.15) and nowhere else in the proofs of Section 4.2.2. This could also be done by

$$\begin{aligned}&\sum_{k=1}^{2s_n}\Big(1-\frac{k}{r_n}\Big)P(X_k > u_n, M_{k+1,k+s_n-1} \le u_n, M_{k+s_n,2s_n} > u_n \mid X_1 > u_n)\\&\le 2s_n P(M_{s_n,2s_n} > u_n \mid X_1 > u_n) \to 0.\end{aligned}$$

This proves the assertion. □

The method of the proof of Theorem 4.2.10 is the same as for the proof of the asymptotic behavior of $\hat{\theta}^d_n$ and $\hat{\theta}^s_n$ and combines the results of Proposition 4.2.8 and Lemma 3.3.5.

Proof of Theorem 4.2.10. With

$$Z^1_n = \frac{1}{nv_n}\sum_{i=1}^{n-s_n+1} \mathbb{1}_{\{X_i>u_n, M_{2,s_n}\le u_n\}} \quad \text{and} \quad Z^2_n = \frac{1}{nv_n}\sum_{i=1}^{n-s_n+1} \mathbb{1}_{\{X_i>u_n\}}$$

one has $E[Z^2_n] = (nv_n)^{-1}(n-s_n+1)v_n = (n-s_n+1)/n \to 1$ and $E[Z^1_n] = (nv_n)^{-1}(n-s_n+1)P(M_{2,s_n} \le u_n, X_1 > u_n) = (n-s_n+1)/nP(M_{2,s_n} \le u_n \mid X_1 > u_n)$.
With this definitions of Z^1_n and Z^2_n it follows

$$\sqrt{nv_n}\Big(E[Z^1_n] - \theta E[Z^2_n]\Big) = \sqrt{nv_n}\left(P(M_{2,s_n} \le u_n \mid X_1 > u_n) - \theta\right)\frac{n-s_n+1}{n} \to 0$$

due to the bias condition (B$_r$) and $(n-s_n+1)/n \to 1$. Thus, all conditions of Lemma 3.3.5 are satisfied, the required joint convergence is given by Proposition 4.2.8. Therefore, Lemma 3.3.5 implies

$$\begin{aligned}\sqrt{nv_n}(\hat{\theta}^r_n - \theta) &= \sqrt{nv_n}\Bigg(\frac{\sum_{i=1}^{n-s_n+1}\mathbb{1}_{\{X_i>u_n, M_{2,s_n}\le u_n\}}}{\sum_{i=1}^{n-s_n+1}\mathbb{1}_{\{X_i>u_n\}}} - \theta\Bigg)\\&= \sqrt{nv_n}\Big(\frac{Z^1_n}{Z^2_n} - \theta\Big) \xrightarrow{w} Z(f) - \theta Z(h).\end{aligned}$$

Thus, the centered asymptotic normality of the estimator $\hat{\theta}^r_n$ is shown. The asymptotic variance of the estimator is given by

$$\begin{aligned}Var(Z^R - \theta Z^N) &= Var(Z^R) + \theta^2 Var(Z^N) - 2\theta Cov(Z^R, Z^N)\\&= \theta + c\theta^2 - 2\theta(1) = \theta(1 + c\theta - 2) = \theta(c\theta - 1).\end{aligned}$$

□

4.4.3 Proofs for Section 4.2.3

The lemmas of this section and also the proof of Proposition 4.2.11 are preparations for the central proof of Theorem 4.2.12 at the end of this section.
In this section, we have

$$\begin{aligned}p_n^s &:= P(M_{1,r_n+s_n-1} > (1-\varepsilon)u_n) = (1-\varepsilon)^{-\alpha} P(M_{1,r_n+s_n-1} > u_n)(1+\mathrm{o}(1)) \\ &= (1-\varepsilon)^{-\alpha} r_n v_n \theta(1+\mathrm{o}(1))\end{aligned}$$

by (4.4.1) and regular variation.
The following technical lemma is used in the proof about the convergence of the covariances in Lemma 4.4.4. Note that the idea of the proof of the next lemma is similar to Lemma 5.2.2 in combination with Lemma 2.1.9, where a more general assumption (PP) is used to verify (AC) and to prove weak convergence of a growing segment of a regular varying time series to the tail process. For this more general statement the assumption (PP) is necessary, which needs to hold for all $c \in (0,1)$, while (θPR) needs to hold only for one threshold $(1-\varepsilon)u_n$. Recall the notation $U^*_{s,t} = \max_{s\le j\le t} \|U_j\|$ for $-\infty \le s \le t \le \infty$ and a stochastic process $(U_j)_{j\in\mathbb{Z}}$.

Lemma 4.4.3. *Suppose the conditions (θPR) and (R) are satisfied. Then, for all sequences $t_n, \tilde{t}_n, t_n^* \to \infty$, $t_n, \tilde{t}_n, t_n^* \le r_n$ and all $c, d, d^* \in [1-\varepsilon, 1+\varepsilon]$,*

$$P\Big(M_{-t_n,\tilde{t}_n} > du_n, M_{1,t_n^*} \le d^* u_n \mid X_0 > cu_n\Big) \to P\Big(Y^*_{-\infty,\infty} > d/c, Y^*_{1,\infty} \le d^*/c\Big), \quad (4.4.16)$$

$$P\Big(M_{-t_n,\tilde{t}_n} > du_n \mid X_0 > cu_n\Big) \to P\Big(Y^*_{-\infty,\infty} > d/c\Big). \quad (4.4.17)$$

Proof. First note that under (θPR) and (R), the tail process will finally not exceed $(1-\varepsilon)/(1+\varepsilon)$, i.e., $\lim_{l\to\infty} P\Big(\sup_{|t|>l} Y_t > (1-\varepsilon)/(1+\varepsilon)\Big) = 0$. To see this, check that by the Definition 2.1.6 of the tail process for all $1 \le l \le m$

$$\begin{aligned}P\Big(Y^*_{l,m} > \frac{1-\varepsilon}{1+\varepsilon}\Big) &= \lim_{n\to\infty} P\Big(M_{l,m} > (1-\varepsilon)u_n \mid X_0 > (1+\varepsilon)u_n\Big) \\ &\le \limsup_{n\to\infty} \sum_{j=l}^{m} P(X_j > (1-\varepsilon)u_n \mid X_0 > (1+\varepsilon)u_n) \\ &\le 2\sum_{j=l}^{\infty} \lim_{n\to\infty} e_n(j) \Big(\frac{1-\varepsilon}{1+\varepsilon}\Big)^{-\alpha} < \infty.\end{aligned}$$

In the last step we applied (4.2.7).
By monotone convergence, we conclude $\lim_{l\to\infty} P\Big(Y^*_{l,\infty} > (1-\varepsilon)/(1+\varepsilon)\Big) = 0$. The proof of $\lim_{l\to\infty} P\Big(Y^*_{-\infty,-l} > (1-\varepsilon)/(1+\varepsilon)\Big) = 0$ works the same way.
Hence, for any fixed $\eta > 0$, there exists $m_\eta \in \mathbb{N}$ such that for all $m \ge m_\eta$

$$\Big|P\Big(Y^*_{-\infty,\infty} > d/c, Y^*_{1,\infty} \le d^*/c\Big) - P\Big(Y^*_{-m,m} > d/c, Y^*_{1,m} \le d^*/c\Big)\Big| < \eta/3.$$

Moreover,

$$\begin{aligned}
&\Big|P\Big(M_{-t_n,\tilde{t}_n} > du_n, M_{1,t_n^*} \le d^* u_n \mid X_0 > cu_n\Big) \\
&\qquad - P\Big(M_{-m,m} > du_n, M_{1,m} \le d^* u_n \mid X_0 > cu_n\Big)\Big| \\
&\le \sum_{k=m+1}^{\max(\tilde{t}_n, t_n^*)} P(X_k > (1-\varepsilon)u_n \mid X_0 > (1-\varepsilon)u_n) \cdot \frac{P(X_0 > (1-\varepsilon)u_n)}{P(X_0 > cu_n)} \\
&\qquad + \sum_{k=m+1}^{t_n} P(X_0 > cu_n \mid X_{-k} > (1-\varepsilon)u_n) \cdot \frac{P(X_{-k} > (1-\varepsilon)u_n)}{P(X_0 > cu_n)} \\
&\le 3\Big(\frac{1-\varepsilon}{1+\varepsilon}\Big)^{-\alpha} \sum_{k=m+1}^{\max(t_n, \tilde{t}_n, t_n^*)} e_n(k) \\
&\le \frac{\eta}{3}
\end{aligned}$$

for sufficiently large m and n. The penultimate step holds due to (4.2.7) and due to the regular variation which implies $P(X_0 > (1-\varepsilon)u_n)/P(X_0 > cu_n) \le (3/2)((1-\varepsilon)/c)^{-\alpha} \le (3/2)((1-\varepsilon)/(1+\varepsilon))^{-\alpha}$ for n large enough. The last step holds due to (θPR). Therefore, one has eventually

$$\begin{aligned}
&\Big|P\Big(M_{-t_n,\tilde{t}_n} > du_n, M_{1,t_n^*} \le d^* u_n \mid X_0 > cu_n\Big) - P\Big(Y^*_{-\infty,\infty} > d/c, Y^*_{1,\infty} \le d^*/c\Big)\Big| \\
&< \Big|P\Big(M_{-m,m} > du_n, M_{1,m} \le d^* u_n \mid X_0 > cu_n\Big) - P\Big(Y^*_{-m,m} > d/c, Y^*_{1,m} \le d^*/c\Big)\Big| + \frac{2}{3}\eta \\
&< \eta
\end{aligned}$$

by the definition of the tail process (2.1.1). Since $\eta > 0$ is arbitrary, this proves (4.4.16). The second assertion follows with exactly the same arguments, just consider the event $\{Y^*_{-\infty,\infty} > d/c\}$ instead of $\{Y^*_{-\infty,\infty} > d/c, Y^*_{1,\infty} \le d^*/c\big\}$. □

Using this lemma, we show in the next step the convergence of the standardized covariance of $(Z_n(\cdot))_{\cdot \in \mathcal{G}}$. The proof of the next lemma is similar to the proof of Lemma 4.4.1, but more complicated since the regular variation and the different thresholds cu_n, du_n have to be taken into account.

Lemma 4.4.4. *If the conditions (θ1), (θPR) and (R) are met, then the following three limits exists for all $c, d \in [1-\varepsilon, 1+\varepsilon]$:*

(i)

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{r_n v_n} Cov\Big(\sum_{i=1}^{r_n} \mathbb{1}_{\{X_i > cu_n\}}, \sum_{j=1}^{r_n} \mathbb{1}_{\{X_j > du_n\}}\Big) \\
&\quad = \sum_{k=1}^{\infty} P\Big(Y_k > \frac{c}{d}\Big) d^{-\alpha} + \sum_{k=1}^{\infty} P\Big(Y_k > \frac{d}{c}\Big) c^{-\alpha} + (\max(c,d))^{-\alpha}
\end{aligned}$$

(ii)

$$\lim_{n\to\infty}\frac{1}{r_ns_nv_n}Cov\Big(\sum_{i=1}^{r_n}\mathbb{1}_{\{M_{i,i+s_n-1}>cu_n\}},\sum_{j=1}^{r_n}\mathbb{1}_{\{X_j>du_n\}}\Big)=P\Big(Y^*_{-\infty,\infty}>\frac{c}{d}\Big)d^{-\alpha}$$

(iii)

$$\lim_{n\to\infty}\frac{1}{r_ns_n^2v_n}Cov\Big(\sum_{i=1}^{r_n}\mathbb{1}_{\{M_{i,i+s_n-1}>cu_n\}},\sum_{j=1}^{r_n}\mathbb{1}_{\{M_{j,j+s_n-1}>du_n\}}\Big)$$
$$=\frac{1}{2}\Big(P\Big(Y^*_{1,\infty}\le 1,Y^*_{-\infty,\infty}>\frac{d}{c}\Big)c^{-\alpha}+P\Big(Y^*_{1,\infty}\le 1,Y^*_{-\infty,\infty}>\frac{c}{d}\Big)d^{-\alpha}\Big).$$

Proof of Lemma 4.4.4. To prove assertion (i), note that by regular variation $P(X_0>du_n)=d^{-\alpha}v_n(1+\mathrm{o}(1))$ for all $d>0$. Hence, by stationarity,

$$\frac{1}{r_nv_n}Cov\Big(\sum_{i=1}^{r_n}\mathbb{1}_{\{X_i>cu_n\}},\sum_{j=1}^{r_n}\mathbb{1}_{\{X_j>du_n\}}\Big)$$
$$=\frac{1}{r_nv_n}\sum_{i=1}^{r_n}\sum_{j=1}^{r_n}P(X_i>cu_n,X_j>du_n)+\mathrm{O}(r_nv_n)$$
$$=\sum_{k=1}^{r_n-1}\Big(1-\frac{k}{r_n}\Big)P(X_k>cu_n\mid X_0>du_n)\frac{P(X_0>du_n)}{v_n}$$
$$+\sum_{k=1}^{r_n-1}\Big(1-\frac{k}{r_n}\Big)P(X_k>du_n\mid X_0>cu_n)\frac{P(X_0>cu_n)}{v_n}$$
$$+\frac{1}{v_n}P(X_0>\max(c,d)u_n)+\mathrm{O}(r_nv_n)$$
$$\to\sum_{k=1}^{\infty}P\Big(Y_k>\frac{c}{d}\Big)d^{-\alpha}+\sum_{k=1}^{\infty}P\Big(Y_k>\frac{d}{c}\Big)c^{-\alpha}+\Big(\max(c,d)\Big)^{-\alpha}.$$

In the last step we have used regular variation and Pratt's lemma that can be applied due to Condition (θPR).
Next, note that the following generalizations of (4.4.2) and (4.4.3) hold for all $c,d\in[1-\varepsilon,1+\varepsilon]$:

$$\sum_{j=\ell+s_n+1}^{r_n}P\Big(M_{\ell+1,\ell+s_n}>cu_n,X_j>du_n\Big)$$
$$\le\sum_{j=\ell+s_n+1}^{r_n}P\Big(M_{\ell+1,\ell+s_n}>(1-\varepsilon)u_n,X_j>(1-\varepsilon)u_n\Big)=\mathrm{o}(s_nv_n) \tag{4.4.18}$$

and

$$\sum_{j=1}^{\ell}P\Big(M_{\ell+1,\ell+s_n}>cu_n,X_j>du_n\Big)=\mathrm{o}(s_nv_n) \tag{4.4.19}$$

uniformly for $1\le\ell\le r_n-s_n$. The proof is the same as for (4.4.2) and (4.4.3).

It follows for the left hand side in (ii) that

$$\begin{aligned}
&\frac{1}{r_n s_n v_n} Cov\Bigg(\sum_{i=1}^{r_n} \mathbb{1}_{\{M_{i,i+s_n-1}>cu_n\}}, \sum_{j=1}^{r_n} \mathbb{1}_{\{X_j>du_n\}}\Bigg)\\
&= \frac{1}{r_n s_n v_n}\sum_{i=1}^{r_n}\sum_{j=1}^{r_n} P(M_{i,i+s_n-1} > cu_n, X_j > du_n) + \mathrm{O}(r_n v_n)\\
&= \frac{1}{r_n s_n v_n}\sum_{i=1}^{r_n}\sum_{j=i}^{\min(i+s_n-1,r_n)} P(M_{i,i+s_n-1} > cu_n, X_j > du_n) + \mathrm{o}(1)\\
&= \frac{1}{s_n}\sum_{k=-s_n+1}^{0}\Big(1-\frac{|k|}{r_n}\Big)P(M_{k,k+s_n-1} > cu_n \mid X_0 > du_n)\frac{P(X_0 > du_n)}{v_n} + \mathrm{o}(1).
\end{aligned}$$

Moreover, for any sequence $t_n \to \infty$ with $t_n = \mathrm{o}(s_n)$, we obtain

$$\begin{aligned}
&\frac{1}{s_n}\sum_{k=-s_n+t_n+1}^{-t_n} P(M_{k,k+s_n-1} > cu_n \mid X_0 > du_n)\\
&\quad\le \frac{s_n-2t_n}{s_n}P(M_{-s_n,s_n} > cu_n \mid X_0 > du_n) \to P\Big(Y^*_{-\infty,\infty} > \frac{c}{d}\Big),
\end{aligned}$$

where (4.4.17) was applied in the last step. Likewise, for sufficiently large n,

$$\begin{aligned}
&\frac{1}{s_n}\sum_{k=-s_n+t_n+1}^{-t_n} P(M_{k,k+s_n-1} > cu_n \mid X_0 > du_n)\\
&\quad\ge \frac{s_n-2t_n}{s_n}P(M_{-t_n,t_n} > cu_n \mid X_0 > du_n) \to P\Big(Y^*_{-\infty,\infty} > \frac{c}{d}\Big).
\end{aligned}$$

Thus,

$$\begin{aligned}
&\frac{1}{s_n}\sum_{k=-s_n+1}^{0}\Big(1-\frac{|k|}{r_n}\Big)P(M_{k,k+s_n-1} > cu_n \mid X_0 > du_n)\frac{P(X_0 > du_n)}{v_n} + \mathrm{o}(1)\\
&= \frac{1}{s_n}\sum_{k=-s_n+t_n+1}^{-t_n} P(M_{k,k+s_n-1} > cu_n \mid X_0 > du_n)\frac{P(X_0 > du_n)}{v_n} + \mathrm{O}\left(\frac{t_n}{s_n}\right) + \mathrm{o}(1)\\
&\to P\Big(Y^*_{-\infty,\infty} > \frac{c}{d}\Big)d^{-\alpha}.
\end{aligned}$$

This proves (ii). Finally, we turn to (iii). The arguments are similar to the arguments used in the proof of Lemma 4.4.1. By stationarity,

$$\begin{aligned}
&Cov\Bigg(\sum_{i=1}^{r_n} \mathbb{1}_{\{M_{i,i+s_n-1}>cu_n\}}, \sum_{i=1}^{r_n} \mathbb{1}_{\{M_{i,i+s_n-1}>du_n\}}\Bigg)\\
&\quad= \sum_{i=1}^{r_n}\sum_{j=1}^{r_n} P(M_{i,i+s_n-1} > cu_n, M_{j,j+s_n-1} > du_n) + \mathrm{O}\Big((r_n s_n v_n)^2\Big)\\
&\quad= \sum_{i=1}^{r_n}\sum_{j=i}^{r_n} P(M_{i,i+s_n-1} > cu_n, M_{j,j+s_n-1} > du_n)
\end{aligned}$$

$$+\sum_{i=1}^{r_n}\sum_{j=i}^{r_n} P(M_{i,i+s_n-1} > du_n, M_{j,j+s_n-1} > cu_n) + \mathrm{o}(r_n s_n^2 v_n). \tag{4.4.20}$$

For all $c, d \in [1-\varepsilon, 1+\varepsilon]$ one can decompose the first term as follows:

$$\begin{aligned}
&\sum_{i=1}^{r_n}\sum_{j=i}^{r_n} P(M_{i,i+s_n-1} > cu_n, M_{j,j+s_n-1} > du_n)\\
&\quad = \sum_{i=1}^{r_n-3s_n}\sum_{j=i}^{i+s_n-1} P(M_{i,i+s_n-1} > cu_n, M_{j,j+s_n-1} > du_n)\\
&\qquad\quad + \sum_{i=r_n-3s_n+1}^{r_n}\sum_{j=i}^{r_n} P(M_{i,i+s_n-1} > cu_n, M_{j,j+s_n-1} > du_n)\\
&\qquad\quad + \sum_{i=1}^{r_n-3s_n}\sum_{j=i+s_n}^{r_n-s_n} P(M_{i,i+s_n-1} > cu_n, M_{j,j+s_n-1} > du_n)\\
&\qquad\quad + \sum_{i=1}^{r_n-3s_n}\sum_{j=r_n-s_n+1}^{r_n} P(M_{i,i+s_n-1} > cu_n, M_{j,j+s_n-1} > du_n)\\
&\quad =: I + II + III + IV.
\end{aligned}$$

It can be directly seen that term II is of order $s_n^2 s_n v_n = \mathrm{o}(r_n s_n^2 v_n)$. Term III can be bounded by

$$\begin{aligned}
&\sum_{i=1}^{r_n-3s_n}\sum_{j=i+s_n}^{r_n-s_n}\sum_{k=j}^{j+s_n-1} P(M_{i,i+s_n-1} > cu_n, X_k > du_n)\\
&\quad \le s_n \sum_{i=1}^{r_n-3s_n}\sum_{k=i+s_n}^{r_n} P(M_{i,i+s_n-1} > cu_n, X_k > du_n) = \mathrm{o}(r_n s_n^2 v_n)
\end{aligned}$$

due to (4.4.18). Moreover, by $(\theta 1)$,

$$\begin{aligned}
IV &\le \sum_{i=1}^{r_n-3s_n}\sum_{j=r_n-s_n+1}^{r_n} \Big(P(M_{i,i+s_n-1} > cu_n)\cdot P(M_{j,j+s_n-1} > du_n) + \beta^X_{n,s_n-1}\Big)\\
&= \mathrm{O}\Big(r_n s_n\big((s_n v_n)^2 + \beta^X_{n,s_n-1}\big)\Big) = \mathrm{o}(r_n s_n^2 v_n)
\end{aligned}$$

because $r_n/(nv_ns_n) \to 0$ and therefore $(r_ns_n/(r_ns_n^2v_n))\beta^X_{n,s_n-1} = (1/(s_nv_n))\beta^X_{n,s_n-1} \to 0$. Thus, II, III, and IV are of smaller order than the normalization $r_ns_n^2v_n$.
Next, we show that

$$\begin{aligned}
\frac{I}{r_n s_n^2 v_n} &= \frac{1+\mathrm{o}(1)}{s_n^2 v_n}\sum_{k=1}^{s_n} P(M_{1,s_n} > cu_n, M_{k,k+s_n-1} > du_n)\\
&\to \frac{1}{2} P\Big(Y^*_{1,\infty} \le 1, Y^*_{-\infty,\infty} > \frac{d}{c}\Big)c^{-\alpha}.
\end{aligned} \tag{4.4.21}$$

Distinguish the maximum according to the last exceedance in $\{1,\ldots,s_n\}$ to conclude

$$\begin{aligned}
&\sum_{k=1}^{s_n} P(M_{1,s_n} > cu_n, M_{k,k+s_n-1} > du_n) \\
&\quad= \sum_{k=1}^{s_n}\sum_{i=1}^{s_n} P(X_i > cu_n, M_{i+1,s_n} \le cu_n, M_{k,k+s_n-1} > du_n) \\
&\quad= \sum_{k=1}^{s_n}\sum_{i=k}^{s_n} P(X_i > cu_n, M_{i+1,s_n} \le cu_n, M_{k,k+s_n-1} > du_n) \\
&\qquad\quad + \mathrm{O}\bigg(\sum_{k=1}^{s_n}\sum_{i=1}^{k-1} P(X_i > cu_n, M_{k,k+s_n-1} > du_n)\bigg) \\
&\quad= \sum_{k=1}^{s_n}\sum_{i=k}^{s_n} P(X_0 > cu_n, M_{1,s_n-i} \le cu_n, M_{k-i,k-i+s_n-1} > du_n) + \mathrm{o}(s_n^2 v_n) \\
&\quad= \sum_{k=1}^{s_n}\sum_{j=k-s_n}^{0} P(X_0 > cu_n, M_{1,s_n+j-k} \le cu_n, M_{j,j+s_n-1} > du_n) + \mathrm{o}(s_n^2 v_n) \\
&\quad= \sum_{j=1-s_n}^{0}\sum_{k=1}^{s_n+j} P(X_0 > cu_n, M_{1,s_n+j-k} \le cu_n, M_{j,j+s_n-1} > du_n) + \mathrm{o}(s_n^2 v_n),
\end{aligned}$$

where in the third step we have employed (4.4.19). For any sequence $t_n \to \infty$, $t_n = \mathrm{o}(s_n)$, this last sum can eventually be bounded from below by

$$\begin{aligned}
&\sum_{j=-s_n+t_n}^{-t_n}\sum_{k=1}^{s_n+j} P(X_0 > cu_n, M_{1,s_n+j-k} \le cu_n, M_{j,j+s_n-1} > du_n) + \mathrm{O}(t_n s_n v_n) \\
&\qquad \ge \sum_{j=-s_n+t_n}^{-t_n} (s_n+j-1) P(X_0 > cu_n, M_{1,s_n} \le cu_n, M_{-t_n,t_n-1} > du_n) + \mathrm{o}(s_n^2 v_n) \\
&\qquad = \frac{s_n^2 v_n}{2}\frac{P(X_0 > cu_n)}{v_n} P\Big(Y^*_{1,\infty} \le 1, Y^*_{-\infty,\infty} > \frac{d}{c}\Big) + \mathrm{o}(s_n^2 v_n)
\end{aligned}$$

due to (4.4.16). Similarly, the sum has the upper bound

$$\begin{aligned}
&\sum_{j=t_n-s_n+1}^{0}\sum_{k=1}^{s_n+j-t_n} P(X_0 > cu_n, M_{1,s_n+j-k} \le cu_n, M_{j,j+s_n-1} > du_n) + \mathrm{O}(t_n s_n v_n) \\
&\qquad \le \sum_{j=-s_n}^{0} (s_n+j-t_n) P(X_0 > cu_n, M_{1,t_n} \le cu_n, M_{-s_n,s_n} > du_n) + \mathrm{o}(s_n^2 v_n) \\
&\qquad = \frac{s_n^2 v_n}{2} c^{-\alpha} P\Big(Y^*_{1,\infty} \le 1, Y^*_{-\infty,\infty} > \frac{d}{c}\Big) + \mathrm{o}(s_n^2 v_n).
\end{aligned}$$

Hence, convergence (4.4.21) follows, which gives the asymptotic behavior of the first term in (4.4.20). Interchanging the role of c and d yields the analogous result for the second term, which concludes the proof of (iii). □

With this preparation we can conclude Proposition 4.2.11 from Theorem 3.2.1.

Proof of Proposition 4.2.11. Since $0 \le g_d, h_d \le 1$ we can apply part (b) of Theorem 3.2.1.

The conditions (A), (A2) and (MX) follow directly from ($\theta 1$) and (R). Indeed, $P(X_0 > cu_n) = c^{-\alpha} P(X_0 > u_n)(1 + o(1))$ and $P(M_{1,s_n} > cu_n) = c^{-\alpha} P(M_{1,s_n} > u_n)(1 + o(1))$ hold by (R) for $c \in [1-\varepsilon, 1+\varepsilon]$. Condition (D0) holds by the separability of the process.
Condition (3.2.4) for $g_{1-\varepsilon}$, and hence for all g_d, $d \in [1-\varepsilon, 1+\varepsilon]$, follows from ($\theta$PR) in the same way as in the proof of Proposition 4.2.6. Condition (3.2.4) for h_d, $d \in [1-\varepsilon, 1+\varepsilon]$ can be verified in the same way as in Proposition 4.2.6 by using (4.2.7). The convergence of the covariance function is shown in Lemma 4.4.4. Thus, condition (C) is satisfied. In particular, for $c = d = 1$ one achieves the same covariances as in Proposition 4.2.6, since $P(Y_0 > 1) = 1$. (And the covariances for $d = 1$ are the only ones we need for further calculations.)
It remains to check the conditions for convergence of the process, i.e. (D1) and (D3).
The functions $(g_d)_{d \in [1-\varepsilon, 1+\varepsilon]}$ are linearly ordered, since $g_d \geq g_{d'}$ for $d \leq d'$. Moreover, the sets $\{\max_{1 \leq i < \infty} x_i > du_n\}$ are totally ordered w.r.t. inclusion. Hence, $(g_d)_{d \in (1-\varepsilon, 1+\varepsilon)}$ forms a VC(2)-class (compare Van der Vaart and Wellner (1996), Example 2.6.1) and by a remark direct after the definition of (D3) this is enough to fulfill (D3) (cf. Drees and Rootzén (2010), Remark 2.11). The same argument holds for $(h_d)_{d \in [1-\varepsilon, 1+\varepsilon]}$. Therefore, condition (D3) is satisfied. For sake of completeness we repeat the argument here for $(g_d)_{d \in [1-\varepsilon, 1+\varepsilon]}$:
The envelope function of $(g_d)_{d \in [1-\varepsilon, 1+\varepsilon]}$ is given by $g_{1-\varepsilon}$. Define the metric d_n as in condition (D3) by

$$d_n(g, h) = \left(\frac{1}{n v_n s_n^2} \sum_{j=1}^{\lfloor n/r_n \rfloor} \left(\sum_{i=1}^{r_n} g(\tilde{W}_{n,(j-1)r_n+i}) - h(\tilde{W}_{n,(j-1)r_n+i}) \right)^2 \right)^{1/2},$$

where $W_{n,t} = (X_{n,i})_{t \leq i \leq t+s_n-1}$, $X_{n,i} = X_i/u_n$ and $(\tilde{W}_{n,(j-1)r_n+i})_{1 \leq i \leq r_n}$, $1 \leq j \leq \lfloor n/r_n \rfloor$ are iid copies of the random variables $(W_{n,t})_{1 \leq t \leq r_n}$. Define Q_n as the (random) discrete probability measure which has uniform distributed mass in the points $(\tilde{W}_{n,i})_{(j-1)r_n+1 \leq i \leq jr_n}$ such that $d_n(f, g) = \sqrt{\int (f-g)^2 dQ_n}$, i.e. d_n is the $L_2(Q_n)$ semi metric.
It is important to note, that $\int \sup_{g \in \mathcal{G}} g^2 dQ_n \leq 1$ for some discrete measure Q_n. This last feature is needed to apply the VC-Theory (cf. Van der Vaart and Wellner (1996), Theorem 2.6.7). For the L_2-norm $\| \cdot \|_{Q_n,2}$ w.r.t. Q_n one has

$$\|g_{1-\varepsilon}\|_{Q_n,2}^2 = \frac{1}{n v_n s_n^2} \sum_{j=1}^{\lfloor n/r_n \rfloor} \left(\sum_{i=1}^{r_n} g_{1-\varepsilon}(\tilde{W}_{n,(j-1)r_n+i}) \right)^2 > 0$$

whenever there exists an $1 \leq j \leq \lfloor n/r_n \rfloor r_n$ with $\min(\tilde{W}_{n,j}) > 1$. If $\min(\tilde{W}_{n,j}) \leq 1$ for all $1 \leq j \leq \lfloor n/r_n \rfloor r_n$, then $d_n(g, h) = 0$ for all $g, h \in \mathcal{G}$ and the covering number is $N(\varepsilon, (g_d)_{d \in (1-\varepsilon, 1+\varepsilon)}, d_n) = 1$. In this case the entropy condition is fulfilled. For

$\|g_{1-\varepsilon}\|^2_{Q_n,2} > 0$ and with Van der Vaart and Wellner (1996),Theorem 2.6.7, it holds

$$N(\varepsilon \|g_{1-\varepsilon}\|_{Q_n,2}, (g_d)_{d\in(1-\varepsilon,1+\varepsilon)}, d_n) \le 2K(16e)^2 \left(\frac{1}{\varepsilon}\right)^{2(2-1)}$$

and thus

$$\begin{aligned}\log(N(\varepsilon, (g_d)_{d\in(1-\varepsilon,1+\varepsilon)}, d_n)) &\le \log(2K(16e)^2\|g_{1-\varepsilon}\|^2_{Q_n,2}) - 2\log(\varepsilon)\\ &\le \log(2K(16e)^2) + \log(a_n) - 2\log(\varepsilon)\end{aligned}$$

for any increasing sequence $a_n \to \infty$. The last inequality holds with probability tending to 1, since

$$\begin{aligned}&P\left(\|g_{1-\varepsilon}\|^2_{Q_n,2} > a_n\right) = P\left(\frac{1}{nv_ns_n^2}\sum_{j=1}^{\lfloor n/r_n\rfloor}\left(\sum_{i=1}^{r_n} g_{1-\varepsilon}(\tilde{W}_{n,(j-1)r_n+i})\right)^2 > a_n\right)\\ &= P\left(\sum_{j=1}^{m_n}\left(\sum_{i=1}^{r_n}\mathbb{1}_{\{\tilde{M}_{(j-1)r_n+i,(j-1)r_n+i+s_n-1}>(1-\varepsilon)u_n\}}\right)^2 > nv_ns_n^2a_n\right)\\ &\le \frac{1}{nv_ns_n^2a_n}E\left[\sum_{i=1}^{m_n}\left(\sum_{j=1}^{r_n}\mathbb{1}_{\{\tilde{M}_{(i-1)r_n+j,(i-1)r_n+j+s_n-1}>(1-\varepsilon)u_n\}}\right)^2\right]\\ &= \frac{1}{a_n}\frac{1}{r_nv_ns_n^2}E\left[\left(\sum_{j=1}^{r_n}\mathbb{1}_{\{M_{j,j+s_n-1}>(1-\varepsilon)u_n\}}\right)^2\right]\\ &= \frac{1}{a_n}\theta(1+\mathrm{o}(1)) \to 0.\end{aligned}$$

In the fourth line we applied Lemma 4.4.1 (and $r_n^2P(M_{1,s_n} > (1-\varepsilon)u_n)^2 = \mathrm{o}(r_ns_n^2v_n)$) for $(1-\varepsilon)u_n$ instead of u_n, which shows $E[(\sum_{j=1}^{r_n}\mathbb{1}_{\{M_{j,j+s_n-1}>(1-\varepsilon)u_n\}})^2]/(r_nnv_ns_n^2) \to \theta$. Therefore, since $\log(2K(16e)^2) + \log(a_n) - 2\log(\varepsilon) \ge 1$ for $\varepsilon < 1$ and $K > 1$ (K can be enlarged, if needed), it directly follows that

$$\begin{aligned}&\int_0^{\delta_n}\sqrt{\log(N(\varepsilon, (g_d)_{d\in(1-\varepsilon,1+\varepsilon)}, d_n))}d\varepsilon \le \int_0^{\delta_n}\sqrt{\log(2K(16e)^2) + \log(a_n) - 2\log(\varepsilon)}d\varepsilon\\ &\le \int_0^{\delta_n}\log(2K(16e)^2) + \log(a_n) - 2\log(\varepsilon)d\varepsilon\\ &= \delta_n\log(2K(16e)^2) + \delta_n\log(a_n) - 2\delta_n\log(\delta_n) - 2\delta_n \to 0\end{aligned}$$

for any sequence $\delta_n \downarrow 0$ and with a_n chosen so that $a_n = \mathrm{o}(1/\delta_n)$. The last convergence holds, since $\delta_n\log(\delta_n) \to 0$. The boundary holds with probability tending to 1. Hence,

$$\int_0^{\delta_n}\sqrt{\log(N(\varepsilon, (g_d)_{d\in(1-\varepsilon,1+\varepsilon)}, d_n))}d\varepsilon \xrightarrow{P^*} 0$$

for all sequences $\delta_n \downarrow 0$ and, therefore, (D3) for $(g_d)_{d\in(1-\varepsilon,1+\varepsilon)}$ is fulfilled.
The functionals $(h_d)_{d\in(1-\varepsilon,1+\varepsilon)}$ are linearly ordered, since $h_d \ge h_{d'}$ for $d \le d'$. So, $(h_d)_{d\in(1-\varepsilon,1+\varepsilon)}$ forms a VC-class and as for g_d this is enough to fulfill (D3), since $h_d \le 1$

too. From this it follows, that $\mathcal{G}$ is a VC-class (Van der Vaart and Wellner (1996), Section 2.6). Therefore, condition (D3) is satisfied.

It remains to establish condition (D1). Note, that it is enough to verify this condition for the sets of functions $(g_d)_{d\in(1-\varepsilon,1+\varepsilon)}$ and $(h_d)_{d\in(1-\varepsilon,1+\varepsilon)}$ separately. For both sets we define a semi-metric ρ_g and ρ_h. If (D1) is satisfied for both sets, one can define the semi-metric $\rho_{\mathcal{G}}$ on $\mathcal{G}$ by

$$\rho_{\mathcal{G}}(a,b) = \begin{cases} \rho_g(a,b) & \text{if } a,b \in (g_d)_{d\in(1-\varepsilon,1+\varepsilon)}, \\ \rho_h(a,b) & \text{if } a,b \in (h_d)_{d\in(1-\varepsilon,1+\varepsilon)}, \\ 1 & \text{else.} \end{cases}$$

Then (D1) is fulfilled, since for $\delta < 1$ one can consider ρ_g and ρ_h separately. In the remaining parts we will establish (D1) for the set of functions $(g_d)_{d\in(1-\varepsilon,1+\varepsilon)}$. For h_d the assertions will follow along the same lines.

Define $\rho_g(d,d') := |d^{-\alpha} - d'^{-\alpha}|$. Obviously, $(g_d)_{d\in(1-\varepsilon,1+\varepsilon)}$ is totally bounded with respect to $\rho_g(d,d')$. To ease the notation we assume w.l.o.g. $d \le d'$. Lemma 4.4.4 (iii) yields for $W_{n,t} = (X_{n,i})_{t\le i\le t+s_n-1}$

$$\begin{aligned} &\frac{1}{r_n s_n^2 v_n} E\Big[\sum_{i=1}^{r_n} g_d(W_{n,i}) \cdot \sum_{i=1}^{r_n} g_{d'}(W_{n,i})\Big] \\ &\qquad = \frac{1}{r_n s_n^2 v_n} Cov\Big(\sum_{i=1}^{r_n} g_d(W_{n,i}), \sum_{i=1}^{r_n} g_{d'}(W_{n,i})\Big) + \mathrm{O}(r_n v_n) \\ &\qquad \to \frac{1}{2}\Big(P\Big(Y^*_{1,\infty} \le 1, Y^*_{-\infty,\infty} > d'/d\Big)d^{-\alpha} + P\Big(Y^*_{1,\infty} \le 1, Y^*_{-\infty,\infty} > d/d'\Big)d'^{-\alpha}\Big) \\ &\qquad =: D(d,d'). \end{aligned} \tag{4.4.22}$$

The convergence holds due to the regular variation of the time series $(X_t^+)_{t\in\mathbb{Z}}$. Because

$$\begin{aligned} P\Big(Y^*_{1,\infty} \le 1, Y^*_{-\infty,\infty} > c\Big) &= P\Big(Y_0 \le \frac{1}{\Theta^*_{1,\infty}}, Y_0 > \frac{c}{\Theta^*_{-\infty,\infty}}\Big) \\ &= \int \Big(\Big(\max\Big(\frac{c}{t},1\Big)\Big)^{-\alpha} - \Big(\frac{1}{s}\Big)^{-\alpha}\Big)^+ P^{(\Theta^*_{1,\infty},\Theta^*_{-\infty,\infty})}(ds,dt) \end{aligned}$$

for all $c > 0$, the limit in (4.4.22) is a continuous function of $(d,d') \in [1-\varepsilon,1+\varepsilon]^2$. Moreover, the left-hand side of (4.4.22) is monotone in d and d'. Hence, convergence (4.4.22) holds uniformly on $[1-\varepsilon,1+\varepsilon]^2$.

Since $Y^*_{-\infty,\infty} > 1$ a.s., we may conclude from (4.4.22), uniformly for $1-\varepsilon \le d \le d' \le 1+\varepsilon$ that

$$\begin{aligned} &\frac{1}{r_n s_n^2 v_n} E\Big[\Big(\sum_{i=1}^{r_n}(g_d(W_{n,i}) - g_{d'}(W_{n,i}))\Big)^2\Big] \\ &\qquad = \frac{1}{r_n s_n^2 v_n}\Bigg(E\Big[\Big(\sum_{i=1}^{r_n} g_d(W_{n,i})\Big)^2\Big] + E\Big[\Big(\sum_{i=1}^{r_n} g_{d'}(W_{n,i})\Big)^2\Big] \end{aligned}$$

$$-2E\Big[\sum_{i=1}^{r_n} g_d(W_{n,i})\sum_{j=1}^{r_n} g_{d'}(W_{n,j})\Big]\Big)$$
$$\begin{aligned}
&\to D(d,d)+D(d',d')-2D(d,d')\\
&=P\Big(Y^*_{1,\infty}\le 1\Big)(d^{-\alpha}+d'^{-\alpha})-P\Big(Y^*_{1,\infty}\le 1, Y^*_{-\infty,\infty}>d'/d\Big)d^{-\alpha}\\
&\quad -P\Big(Y^*_{1,\infty}\le 1, Y^*_{-\infty,\infty}>d/d'\Big)d'^{-\alpha}\\
&=P\Big(Y^*_{1,\infty}\le 1\Big)d^{-\alpha}-P\Big(Y^*_{1,\infty}\le 1, Y^*_{-\infty,\infty}>d'/d\Big)d^{-\alpha}\\
&=P\Big(Y^*_{1,\infty}\le 1, Y^*_{-\infty,\infty}\le d'/d\Big)d^{-\alpha}\\
&\le P\Big(Y_0\le d'/d\Big)d^{-\alpha}=\Big(1-(d'/d)^{-\alpha}\Big)d^{-\alpha}=\rho(g_d,g_{d'}).
\end{aligned}$$

Hence,

$$\limsup_{n\to\infty}\sup_{d,d'\in[1-\varepsilon,1+\varepsilon],\rho(g_d,g_{d'})<\delta}\frac{1}{r_n s_n^2 v_n}E\Bigg[\Bigg(\sum_{i=1}^{r_n}(g_d(W_{n,i})-g_{d'}(W_{n,i}))\Bigg)^2\Bigg]\le\delta,$$

i.e. Condition (D1) is satisfied. Thus, all conditions for the process convergence are fulfilled and Theorem 3.2.1, part (b), provides the assertion. □

Now, Proposition 4.2.11, Slutsky's Lemma and a more involved continuous mapping argument are used to prove Theorem 4.2.12.

Proof of Theorem 4.2.12. For $d\in(1-\varepsilon,1+\varepsilon)$ define

$$\theta^S_{n,d}:=\frac{(\sqrt{nv_n}s_n)^{-1}\sum_{i=1}^{n-s_n+1}g_d(X_{n,i},...,X_{n,i+s_n-1})}{\sqrt{nv_n}^{-1}\sum_{i=1}^{n-s_n+1}h_d(X_{n,i},...,X_{n,i+s_n-1})}.$$

With this notation we have $\hat\theta^s_{n,\hat u_n}=\theta^S_{n,D_n}$ if $D_n\in[1-\varepsilon,1+\varepsilon]$, which holds with probability tending to 1, since $D_n\to 1$, i.e. $P(D_n\in[1-\varepsilon,1+\varepsilon])\to 1$. In addition define

$$\theta_n(d):=\frac{P(M_{1,s_n}>du_n)}{s_nP(X_0>du_n)}.$$

By Proposition 4.2.11 we know

$$(\bar Z_n(g_d),\bar Z_n(h_d))_{d\in[1-\varepsilon,1+\varepsilon]}\xrightarrow{w}(Z(g_d),Z(h_d))_{d\in[1-\varepsilon,1+\varepsilon]}.$$

The convergence $D_n\to 1$ in probability and Slutsky's Lemma yield

$$\Big((\bar Z_n(g_d),\bar Z_n(h_d))_{d\in[1-\varepsilon,1+\varepsilon]},D_n\Big)\xrightarrow{w}\Big((Z(g_d),Z(h_d))_{d\in[1-\varepsilon,1+\varepsilon]},1\Big).\qquad(4.4.23)$$

Skorohod's theorem provides the existence of versions of these processes which converge almost surely.

Next we show that the sample paths of the limit processes $Z(g_d)_{d\in[1-\varepsilon,1+\varepsilon]}$ are almost surely continuous. In the proof of Proposition 4.2.11 we have shown the asymptotic equicontinuity of $Z_n(g_d)_{d\in[1-\varepsilon,1+\varepsilon]}$ by establishing the condition (D1) and (D3) and using Theorem 3.2.1, part (b). From this it follows that $Z(g_d)_{d\in[1-\varepsilon,1+\varepsilon]}$ is tight (Kosorok (2008), Theorem 2.1). Applying Addendum 1.5.8 of Van der Vaart and Wellner (1996) gives the a.s. continuous sample paths with respect to the metric ρ_g, which was introduced in the proof of Proposition 4.2.11. Analogously $(Z(h_d))_{d\in[1-\varepsilon,1+\varepsilon]}$ has almost surely continuous sample paths.

Moreover,

$$\rho_g(D_n, 1) = |D_n^{-\alpha} - 1| \to 0,$$

since $D_n \to 1$ in probability, i.e. $D_n \to 1$ in probability with respect to the metric ρ_g, analogously for ρ_h.

Therefore, in view of the almost sure version of (4.4.23),

$$\begin{aligned} |\bar{Z}_n(g_{D_n}) - Z(g_1)| &\le |\bar{Z}_n(g_{D_n}) - Z(g_{D_n})| + |Z(g_{D_n}) - Z(g_1)| \\ &\le \sup_{d\in[1-\varepsilon,1+\varepsilon]} |\bar{Z}_n(g_d) - Z(g_d)| + |Z(g_{D_n}) - Z(g_1)| \to 0 \end{aligned} \tag{4.4.24}$$

almost surely. The convergence of the first term holds due to (4.4.23) the convergence of the second term holds due to the almost sure continuous sample paths. Likewise

$$|\bar{Z}_n(h_{D_n}) - Z(h_1)| \to 0 \tag{4.4.25}$$

almost surely. Then, for any sequence $d_n \to 1$,

$$\begin{aligned} &\sqrt{nv_n}(\theta^s_{n,d_n} - \theta_n(d_n)) \\ &\quad= \sqrt{nv_n}\left(\frac{\sqrt{nv_n}^{-1}\bar{Z}_n(g_{d_n}) + (n-s_n+1)/(nv_ns_n)P(M_{1,s_n} > d_nu_n)}{\sqrt{nv_n}^{-1}\bar{Z}_n(h_{d_n}) + (n-s_n+1)/(nv_n)P(X_0 > d_nu_n)} - \theta_n(d_n)\right) \\ &\quad= \frac{\bar{Z}_n(g_{d_n}) - \theta_n(d_n)\bar{Z}_n(h_{d_n})}{\sqrt{nv_n}^{-1}\bar{Z}_n(h_{d_n}) + (n-s_n+1)/(nv_n)P(X_0 > d_nu_n)} \\ &\quad= \frac{\bar{Z}_n(g_{d_n}) - \theta_n(d_n)\bar{Z}_n(h_{d_n})}{1 + o_P(1)}. \end{aligned}$$

The last equation holds, since the denominator tends to 1 by (4.4.23), $d_n \to 1$ and regular variation. The last two arguments imply $((n-s_n+1)/(nv_n))P(X_0 > d_nu_n) \to 1$.

Since $\theta_n(D_n) \to \theta$ by the bias condition (B_bR), using $D_n \to 1$, (4.4.24) and (4.4.25) we conclude

$$\sqrt{nv_n}(\theta^s_{n,D_n} - \theta_n(D_n)) \to Z(g_1) - \theta Z(h_1)$$

almost surely. Due to the bias condition (B_bR) it also holds that

$$\sqrt{nv_n}(\theta_n(D_n) - \theta) \to 0$$

and therefore

$$\sqrt{nv_n}(\hat{\theta}^s_{n,\hat{u}_n} - \theta) = \sqrt{nv_n}(\theta^S_{n,D_n} - \theta_n(D_n)) + o_P(1) \xrightarrow{w} Z(g_1) - \theta Z(h_1).$$

Hence, $Z(g_1) - \theta Z(h_1)$ is a centered normal distributed random variable with variance

$$Var(Z(g_1)) + \theta^2 Var(Z(h_1)) - 2\theta Cov(Z(g_1), Z(h_1)) = \theta(\theta c - 1),$$

the same variance as obtained in Theorem 4.2.7. □

4.4.4 Proofs for Section 4.3.1

Before we prove Theorem 4.3.1 we first state a lemma which shows the convergence of covariances.

Lemma 4.4.5. *If the conditions (S1), (SP) and (AC) hold, then*

(i) for all $s, t \in [S_0, \infty)$ one has

$$\lim_{n\to\infty} \frac{1}{r_n v_n} Cov\Big(\sum_{j=1}^{r_n} g_s(X_{n,j-s_n}, ..., X_{n,j+s_n}), \sum_{k=1}^{r_n} g_t(X_{n,k-s_n}, ..., X_{n,k+s_n})\Big)$$
$$= \theta_{sl}(\max(s,t)),$$

(ii) for all $S \in [S_0, \infty)$ one has

$$\lim_{n\to\infty} \frac{1}{r_n v_n} Cov\Big(\sum_{j=1}^{r_n} g_S(X_{n,j-s_n}, ..., X_{n,j+s_n}), \sum_{k=1}^{r_n} \mathbb{1}_{\{X_k > u_n\}}\Big)$$
$$= \sum_{k=0}^{\infty} P\left(\sum_{i=0}^{\infty} (Y_i - 1)^+ > S, Y^*_{-\infty,-1} \le 1, Y_k > 1\right).$$

Proof of Lemma 4.4.5. We start with part (i). Inserting the definition of g_s and $X_{n,i} = X_i/u_n$ yields

$$\frac{1}{r_n v_n} Cov\Big(\sum_{j=1}^{r_n} \mathbb{1}_{\{\sum_{i=j}^{j+s_n}(X_i - u_n)^+ > su_n\}} \mathbb{1}_{\{X_j > u_n\}} \mathbb{1}_{\{M_{j-s_n,j-1} \le u_n\}},$$
$$\sum_{j=1}^{r_n} \mathbb{1}_{\{\sum_{i=j}^{j+s_n}(X_i - u_n)^+ > tu_n\}} \mathbb{1}_{\{X_j > u_n\}} \mathbb{1}_{\{M_{j-s_n,j-1} \le u_n\}}\Big)$$
$$= \frac{1}{r_n v_n} E\Big[\sum_{j=1}^{r_n}\sum_{k=1}^{r_n} \mathbb{1}_{\{\sum_{i=k}^{k+s_n}(X_i - u_n)^+ > su_n\}} \mathbb{1}_{\{X_k > u_n\}} \mathbb{1}_{\{M_{k-s_n,k-1} \le u_n\}}$$
$$\mathbb{1}_{\{\sum_{i=j}^{j+s_n}(X_i - u_n)^+ > tu_n\}} \mathbb{1}_{\{X_j > u_n\}} \mathbb{1}_{\{M_{j-s_n,j-1} \le u_n\}}\Big]$$
$$+ \frac{1}{r_n v_n} r_n^2 P\Big(\sum_{i=0}^{s_n}(X_i - u_n)^+ > su_n, X_0 > u_n, M_{-s_n,-1} \le u_n\Big)$$

$$
\begin{aligned}
&\times P\Big(\sum_{i=0}^{s_n}(X_i-u_n)^+ > tu_n, X_0>u_n, M_{-s_n,-1}\le u_n\Big)\\
&= \frac{1}{r_nv_n} r_n P\Big(\sum_{i=0}^{s_n}(X_i-u_n)^+ > \max(s,t)u_n, X_0>u_n, M_{-s_n,-1}\le u_n\Big)\\
&\quad + \frac{1}{r_nv_n}\sum_{k=1}^{r_n}\sum_{j=1}^{k-1} P\Big(\sum_{i=k}^{k+s_n}(X_i-u_n)^+ > su_n, X_k>u_n, M_{k-s_n,k-1}\le u_n,\\
&\qquad\qquad \sum_{i=j}^{j+s_n}(X_i-u_n)^+ > tu_n, X_j>u_n, M_{j-s_n,j-1}\le u_n\Big)\\
&\quad + \frac{1}{r_nv_n}\sum_{k=1}^{r_n}\sum_{j=k+1}^{r_n} P\Big(\sum_{i=k}^{k+s_n}(X_i-u_n)^+ > su_n, X_k>u_n, M_{k-s_n,k-1}\le u_n,\\
&\qquad\qquad \sum_{i=j}^{j+s_n}(X_i-u_n)^+ > tu_n, X_j>u_n, M_{j-s_n,j-1}\le u_n\Big)\\
&\quad + \mathrm{O}(r_nv_n)\\
&=: I + II + III \to \theta_{sl}(\max(s,t)).
\end{aligned}
$$

For the last step we need to argue that the convergences holds. One has

$$
I = P\Big(\sum_{i=0}^{s_n}(X_i-u_n)^+ > \max(s,t)u_n, M_{-s_n,-1}\le u_n \mid X_0>u_n\Big) \to \theta_{sl}(\max(s,t))
$$

by Proposition 4.1.3 (which can be applied since (AC) holds). Thus, we only need to show that the emerging sums II and III converge to 0. One has

$$
\begin{aligned}
&II + III\\
&\le \frac{1}{r_nv_n}\sum_{k=1}^{r_n}\sum_{j=1}^{k-1} P(X_k>u_n, M_{k-s_n,k-1}\le u_n, X_j>u_n, M_{j-s_n,j-1}\le u_n)\\
&\quad + \frac{1}{r_nv_n}\sum_{k=1}^{r_n}\sum_{j=k+1}^{r_n} P(X_k>u_n, M_{k-s_n,k-1}\le u_n X_j>u_n, M_{j-s_n,j-1}\le u_n)\\
&= \frac{2}{r_nv_n}\sum_{k=1}^{r_n-s_n}\sum_{j=k+1+s_n}^{r_n} P\left(X_k>u_n, M_{k-s_n,k-1}\le u_n, X_j>u_n, M_{j-s_n,j-1}\le u_n\right),
\end{aligned}
$$

where for the last step we interchanged the role of j and k in the second sum. Therefore, $II + III \to 0$ follows from the calculations in (4.4.13) and (4.4.14).
Next, we show the convergence of the covariance between g_S and h in part (ii):

$$
\begin{aligned}
&\frac{1}{r_nv_n} Cov\Big(\sum_{j=1}^{r_n}\mathbb{1}_{\{\sum_{i=j}^{j+s_n}(X_i-u_n)^+>Su_n\}}\mathbb{1}_{\{X_j>u_n\}}\mathbb{1}_{\{M_{j-s_n,j-1}\le u_n\}}, \sum_{k=1}^{r_n}\mathbb{1}_{\{X_k>u_n\}}\Big)\\
&= \frac{1}{r_nv_n} E\Big[\sum_{j=1}^{r_n}\mathbb{1}_{\{\sum_{i=j}^{j+s_n}(X_i-u_n)^+>Su_n\}}\mathbb{1}_{\{X_j>u_n\}}\mathbb{1}_{\{M_{j-s_n,j-1}\le u_n\}} \sum_{k=1}^{r_n}\mathbb{1}_{\{X_k>u_n\}}\Big]\\
&\quad - \frac{1}{r_nv_n} r_n P\Big(\sum_{i=j}^{j+s_n}(X_i-u_n)^+ > Su_n, X_j>u_n, M_{j-s_n,j-1}\le u_n\Big) r_nv_n
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{r_n v_n} r_n E\left[\mathbb{1}_{\{\sum_{i=0}^{s_n}(X_i-u_n)^+>Su_n\}}\mathbb{1}_{\{X_0>u_n\}}\mathbb{1}_{\{M_{-s_n,-1}\le u_n\}}\right] \\
&\quad + \frac{1}{r_n v_n}\sum_{k=1}^{r_n}(r_n-k)E\left[\mathbb{1}_{\{\sum_{i=0}^{s_n}(X_i-u_n)^+>Su_n\}}\mathbb{1}_{\{X_0>u_n\}}\mathbb{1}_{\{M_{-s_n,-1}\le u_n\}}\mathbb{1}_{\{X_k>u_n\}}\right] \\
&\quad + \frac{1}{r_n v_n}\sum_{k=1}^{r_n}(r_n-k)E\left[\mathbb{1}_{\{\sum_{i=k}^{s_n+k}(X_i-u_n)^+>Su_n\}}\mathbb{1}_{\{X_k>u_n\}}\mathbb{1}_{\{M_{k-s_n,k-1}\le u_n\}}\mathbb{1}_{\{X_0>u_n\}}\right] \\
&\quad + \mathrm{o}(1) \\
&= P\Big(\sum_{i=0}^{s_n}(X_i-u_n)^+ > Su_n, M_{-s_n,-1}\le u_n | X_0>u_n\Big) \\
&\quad + \sum_{k=1}^{r_n}\Big(1-\frac{k}{r_n}\Big)P\Big(\sum_{i=0}^{s_n}(X_i-u_n)^+ > Su_n, M_{-s_n,-1}\le u_n, X_k>u_n | X_0>u_n\Big) \\
&\quad + \sum_{k=1}^{r_n}\Big(1-\frac{k}{r_n}\Big)P\Big(\sum_{i=k}^{s_n+k}(X_i-u_n)^+ > Su_n, X_k>u_n, M_{k-s_n,k-1}\le u_n | X_0>u_n\Big) \qquad (4.4.26)\\
&\quad + \mathrm{o}(1) \\
&\to \theta_{sl}(S) + \sum_{k=1}^{\infty} P\left(\sum_{i=0}^{\infty}(Y_i-1)^+ > S, Y^*_{-\infty,-1}\le 1, Y_k>1\right) \\
&= \sum_{k=0}^{\infty} P\left(\sum_{i=0}^{\infty}(Y_i-1)^+ > S, Y^*_{-\infty,-1}\le 1, Y_k>1\right).
\end{aligned}
$$

For the convergence we used condition (SP) which allows us to apply Pratt's Lemma to interchange the limes and the summation. Note that the probabilities in line (4.4.26) equal 0 for all $k \le s_n$, which is why this summands do not appear in the limit. The convergences of the other summands follows from Proposition 4.1.3. This concludes the proof. □

With this preparation we can prove Theorem 4.3.1. Note that the structure of the proof is the same as for Theorem 4.2.10. Here, we establish the joint convergence of numerator and denominator directly in this proof and do not state it as an extra proposition.

Proof of Theorem 4.3.1. In a first step we want to prove

$$
\begin{pmatrix} \sqrt{nv_n}(\tilde{\theta}^r_{sl,n}(S) - E[\tilde{\theta}^r_{sl,n}(S)])_{S\in[S_0,\infty)} \\ \sqrt{nv_n}^{-1}\sum_{i=s_n+1}^{n-s_n}\left(\mathbb{1}_{\{X_i>u_n\}} - v_n\right) \end{pmatrix} \xrightarrow{w} \begin{pmatrix} (Z^r_S)_{S\in[S_0,\infty)} \\ Z^c \end{pmatrix} \qquad (4.4.27)
$$

where $((Z^r_S)_{S\in[S_0,\infty)}, Z^c)$ is a centered Gaussian process with covariance $Var(Z^c) = c$, $Cov(Z^r_s, Z^r_t) = \theta_{sl}(\max(s,t))$ and for $Cov(Z^r_S, Z^c)$ specified in Lemma 4.4.5, part (ii). Here, we have $c = \lim_{n\to\infty} E\left[\left(\sum_{j=1}^{r_n}\mathbb{1}_{\{X_j>u_n\}}\right)^2\right]/(r_n v_n) = \sum_{k\in\mathbb{Z}} P(Y_k>1)$, since $(X_t)_{t\in\mathbb{Z}}$ is a regularly varying time series, see (4.2.3).

For the proof of the asymptotic normality we apply part (b) of Theorem 3.2.1. First note that g_S and h are bounded by $0 \le g_S, h \le 1$. The conditions (A), (A2) and (MX) are direct consequences of (S1) for $s'_n = 2s_n + 1$ and $l_n = 4s_n + 1$. Moreover, note that $p_n = P(M_{1,r_n} > u_n)$, $b_n(g_S) = b_n(h) = (nv_n/p_n)^{1/2}$ and $p_n = r_n v_n \theta(1+\mathrm{o}(1))$ by (4.4.1) under assumption (S1) and (SP). Therefore, all conditions in (A) and (A2) are directly given by (S1). The mixing condition in (MX) follows from (S1) and $\beta^X_{n,l_n-2s_n-1} \leq \beta^X_{n,s_n-1}$ with with l_n chosen above. Condition (D0) is a direct consequence of the separability of the index set $[S_0,\infty)$.

The proof of Proposition 4.2.3 shows (3.2.4) for the function $\tilde{h}$ with $\tilde{h}(x_1,...x_s) = \mathbb{1}_{\{x_1>1\}}$. Thus, this condition is satisfied for the function h with $h(x_{-s},...,x_s) = \mathbb{1}_{\{x_0>1\}}$ by the same arguments. Since $\mathbb{1}_{\{g_S(X_{n,-s_n},...,X_{n,s_n})\neq 0\}} \leq \mathbb{1}_{\{X_0>u_n\}}$ for all $S>0$ it follows

$$E\Bigg[\Big(\sum_{j=1}^{r_n}\mathbb{1}_{\{g_S(X_{n,j-s_n},...,X_{n,j+s_n})\neq 0\}}\Big)^2\Bigg] \leq E\Bigg[\Big(\sum_{j=1}^{r_n}\mathbb{1}_{\{X_j>u_n\}}\Big)^2\Bigg] = \mathrm{O}(r_n v_n).$$

The last equation follows since (3.2.4) is satisfied for h and by the definition of $b_n(h)$. Thus, (3.2.4) holds for all $S>0$.

For the application of Theorem 3.2.1 it remains to check (C), (D1) and (D3). The convergence of the variance for the function h has been verified in Proposition 4.2.8:

$$\frac{1}{r_n v_n} Var\left(\sum_{j=1}^{r_n}\mathbb{1}_{\{X_j>u_n\}}\right) \to c.$$

The remaining parts of (C) are a direct consequence of Lemma 4.4.5. Since h is a single function, it is enough to verify the conditions (D1) and (D3) for $\{g_S : S\in[S_0,\infty)\}$.

For condition (D3) we note, that we consider the functions $(g_S)_{S\in[S_0,\infty)}$ for which $g_S \leq g_{S'}$ for $S>S'$ follows directly by the definition and, thus, the subgraphs are ordered linearly. Therefore, the subgraphs of $(g_S)_{S\in(0,\infty)}$ form a VC(2)-class. According to a remark after the definition of (D3), this is enough that (D3) is fulfilled (cf. Drees and Rootzén (2010), Remark 2.11). For the detailed arguments see also the verification of the condition (D3) in the proof of Proposition 4.2.11. For details about the VC theory see Van der Vaart and Wellner (1996), Section 2.6.

Next we turn to (D1) and define the semi-metric ρ on $\mathcal{G} = \{g_s : s\in(0,\infty)\}$ by

$$\rho(g_s,g_t) := P\Big(\sum_{j=0}^{\infty}(Y_j-1)^+ \in (s,t], Y^*_{-\infty,1}\leq 1\Big)$$

for $s\leq t$. Regarding this semi-metric $\mathcal{G}$ is obviously totally bounded. Furthermore, for $s\leq t$ the following applies for $W_{n,t} = (X_{n,t})_{|t|\leq s_n}$ (with $X_{n,t} = X_t/u_n$):

$$\begin{aligned}
&\frac{1}{r_n v_n}E\Bigg[\Big(\sum_{j=1}^{r_n}(g_t(W_{n,j})-g_s(W_{n,j}))\Big)^2\Bigg]\\
&= \frac{1}{r_n v_n}E\Bigg[\Big(\sum_{j=1}^{r_n}\mathbb{1}_{\{X_j>u_n, M_{j-s_n,j-1}\leq u_n\}}\Big(\mathbb{1}_{\{\sum_{i=j}^{j+s_n}(X_i-u_n)^+>su_n\}} - \mathbb{1}_{\{\sum_{i=j}^{j+s_n}(X_i-u_n)^+>tu_n\}}\Big)\Big)^2\Bigg]\\
&= \frac{1}{r_n v_n}E\Bigg[\Big(\sum_{j=1}^{r_n}\mathbb{1}_{\{X_j>u_n, M_{j-s_n,j-1}\leq u_n\}}\mathbb{1}_{\{su_n<\sum_{i=j}^{j+s_n}(X_i-u_n)^+\leq tu_n\}}\Big)^2\Bigg]
\end{aligned}$$

$$= \frac{r_n}{r_n v_n} P\Big(X_0 > u_n, M_{-s_n,-1} \le u_n, su_n < \sum_{i=0}^{s_n}(X_i - u_n)^+ \le tu_n\Big)$$
$$+ \frac{2r_n}{r_n v_n} \sum_{k=1}^{r_n-1}\Big(1 - \frac{k}{r_n}\Big) P\Bigg(M_{-s_n,-1} \le u_n, su_n < \sum_{i=0}^{s_n}(X_i - u_n)^+ \le tu_n,$$
$$X_0 > u_n, X_k > u_n, M_{k-s_n,k-1} \le u_n, su_n < \sum_{i=k}^{k+s_n}(X_i - u_n)^+ \le tu_n\Bigg). \quad (4.4.28)$$

From this we first consider the first M summands for a $M \in \mathbb{N}$:

$$\frac{1}{v_n}\sum_{k=0}^{M}\Big(1 - \frac{k}{r_n}\Big) P\Bigg(M_{-s_n,-1} \le u_n, su_n < \sum_{i=0}^{s_n}(X_i - u_n)^+ \le tu_n,$$
$$X_0 > u_n, X_k > u_n, M_{k-s_n,k-1} \le u_n, su_n < \sum_{i=k}^{k+s_n}(X_i - u_n)^+ \le tu_n\Bigg)$$
$$\le (M+1) P\left(su_n < \sum_{i=0}^{s_n}(X_i - u_n)^+ \le tu_n, M_{-s_n,-1} \le u_n \,\middle|\, X_0 > u_n\right)$$
$$\to (M+1) P\left(\sum_{j=0}^{\infty}(Y_j - 1)^+ \in (s,t], Y^*_{-\infty,1} \le 1\right) = (M+1)\rho(g_s, g_t). \quad (4.4.29)$$

By Proposition 4.1.3 the convergence holds only for points s,t where the map $s \mapsto P\left(\sum_{j=0}^{\infty}(Y_j - 1)^+ \le s, Y^*_{-\infty,1} \le 1\right)$ is continuous.
Since $Y_t \overset{d}{=} \|Y_0\|\Theta_t$ and $\|Y_0\|$ and Θ are independent, one has

$$P\Big(\sum_{j=0}^{\infty}(Y_j - 1)^+ = s, Y^*_{-\infty,1} \le 1\Big) = \int\int \mathbb{1}_{\left\{\sum_{j=0}^{\infty}(y\theta_j - 1)^+ = s, y \le (\theta^*_{-\infty,1})^{-1}\right\}} P^{\|Y_0\|}(dy) P^{\Theta}(d\theta).$$

Fix some $\theta = (\theta_j)_{j\in\mathbb{Z}}$. The expression $\sum_{j=0}^{\infty}(y\theta_j - 1)^+$ is strictly monotonously increasing in $y > y_0$ if $\sum_{j=0}^{\infty}(y_0\theta_j - 1)^+ > 0$. Therefore, for $s > 0$ there exist at most one $y_s > 1$ such that $\sum_{j=0}^{\infty}(y_s\theta_j - 1)^+ = s$. Hence, and since $\|Y_0\|$ is Pareto(α) distributed, we obtain

$$\int \mathbb{1}_{\left\{\sum_{j=0}^{\infty}(y\theta_j - 1)^+ = s, y \le (\theta^*_{-\infty,1})^{-1}\right\}} P^{\|Y_0\|}(dy) = \int_1^{(\theta^*_{-\infty,1})^{-1}} \mathbb{1}_{\{y = y_s\}} \alpha y^{-\alpha-1} dy = 0.$$

Thus $P\left(\sum_{j=0}^{\infty}(Y_j - 1)^+ = s, Y^*_{-\infty,1} \le 1\right) = \int 0 P^{\Theta}(d\theta) = 0$, which in particular implies that the map $s \mapsto P\left(\sum_{j=0}^{\infty}(Y_j - 1)^+ \le s, Y^*_{-\infty,1} \le 1\right)$ is continuous for all $s > 0$.
The expression $P\left(\sum_{j=0}^{\infty}(Y_j - 1)^+ \in (s,t], Y^*_{-\infty,1} \le 1\right)$ is monotone in s and t. Therefore, the convergence (4.4.29) above holds uniformly in s,t, since all emerging functions including the limit function are monotone in s,t and uniformly bounded by 1 and the limit function is a measure defining function of a substochastic measure.
The limit in (4.4.29) is less than or equal to $(M+1)\delta$ if $\rho(g_s, g_t) \le \delta$. The remaining

summands in (4.4.28) can be bounded by

$$\frac{1}{v_n}\sum_{k=M+1}^{r_n}\Big(1-\frac{k}{r_n}\Big)P\Bigg(M_{-s_n,-1}\le u_n, su_n<\sum_{i=0}^{s_n}(X_i-u_n)^+<tu_n,$$
$$X_0>u_n, X_k>u_n, M_{k-s_n,k-1}\le u_n, su_n<\sum_{i=k}^{k+s_n}(X_i-u_n)^+<tu_n\Bigg)$$
$$\le\sum_{k=M+1}^{r_n}P(X_k>u_n\mid X_0>u_n)\le\sum_{k=M+1}^{r_n}e_n(k).$$

By condition (SP), one has $\lim_{M\to\infty}\sum_{k=M+1}^{r_n}e_n(k)=0$. So, by choosing M large enough and afterwards δ small enough, the considered expectation is arbitrary small, i.e.

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\sup_{g_s,g_t\in\mathcal{G},\rho(g_s,g_t)<\delta}\frac{1}{r_nv_n}E\left[\Big(\sum_{j=1}^{r_n}(g_t(W_{n,j})-g_s(W_{n,j}))\Big)^2\right]=0$$

and condition (D1) is satisfied. Thus, all conditions of Theorem 3.2.1 are satisfied, such that this theorem implies the joint convergence (4.4.27).
In the next step we prove the asymptotic normality of $\hat\theta^r_{sl,n}(S)$. Along the same lines as in the proof of Lemma 3.3.5 we obtain

$$\begin{aligned}&\sqrt{nv_n}\Big(\hat\theta^r_{sl,n}(S)-\theta_{sl}(S)\Big)_{S\in[s_0,\infty)}\\&=\sqrt{nv_n}\left(\frac{\tilde\theta^r_{sl,n}(S)}{((n-2s_n)v_n)^{-1}\sum_{i=s_n+1}^{n-s_n}\mathbb{1}_{\{X_i>u_n\}}}-\theta_{sl}(S)\right)_{S\in[S_0,\infty)}\\&\xrightarrow{w}(Z^r_S-\theta_{sl}(S)Z^c)_{S\in[S_0,\infty)}=:(Z_S)_{S\in[S_0,\infty)}.\end{aligned}$$

Note that this is not exactly the setting of Lemma 3.3.5, since we consider a whole process. The proof of the statement used here works completely analog as Lemma 3.3.5, just replace Z_n^1 by an process $(Z_n^1(s))_{s\in\mathcal{S}}$ and ξ by $\xi(s)$ for an index set $\mathcal{S}$ and adapt the bias condition to

$$\sup_{s\in\mathcal{S}}\sqrt{nv_n}\left|E[Z_n^1(s)]-\xi(s)E[Z_n^2]\right|\to 0.$$

(For this see also the proof of Theorem 4.3.2.) Here, we used $Z_n^1(S)=\tilde\theta^r_{sl,n}(S)$ and $Z_n^2=((n-2s_n)v_n)^{-1}\sum_{i=s_n+1}^{n-s_n}\mathbb{1}_{\{X_i>u_n\}}$. The joint convergence needed for Lemma 3.3.5 is given by (4.4.27) and the adapted bias condition is given by (SB$_r$). This proves the assertion. The stated covariance can be calculated with Lemma 4.4.5 by (using the notation of (4.4.27))

$$\begin{aligned}Cov(Z_s,Z_t)&=Cov\Big(Z^r_s-\theta_{sl}(s)Z^c, Z^r_t-\theta_{sl}(t)Z^c\Big)\\&=\theta_{sl}(\max(s,t))+\theta_{sl}(s)\theta_{sl}(t)c\\&\quad-\theta_{sl}(s)\sum_{k=0}^{\infty}P\Big(\sum_{j=0}^{\infty}(Y_j-1)^+>t, Y_{-\infty,-1}\le 1, Y_k>1\Big)\end{aligned}$$

$$- \theta_{sl}(t) \sum_{k=0}^{\infty} P\Big(\sum_{j=0}^{\infty} (Y_j - 1)^+ > s, Y_{-\infty,-1} \le 1, Y_k > 1 \Big). \qquad \square$$

4.4.5 Proofs for Section 4.3.2

We can prove the asymptotic normality of $\hat{\theta}^d_{sl,n}(S)$ with methods similar to the methods used for the disjoint blocks estimator $\hat{\theta}^d_n$ for the extremal index. In preparation for the proof of Theorem 4.3.2, we first prove the convergence of the covariances $V^d_{n,i}(s)$ and $V^c_{n,i}$ defined in (4.3.2).

Lemma 4.4.6. *Suppose (S1), (SP) and (SP) are met. Then it holds for all $s, t \in [S_0, \infty)$*

(i) $\displaystyle \lim_{n\to\infty} \frac{m_n}{p_n} Cov\left(V^d_{n,i}(s), V^d_{n,i}(t)\right) = \frac{1}{\theta}\theta_{sl}(\max(s,t)),$

(ii) $\displaystyle \lim_{n\to\infty} \frac{m_n}{p_n} Cov(V^d_{n,i}(s), V^c_{n,i}) = \frac{1}{\theta} P\Big(\sum_{j\in\mathbb{Z}} (Y_j - 1)^+ > s \Big).$

Proof of Lemma 4.4.6. For the covariance of $V^d_{n,i}(s)$ and $V^d_{n,i}(t)$ one obtains by stationarity

$$\begin{aligned}
&\frac{m_n}{p_n} Cov\left(V^d_{n,i}(s), V^d_{n,i}(t)\right) \\
&= \frac{1}{p_n} Cov\Big(\sum_{k=1}^{r_n/s_n} \mathbb{1}_{\{\sum_{i=1+(k-1)s_n}^{ks_n} (X_i - u_n)^+ > su_n\}}, \sum_{k=1}^{r_n/s_n} \mathbb{1}_{\{\sum_{i=1+(k-1)s_n}^{ks_n} (X_i - u_n)^+ > tu_n\}} \Big) \\
&= \frac{1}{p_n}\frac{r_n}{s_n} Cov\left(\mathbb{1}_{\{\sum_{i=1}^{s_n} (X_i - u_n)^+ > su_n\}}, \mathbb{1}_{\{\sum_{i=1}^{s_n} (X_i - u_n)^+ > tu_n\}} \right) \\
&\quad + \frac{1}{p_n} \sum_{1\le k<j\le r_n/s_n} Cov\left(\mathbb{1}_{\{\sum_{i=1+(k-1)s_n}^{ks_n} (X_i - u_n)^+ > su_n\}}, \mathbb{1}_{\{\sum_{i=1+(j-1)s_n}^{js_n} (X_i - u_n)^+ > tu_n\}} \right) \\
&\quad + \frac{1}{p_n} \sum_{1\le k<j\le r_n/s_n} Cov\left(\mathbb{1}_{\{\sum_{i=1+(k-1)s_n}^{ks_n} (X_i - u_n)^+ > tu_n\}}, \mathbb{1}_{\{\sum_{i=1+(j-1)s_n}^{js_n} (X_i - u_n)^+ > su_n\}} \right) \\
&=: I + II + III.
\end{aligned}$$

For the term I it holds that

$$\begin{aligned}
I &= \frac{1}{p_n}\frac{r_n}{s_n} Cov\left(\mathbb{1}_{\{\sum_{i=1}^{s_n} (X_i - u_n)^+ > su_n\}}, \mathbb{1}_{\{\sum_{i=1}^{s_n} (X_i - u_n)^+ > tu_n\}} \right) \\
&= \frac{r_n v_n}{p_n} \frac{1}{s_n v_n} P\Big(\sum_{i=1}^{s_n} (X_i - u_n)^+ > \max(s,t) u_n \Big) \\
&\qquad - \frac{1}{p_n}\frac{r_n}{s_n} P\Big(\sum_{i=1}^{s_n} (X_i - u_n)^+ > su_n \Big) P\Big(\sum_{i=1}^{s_n} (X_i - u_n)^+ > tu_n \Big) \\
&= \frac{1}{\theta}\theta_{sl}(\max(s,t))(1 + \mathrm{o}(1)) + \mathrm{O}\Big(\frac{r_n s_n v_n^2}{p_n} \Big) \to \frac{1}{\theta}\theta_{sl}(\max(s,t)).
\end{aligned}$$

Here we applied (4.3.1) and (4.4.1), which holds under the given conditions (S1) and (SP).

For the second term II we obtain in view of (4.4.4) that

$$\begin{aligned}
II &= \frac{1}{p_n} \sum_{1\le k<j\le r_n/s_n} P\Big(\sum_{i=1+(k-1)s_n}^{ks_n} (X_i-u_n)^+ > su_n, \sum_{i=1+(j-1)s_n}^{js_n} (X_i-u_n)^+ > tu_n \Big) \\
&\quad - \frac{1}{p_n} \sum_{1\le k<j\le r_n/s_n} P\Big(\sum_{i=1+(k-1)s_n}^{ks_n} (X_i-u_n)^+ > su_n \Big) P\Big(\sum_{i=1+(j-1)s_n}^{js_n} (X_i-u_n)^+ > tu_n \Big) \\
&\le \frac{1}{p_n} \sum_{1\le k<j\le r_n/s_n} P\left(M_{(k-1)s_n+1,ks_n} > u_n, M_{(j-1)s_n+1,js_n} > u_n \right) + \mathrm{O}(r_n v_n) = \mathrm{o}(1).
\end{aligned}$$

By the same arguments, just switching the role of t and s one obtains $III = \mathrm{o}(1)$. This proves assertion (i).

The covariance between $V^d_{n,i}(s)$ and $V^c_{n,i}$ can be calculated as follows:

$$\begin{aligned}
&\frac{m_n}{p_n} Cov(V^d_{n,i}(s), V^c_{n,i}) \\
&= \frac{1}{p_n} \sum_{k=1}^{r_n/s_n} \sum_{j=1}^{r_n} \Bigg(P\Big(\sum_{i=(k-1)s_n+1}^{ks_n} (X_i-u_n)^+ > su_n, X_j > u_n \Big) \\
&\qquad\qquad - P\Big(\sum_{i=(k-1)s_n+1}^{ks_n} (X_i-u_n)^+ > su_n \Big) P\left(X_j > u_n\right) \Bigg) \\
&= \frac{1}{p_n} \sum_{k=1}^{r_n/s_n} \sum_{j=(k-1)s_n+1}^{ks_n} P\Big(\sum_{i=(k-1)s_n+1}^{ks_n} (X_i-u_n)^+ > su_n, X_j > u_n \Big) \\
&\quad + \frac{1}{p_n} \sum_{k=1}^{r_n/s_n} \Bigg(\sum_{j=1}^{(k-1)s_n} P\Big(\sum_{i=(k-1)s_n+1}^{ks_n} (X_i-u_n)^+ > su_n, X_j > u_n \Big) \\
&\qquad + \sum_{j=ks_n+1}^{r_n} P\Big(\sum_{i=(k-1)s_n+1}^{ks_n} (X_i-u_n)^+ > su_n, X_j > u_n \Big) \Bigg) + \mathrm{O}\Big(\frac{r_n^2}{p_n s_n} s_n v_n^2 \Big) \\
&= \frac{1}{p_n} \sum_{k=1}^{r_n/s_n} \sum_{j=(k-1)s_n+1}^{ks_n} P\Big(\sum_{i=(k-1)s_n+1}^{ks_n} (X_i-u_n)^+ > su_n, X_j > u_n \Big) + \mathrm{o}(1) \\
&= \frac{1}{p_n} \sum_{k=1}^{r_n/s_n} \sum_{j=1}^{s_n} P\Big(\sum_{i=1}^{s_n} (X_i-u_n)^+ > su_n, X_j > u_n \Big) + \mathrm{o}(1) \\
&= \frac{r_n v_n}{p_n} \frac{1}{s_n v_n} \sum_{j=1}^{s_n} P\Big(\sum_{i=1}^{s_n} (X_i-u_n)^+ > su_n, X_j > u_n \Big) + \mathrm{o}(1).
\end{aligned}$$

In the third to last step we applied (4.4.2) and (4.4.3). Since each probability is at most 1, for the remaining sum it holds

$$\begin{aligned}
&\frac{1}{s_n} \sum_{k=1}^{s_n} P\Big(\sum_{i=1}^{s_n} (X_i-u_n)^+ > su_n \mid X_k > u_n \Big) \\
&= \frac{1}{s_n} \sum_{k=t_n}^{s_n-t_n} P\Big(\sum_{i=1-k}^{s_n-k} (X_i-u_n)^+ > su_n \mid X_0 > u_n \Big) + \mathrm{O}\Big(\frac{t_n}{s_n} \Big). \qquad (4.4.30)
\end{aligned}$$

Here, we used some sequence $t_n \to \infty$ with $t_n = o(s_n)$. This last sum can be bounded by

$$\frac{1}{s_n}\sum_{k=t_n}^{s_n-t_n} P\Big(\sum_{i=1-k}^{s_n-k}(X_i-u_n)^+ > su_n \mid X_0 > u_n\Big)$$
$$\leq \frac{s_n-2t_n}{s_n} P\Big(\sum_{i=1-s_n}^{s_n}(X_i-u_n)^+ > su_n \mid X_0 > u_n\Big) \to P\Big(\sum_{j\in\mathbb{Z}}(Y_j-1)^+ > s\Big).$$

Furthermore, it can be bounded from below by

$$\frac{1}{s_n}\sum_{k=t_n}^{s_n-t_n} P\Big(\sum_{i=1-k}^{s_n-k}(X_i-u_n)^+ > su_n \mid X_0 > u_n\Big)$$
$$\geq \frac{s_n-2t_n}{s_n} P\Big(\sum_{i=1-t_n}^{t_n}(X_i-u_n)^+ > su_n \mid X_0 > u_n\Big) \to P\Big(\sum_{j\in\mathbb{Z}}(Y_j-1)^+ > s\Big).$$

The convergences hold due to Proposition 4.1.3, which can be applied since (AC) is assumed to hold. Since upper and lower bound coincide and $(r_n v_n)/p_n \to 1/\theta$, this proves the assertion. □

The proof of Theorem 4.3.2 works along the same lines as the proof of Theorem 4.2.4, but here we also have to deal with process convergence.

Proof of Theorem 4.3.2. First, we will show the weak convergence

$$\begin{pmatrix} (Z_n^d(S))_{S\in[s_0,\infty)} \\ Z_n^c \end{pmatrix} := \begin{pmatrix} \Big(p_n^{-1/2}\sum_{i=1}^{m_n}\Big(V_{n,i}^d - E[V_{n,i}^d]\Big)\Big)_{S\in[S_0,\infty)} \\ p_n^{-1/2}\sum_{i=1}^{m_n}\Big(V_{n,i}^c - E[V_{n,i}^c]\Big) \end{pmatrix} \to \begin{pmatrix} (Z_S^d)_{S\in[S_0,\infty)} \\ Z^c \end{pmatrix}, \tag{4.4.31}$$

where $((Z_S^d)_{S\in[S_0,\infty)}, Z^c)$ is a centered Gaussian process with covariance $Cov(Z_s^d, Z_t^d) = \theta_{sl}(\max(s,t))/\theta$, $Var(Z^c) = c/\theta$ (c is given in (4.2.3), note that under regular variation the limit there is the same if r_n is replaced by s_n) and for the constants $Cov(Z_S^d, Z^c) = P\left(\sum_{j\in\mathbb{Z}}(Y_j-1)^+ > S\right)/\theta$.
We will apply the abstract limit Theorem 3.1.10. The condition (A1) is directly given by (S1). By the definition of $V_{n,i}^d(D)$, $\tilde V_{n,i}^d(D)$, $V_{n,i}^c$ and $\tilde V_{n,i}^c$ the condition (V) is directly implied by the stationarity of $(X_t)_{t\in\mathbb{Z}}$. Conditions (M$\tilde V$) and (MX$_2$) follow from the β-mixing condition in (S1). The latter condition follows since $r_n - s_n > s_n - 1$ for sufficiently large n. Condition (D0) is a direct consequence of the separability of the index set $[S_0,\infty)$. The conditions (Δ) (or (3.1.4)) and (L) can be checked separately for $V_{n,i}^d(S)$ and $V_{n,i}^c$. For $V_{n,i}^c$ these conditions have been verified in the proof of Proposition 4.2.3. For $V_{n,i}^d(S)$ one has

$$\Delta_n^d(S) = V_{n,i}^d(S) - \tilde V_{n,i}^d(S) \overset{d}{=} \frac{1}{\sqrt{m_n}}\mathbb{1}_{\{\sum_{i=1}^{s_n}(X_i-u_n)^+ > Su_n\}} \leq \frac{1}{\sqrt{m_n}}\mathbb{1}_{\{M_{1,s_n} > u_n\}}.$$

Therefore, (3.1.4) and (L) are implied by the verification of the same conditions in the proof of Proposition 4.2.3.
For the process convergence the conditions (D1) and (D3) have to be satisfied. The condition (D3) is fulfilled since the functions $(f_S)_{S\in[S_0,\infty)}$ with $f_S(x_1,...,x_m) = \mathbb{1}_{\{\sum_{i=1}^m (x_i-1)^+ > S\}}$ are linearly ordered by $f_s \le f_t$ for $s \ge t$. Therefore, the functions form a $VC(2)$-class which is enough such that (D3) holds (cf. proof of Proposition 4.2.11). The single function needed for $V^c_{n,i}$ does not matter, (D3) remains fulfilled if one adds just one single function to the set of functions. Condition (D1) can be established in the same way as in the proof of Theorem 4.3.1 for the runs estimator. The arguments are exactly the same, just a modified function is used and one has to use the semi-metric $\rho(H_s,H_t) := P\left(\sum_{j=0}^\infty (Y_j-1)^+ \in (s,t]\right)$. For this argument one applies Proposition 4.1.3, which is why (AC) is needed.
The convergence of the variances for $V^c_{n,i}$ was already established in Proposition 4.2.3, the remaining covariances of condition (C) converge due to Lemma 4.4.6. Therefore, all conditions of Theorem 3.1.10 are satisfied and the weak convergence (4.4.31) holds.
Next, we establish the weak convergence as stated in the assertion. Similar to the proof of Lemma 3.3.5 it follows

$$\begin{aligned}
&\sqrt{nv_n}\left(\hat\theta^d_{sl,n}(S) - \theta_{sl}(S)\right)_{S\in[s_0,\infty)} \\
&= \sqrt{nv_n}\left(\frac{\sum_{i=1}^{m_n} V^d_{n,i}(S)}{\sum_{i=1}^{m_n} V^c_{n,i}} - \theta_{sl}(S)\right)_{S\in[S_0,\infty)} \\
&= \sqrt{nv_n}\cdot\left(\frac{\sqrt{p_n}(Z^d_n(S) - \theta_{sl}(S)Z^c_n) + m_n(E[V^d_n] - \theta_{sl}(S)E[V^c_n])}{m_nE[V^c_n] + \sqrt{p_n}Z^c_n}\right)_{S\in[S_0,\infty)} \\
&= \sqrt{\frac{nv_np_n}{m_n(r_nv_n)^2}}\cdot\left(\frac{Z^d_n(S) - \theta_{sl}(S)Z^c_n}{1+\sqrt{p_n/m_n}(r_nv_n)^{-1}Z^c_n}\right. \\
&\qquad\left. + \frac{\sqrt{m_n/p_n}r_nv_n\Big(P\left(\sum_{i=1}^{s_n}(X_i-u_n)^+ > Su_n\right)/(s_nv_n) - \theta_{sl}(S)\Big)}{1+\sqrt{p_n/m_n}(r_nv_n)^{-1}Z^c_n}\right)_{S\in[S_0,\infty)} \\
&\to \sqrt{\theta}(Z^d_S - \theta_{sl}(S)Z^c)_{S\in[S_0,\infty)} =: (Z_S)_{S\in[S_0,\infty)}
\end{aligned}$$

where in the last step we have used (4.4.31), the bias condition (SB_b) and (4.4.1). The limit random variable is a centered Gaussian process. The asserted covariance can be calculated by

$$\begin{aligned}
&Cov(Z_s,Z_t) = \theta Cov(Z^d_s - \theta_{sl}(s)Z^c, Z^d_t - \theta_{sl}(t)Z^c) \\
&= \theta Cov(Z^d_s, Z^d_t) + \theta_{sl}(s)\theta_{sl}(t)c - \theta\theta_{sl}(s)Cov(Z^d_t,Z^c) - \theta\theta_{sl}(t)Cov(Z^d_s,Z^c) \\
&= \theta_{sl}(\max(s,t)) + \theta_{sl}(s)\theta_{sl}(t)c \\
&\qquad - \theta_{sl}(s)P\Big(\sum_{j\in\mathbb{Z}}(Y_j-1)^+ > t\Big) - \theta_{sl}(t)P\Big(\sum_{j\in\mathbb{Z}}(Y_j-1)^+ > s\Big). \qquad \square
\end{aligned}$$

As before, for the proof of Theorem 4.3.3 about the asymptotic normality of the sliding blocks estimator the calculation of the covariance is outsourced.

Lemma 4.4.7. *Suppose (S1), (SP) and (AC) hold. Then it holds for all $s, t \in [S_0, \infty)$*

(i) $$\lim_{n\to\infty} \frac{1}{r_n v_n s_n^2} Cov\Big(\sum_{j=1}^{r_n} f_s((X_{n,h})_{j\le h\le j+s_n-1}), \sum_{k=1}^{r_n} f_t((X_{n,h})_{k\le h\le k+s_n-1})\Big) = \theta_{sl}(s\vee t),$$

(ii) $$\lim_{n\to\infty} \frac{1}{r_n v_n s_n} Cov\Big(\sum_{j=1}^{r_n} f_s((X_{n,h})_{j\le h\le j+s_n-1}), \sum_{i=1}^{r_n} \mathbb{1}_{\{X_i>u_n\}}\Big) = P\Big(\sum_{j\in\mathbb{Z}}(Y_j-1)^+ > s\Big).$$

Proof of Lemma 4.4.7. For the covariance of f_s and f_t for $s, t \in [S_0, \infty)$ we obtain

$$\begin{aligned}
&\frac{1}{r_n v_n s_n^2} Cov\Big(\sum_{j=1}^{r_n} \mathbb{1}_{\{\sum_{i=j}^{j+s_n-1}(X_i-u_n)^+>su_n\}}, \sum_{j=1}^{r_n} \mathbb{1}_{\{\sum_{i=j}^{j+s_n-1}(X_i-u_n)^+>tu_n\}}\Big)\\
&= \frac{1}{r_n v_n s_n^2} r_n E\left[\mathbb{1}_{\{\sum_{i=1}^{s_n}(X_i-u_n)^+>su_n\}}\mathbb{1}_{\{\sum_{i=1}^{s_n}(X_i-u_n)^+>tu_n\}}\right]\\
&\quad + \frac{r_n}{r_n v_n s_n^2}\sum_{k=1}^{r_n}\Big(1-\frac{k}{r_n}\Big)E\left[\mathbb{1}_{\{\sum_{i=1}^{s_n}(X_i-u_n)^+>su_n\}}\mathbb{1}_{\{\sum_{i=k+1}^{k+s_n}(X_i-u_n)^+>tu_n\}}\right]\\
&\quad + \frac{r_n}{r_n v_n s_n^2}\sum_{k=1}^{r_n}\Big(1-\frac{k}{r_n}\Big)E\left[\mathbb{1}_{\{\sum_{i=1}^{s_n}(X_i-u_n)^+>tu_n\}}\mathbb{1}_{\{\sum_{i=k+1}^{k+s_n}(X_i-u_n)^+>su_n\}}\right]\\
&\quad - \frac{r_n^2}{r_n v_n s_n^2} E\left[\mathbb{1}_{\{\sum_{i=1}^{s_n}(X_i-u_n)^+>su_n\}}\right]E\left[\mathbb{1}_{\{\sum_{i=1}^{s_n}(X_i-u_n)^+>tu_n\}}\right]\\
&= \frac{1}{s_n^2 v_n} P\Big(\sum_{i=1}^{s_n}(X_i-u_n)^+ > \max(s,t)u_n\Big) + o(1)\\
&\quad + \frac{1}{s_n^2 v_n}\sum_{k=1}^{s_n}\Big(1-\frac{k}{r_n}\Big)P\Big(\sum_{i=1}^{s_n}(X_i-u_n)^+ > su_n, \sum_{i=k+1}^{k+s_n}(X_i-u_n)^+ > tu_n\Big)\\
&\quad + \frac{1}{s_n^2 v_n}\sum_{k=1}^{s_n}\Big(1-\frac{k}{r_n}\Big)P\Big(\sum_{i=1}^{s_n}(X_i-u_n)^+ > tu_n, \sum_{i=k+1}^{k+s_n}(X_i-u_n)^+ > su_n\Big)\\
&\quad + \frac{1}{s_n^2 v_n}\sum_{k=s_n+1}^{r_n}\Big(1-\frac{k}{r_n}\Big)P\Big(\sum_{i=1}^{s_n}(X_i-u_n)^+ > su_n, \sum_{i=k+1}^{k+s_n}(X_i-u_n)^+ > tu_n\Big)\\
&\quad + \frac{1}{s_n^2 v_n}\sum_{k=s_n+1}^{r_n}\Big(1-\frac{k}{r_n}\Big)P\Big(\sum_{i=1}^{s_n}(X_i-u_n)^+ > tu_n, \sum_{i=k+1}^{k+s_n}(X_i-u_n)^+ > su_n\Big)\\
&=: I + II + III + IV + V + o(1).
\end{aligned}$$

We will analyze the summands separately. Note that under our conditions the conditions of Lemma 4.4.1 are satisfied, since due to (4.4.6) and the β-mixing assumption the last condition in (θ1S) is satisfied. Because of

$$P\Big(\sum_{i=1}^{s_n}(X_i-u_n)^+ > tu_n, \sum_{i=k+1}^{k+s_n}(X_i-u_n)^+ > su_n\Big) \le P\left(M_{1,s_n} > u_n, M_{k+1,k+s_n} > u_n\right)$$

for all $k \in \mathbb{Z}$, and with (4.4.11) we obtain $IV \to 0$ and likewise $V \to 0$. For I we have

$$\frac{1}{s_n}\frac{1}{s_n v_n}P\Big(\sum_{i=1}^{s_n}(X_i-u_n)^+ > \max(s,t)u_n\Big) = \mathrm{O}\Big(\frac{1}{s_n}\Big) \to 0.$$

Moreover, for parts of II and III we have

$$\begin{aligned}
&\frac{1}{s_n^2 v_n}\sum_{k=1}^{s_n}\frac{k}{r_n}P\Big(\sum_{i=1}^{s_n}(X_i-u_n)^+ > tu_n, \sum_{i=k+1}^{k+s_n}(X_i-u_n)^+ > su_n\Big)\\
&\le \frac{1}{s_n^2 v_n}\sum_{k=1}^{s_n}\frac{s_n}{r_n}P(M_{1,s_n} > u_n) = \mathrm{O}\Big(\frac{s_n^2}{r_n s_n^2 v_n}s_n v_n\Big) = \mathrm{O}\Big(\frac{s_n}{r_n}\Big) \to 0.
\end{aligned}$$

It remains to calculate

$$\begin{aligned}
&\frac{1}{s_n^2 v_n}\sum_{k=1}^{s_n}P\Big(\sum_{i=1}^{s_n}(X_i-u_n)^+ > su_n, \sum_{i=k+1}^{k+s_n}(X_i-u_n)^+ > tu_n\Big)\\
&= \frac{1}{s_n^2 v_n}\sum_{k=1}^{s_n}\sum_{j=1}^{s_n}P\Big(\sum_{i=1}^{s_n}(X_i-u_n)^+ > su_n, \sum_{i=k+1}^{k+s_n}(X_i-u_n)^+ > tu_n, X_j > u_n, M_{1,j-1} \le u_n\Big)\\
&= \frac{1}{s_n^2}\sum_{k=1}^{s_n}\sum_{j=k+1}^{s_n}P\Big(\sum_{i=1-j}^{s_n-j}(X_i-u_n)^+ > su_n, \sum_{i=k+1-j}^{k+s_n-j}(X_i-u_n)^+ > tu_n,\\
&\qquad\qquad M_{1-j,-1} \le u_n \Big| X_0 > u_n\Big)\\
&\quad + \frac{1}{s_n^2}\sum_{k=1}^{s_n}\sum_{j=1}^{k}P\Big(\sum_{i=1-j}^{s_n-j}(X_i-u_n)^+ > su_n, \sum_{i=k+1-j}^{k+s_n-j}(X_i-u_n)^+ > tu_n,\\
&\qquad\qquad M_{1-j,-1} \le u_n \Big| X_0 > u_n\Big)\\
&= II_1 + II_2.
\end{aligned}$$

This second sum converges to 0, since for some sequence t_n with $t_n \to \infty$, $t_n = \mathrm{o}(s_n)$ one has

$$\begin{aligned}
II_2 &\le \frac{1}{s_n^2}\sum_{j=1}^{s_n}\sum_{k=j+t_n}^{s_n}P\left(M_{t_n,2s_n} > u_n \mid X_0 > u_n\right) + \frac{1}{s_n^2}\sum_{j=1}^{s_n}\sum_{k=j}^{j+t_n-1}1\\
&\le \frac{s_n^2}{s_n^2}P(M_{t_n,2s_n} > u_n \mid X_0 > u_n) + \frac{s_n t_n}{s_n^2} \to 0.
\end{aligned}$$

In the last step we used that $P(M_{t_n,2s_n} > u_n \mid X_0 > u_n) \to 0$ by (SP). The remaining sum II_1 possesses the upper bound

$$\begin{aligned}
&\frac{1}{s_n^2}\sum_{k=1}^{s_n}\sum_{j=k+1+t_n}^{s_n-t_n}P\Big(\sum_{i=0}^{s_n-t_n}(X_i-u_n)^+ > su_n, \sum_{i=0}^{2s_n-t_n}(X_i-u_n)^+ > tu_n,\\
&\qquad\qquad M_{1-t_n,-1} \le u_n \Big| X_0 > u_n\Big) + \mathrm{O}\Big(\frac{t_n s_n}{s_n^2}\Big)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{s_n} \frac{s_n - 2t_n - k}{s_n^2} P\Big(\sum_{i=0}^{s_n - t_n} (X_i - u_n)^+ > su_n, \sum_{i=0}^{2s_n - t_n} (X_i - u_n)^+ > tu_n, \\
&\qquad\qquad M_{1-t_n,-1} \le u_n \Big| X_0 > u_n \Big) + \mathrm{O}\Big(\frac{t_n s_n}{s_n^2} \Big) \\
&= \frac{s_n^2 - 2t_n s_n - s_n(s_n+1)/2}{s_n^2} P\Big(\sum_{i=0}^{s_n - t_n} (X_i - u_n)^+ > su_n, \sum_{i=0}^{2s_n - t_n} (X_i - u_n)^+ > tu_n, \\
&\qquad\qquad M_{1-t_n,-1} \le u_n \Big| X_0 > u_n \Big) + \mathrm{O}\Big(\frac{t_n s_n}{s_n^2} \Big) \\
&\to \frac{1}{2} P\Big(\sum_{i=0}^{\infty} (Y_i - 1)^+ > s, \sum_{i=0}^{\infty} (Y_i - 1)^+ > t, Y^*_{-\infty,-1} \le 1 \Big) = \frac{1}{2} \theta_{sl}(\max(s,t)).
\end{aligned}
$$

The convergence holds by Proposition 4.1.3 since the considered indicator is a.s. continuous by $Y_t \overset{d}{=} Y_0 \Theta_t$, where Y_0 is independent of Θ_t and Par(α)-distributed.
On the other hand II_1 possesses by similar calculation the lower bound

$$
\begin{aligned}
&\frac{1}{s_n^2} \sum_{k=1}^{s_n} \sum_{j=k+1+t_n}^{s_n - t_n} P\Big(\sum_{i=0}^{t_n} (X_i - u_n)^+ > su_n, \sum_{i=0}^{t_n} (X_i - u_n)^+ > tu_n, \\
&\qquad\qquad M_{1-s_n,-1} \le u_n \Big| X_0 > u_n \Big) + \mathrm{O}\Big(\frac{t_n s_n}{s_n^2} \Big) \\
&= \frac{s_n^2 - 2t_n s_n - s_n(s_n+1)/2}{s_n^2} P\Big(\sum_{i=0}^{t_n} (X_i - u_n)^+ > su_n, \sum_{i=0}^{t_n} (X_i - u_n)^+ > tu_n, \\
&\qquad\qquad M_{1-s_n,-1} \le u_n \Big| X_0 > u_n \Big) + \mathrm{O}\Big(\frac{t_n s_n}{s_n^2} \Big) \\
&\to \frac{1}{2} P\Big(\sum_{i=0}^{\infty} (Y_i - 1)^+ > s, \sum_{i=0}^{\infty} (Y_i - 1)^+ > t, Y^*_{-\infty,-1} \le 1 \Big) = \frac{1}{2} \theta_{sl}(\max(s,t)).
\end{aligned}
$$

The convergence to the tail process holds due to Proposition 4.1.3. Putting things together we obtain

$$
II = II_1 + \mathrm{o}(1) \to \frac{1}{2} \theta_{sl}(\max(s,t)).
$$

Likewise $III \to \theta_{sl}(\max(s,t))/2$, since the calculations above holds for all $s,t \in [S_0, \infty)$. Thus, we have for the covariance of f_s and f_t

$$
I + II + III + IV + V \to \frac{1}{2} \theta_{sl}(\max(s,t)) + \frac{1}{2} \theta_{sl}(\max(t,s)) = \theta_{sl}(\max(s,t))
$$

which proves part (i). Next we turn to part (ii), the covariance between f_s and h. By stationarity and the similar arguments as used before we obtain

$$
\begin{aligned}
&\frac{1}{r_n v_n s_n} Cov\Big(\sum_{j=1}^{r_n} \mathbb{1}_{\{\sum_{i=j}^{j+s_n-1} (X_i - u_n)^+ > su_n\}}, \sum_{i=1}^{r_n} \mathbb{1}_{\{X_i > u_n\}} \Big) \\
&\qquad = \frac{1}{v_n s_n} P\Big(\sum_{i=0}^{s_n - 1} (X_i - u_n)^+ > su_n, X_0 > u_n \Big)
\end{aligned}
$$

$$
\begin{aligned}
&+ \frac{1}{v_n s_n} \sum_{k=1}^{r_n-1} \Big(1-\frac{k}{r_n}\Big) P\Big(\sum_{i=1}^{s_n} (X_i-u_n)^+ > su_n, X_k > u_n\Big) \\
&+ \frac{1}{v_n s_n} \sum_{k=1}^{r_n-1} \Big(1-\frac{k}{r_n}\Big) P\Big(\sum_{i=k}^{s_n+k-1} (X_i-u_n)^+ > su_n, X_0 > u_n\Big) + \mathrm{o}(1) \\
&= \mathrm{o}(1) + \frac{1}{v_n s_n} \sum_{k=1}^{s_n} P\left(\sum_{i=1}^{s_n} (X_i-u_n)^+ > su_n, X_k > u_n\right) + \mathrm{O}\Big(\frac{s_n s_n v_n}{r_n s_n v_n}\Big) \\
&+ \frac{1}{v_n s_n} \sum_{k=s_n+1}^{r_n-1} \Big(1-\frac{k}{r_n}\Big) P\left(\sum_{i=1}^{s_n} (X_i-u_n)^+ > su_n, X_k > u_n\right) \\
&+ \frac{1}{v_n s_n} \sum_{k=1}^{r_n-1} \Big(1-\frac{k}{r_n}\Big) P\left(\sum_{i=k}^{s_n+k-1} (X_i-u_n)^+ > su_n, X_0 > u_n\right) + \mathrm{o}(1) \\
&\to P\left(\sum_{j\in\mathbb{Z}} (Y_j-1)^+ > s\right) + c_{sl}^S(s).
\end{aligned}
$$

The convergence of the first summand was already proven in (4.4.30) and we define

$$
\begin{aligned}
c_{sl}^S(s) := \lim_{n\to\infty} \Bigg(&\frac{1}{v_n s_n} \sum_{k=s_n+1}^{r_n-1} \Big(1-\frac{k}{r_n}\Big) P\Big(\sum_{i=1}^{s_n} (X_i-u_n)^+ > su_n, X_k > u_n\Big) \\
&+ \frac{1}{v_n s_n} \sum_{k=1}^{r_n-1} \Big(1-\frac{k}{r_n}\Big) P\Big(\sum_{i=k}^{s_n+k-1} (X_i-u_n)^+ > su_n, X_0 > u_n\Big)\Bigg).
\end{aligned}
$$

Obviously $c_{sl}^S(s) \le c^S$ with c^S defined in (4.4.12). Thus, since the conditions of the Lemma 4.4.1 are fulfilled, this implies $c_{sl}^S(s) = 0$, which completes the proof. □

Finally, as last proof of this section we prove Theorem 4.3.3.

Proof of Theorem 4.3.3. In the first step we will prove the weak convergence

$$
\begin{pmatrix}
(\sqrt{nv_n}s_n)^{-1}(\sum_{j=1}^{n-s_n+1} \mathbb{1}_{\{\sum_{i=j}^{j+s_n-1}(X_i-u_n)^+>Su_n\}} - P(\sum_{i=1}^{s_n}(X_i-u_n)^+ > Su_n))_{S\in[S_0,\infty)} \\
\sqrt{nv_n}^{-1} \sum_{i=1}^{n-s_n+1} \Big(\mathbb{1}_{\{X_i>u_n\}} - v_n\Big)
\end{pmatrix}
\xrightarrow{w} \begin{pmatrix} (Z_S^s)_{S\in[S_0,\infty)} \\ Z^c \end{pmatrix}, \tag{4.4.32}
$$

where $((Z_S^s)_{S\in[S_0,\infty)}, Z^c)$ is a centered Gaussian process with covariance $Cov(Z_s^s, Z_t^s) = \theta_{sl}(\max(s,t))$, $Var(Z^c) = c$ (cf. (4.2.3)) and $Cov(Z_S^s, Z^c) = P\left(\sum_{j\in\mathbb{Z}}(Y_j-1)^+ > s\right)$.
We will show this convergence with Theorem 3.2.1, part (b). Note that the conditions (A), (A2) and (MX) are directly implied by (S1) (see also the proof of Proposition 4.3.1). To this end, note that $b_n(g_s) = (nv_n/p_n^s)^{1/2} s_n$, $b_n(h) = (nv_n/p_n^s)^{1/2}$ and $p_n^s = r_n v_n \theta(1+\mathrm{o}(1))$. Then the conditions readily stand in (S1). Condition (D0) is a direct consequence of the separability of the index set $[S_0,\infty)$.
For condition (3.2.4) note that $f_S(X_{n,1},...,X_{n,s_n}) \neq 0$ implies $\mathbb{1}_{\{M_{1,s_n}>u_n\}} \neq 0$ and for the latter function the condition is satisfied as shown in the proof of Proposition 4.2.6.

Moreover, we have $h(X_{n,1},...,X_{n,s_n}) = \mathbb{1}_{\{X_1>u_n\}}$ and for this function condition (3.2.4) has been verified in the proof of Proposition 4.2.3. Thus, condition (3.2.4) is satisfied. For the variance with function h one obtains

$$\frac{1}{r_n v_n} Var\left(\sum_{i=1}^{r_n} \mathbb{1}_{\{X_i>u_n\}}\right) = \frac{1}{r_n v_n} E\left[\left(\sum_{i=1}^{r_n} \mathbb{1}_{\{X_i>u_n\}}\right)^2\right] + o(1) \to c$$

by the definition of c. As shown in (4.2.3), one has $c = \sum_{k\in\mathbb{Z}} P(Y_k > 1)$. The other two covariance convergences are established in Lemma 4.4.7, which is why Condition (C) holds.

Condition (D1) can be established with exactly the same arguments as in the proof of Theorem 4.3.1, we omit the details here. Since $\{f_S : S \in [S_0,\infty)\}$ is linearly ordered, this functions form a VC(2)-class and thereby condition (D3) is satisfied. (A detailed argumentation for this is given in the proof of Proposition 4.2.11). Theorem 3.2.1 now implies the weak convergence (4.4.32).

Next we turn to the asymptotic normality of $\hat{\theta}^s_{sl,n}(S)$. Along the same lines as in the proof of Lemma 3.3.5 we obtain

$$\begin{aligned}
&\sqrt{nv_n}\left(\hat{\theta}^s_{sl,n}(S) - \theta_{sl}(S)\right)_{S\in[s_0,\infty)} \\
&= \sqrt{nv_n}\left(\frac{\sum_{j=1}^{n-s_n+1} \mathbb{1}_{\{\sum_{i=j}^{j+s_n-1}(X_i-u_n)^+ > Su_n\}}}{\sum_{j=1}^{n-s_n+1} \mathbb{1}_{\{X_j>u_n\}}} - \theta_{sl}(S)\right)_{S\in[s_0,\infty)} \\
&\xrightarrow{w} (Z^s_S - \theta_{sl}(S)Z^c)_{S\in[s_0,\infty)} =: (Z_S)_{S\in[s_0,\infty)}.
\end{aligned}$$

The convergence is implied by (4.4.32) and the bias condition (SB_b). This shows the assertion. The stated covariance can be calculated by

$$\begin{aligned}
&Cov(Z_s, Z_t) = Cov(Z^s_s - \theta_{sl}(s)Z^c, Z^s_t - \theta_{sl}(t)Z^c) \\
&= \theta_{sl}(\max(s,t)) + \theta_{sl}(s)\theta_{sl}(t)c \\
&\quad - \theta_{sl}(s)P\Big(\sum_{j\in\mathbb{Z}}(Y_j-1)^+ > t\Big) - \theta_{sl}(t)P\Big(\sum_{j\in\mathbb{Z}}(Y_j-1)^+ > s\Big).
\end{aligned}$$

□

Chapter 5

Projection based estimator for the spectral tail process

As mentioned in the introduction and due to Definition 2.1.6, the spectral tail process $(\Theta_t)_{t\in\mathbb{Z}}$ carries all information about the extremal dependence of the underlying process $(X_t)_{t\in\mathbb{Z}}$. The aim of this chapter is to derive estimators for quantities about the extreme behavior of $(X_t)_{t\in\mathbb{Z}}$, and, therefore, we want to estimate quantities depending on the distribution of the spectral tail process $(\Theta_t)_{t\in\mathbb{Z}}$. We will exemplary motivate and derive an estimator for $P(\Theta_i \in A)$ for some Borel sets A in $\mathbb{R}^d$ and some $i \in \mathbb{Z}$ in Section 5.1.
In Sections 5.2 and 5.3, the asymptotic normality of this estimator with known α and estimated index of regular variation, respectively, is shown using the limit theory developed in Section 3.2. A generalization for multiple time points is provided in Section 5.4. Section 5.5 contains an example for the asymptotic variance of the estimator and one example where all conditions of this chapters are verified. Finally, in Section 5.6, a simulation study for the new estimator is presented. All proofs of this chapter are deferred to Section 5.7.
The results of Sections 5.1, 5.2, 5.3 and 5.5 and parts of the simulations in Section 5.6 as well as the corresponding proofs have already been published in advance in cooperation with my PhD-supervisors Holger Drees and Anja Janßen in Drees et al. (2021).

5.1 Motivation and construction

Throughout this section we consider observations $X_{1-s_n}, ..., X_{n+s_n}$ of a stationary $\mathbb{R}^d$-valued time series $(X_t)_{t\in\mathbb{Z}}$, where we start with index $1 - s_n$ and end with $n + s_n$ to simplify the notation. Moreover, we assume that the time series is regularly varying such that the tail process satisfies the summability condition (SC) (cf. inequality (2.2.3)).
The definition of the spectral tail process leads to a straightforward empirical estimator

for $P(\Theta_i \in A)$ given by

$$\hat{p}^f_{n,A} := \frac{1}{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}} \sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}} \mathbb{1}_A \left(X_{t+i}/\|X_t\| \right), \tag{5.1.1}$$

where u_n denotes a suitably high threshold converging to ∞ as $n \to \infty$. This simple empirical estimator for $P(\Theta_i \in A)$ is called *forward estimator*. It was introduced in Davis et al. (2018) and is consistent, basically due to stationarity of $(X_t)_{t\in\mathbb{Z}}$. The asymptotic normality of this estimator was proven in Theorem 3.1 of the same reference. This estimator does not make use of any properties of the spectral tail process, apart from the definition.

For univariate time series and the special sets $A = (-\infty, -x)$ or $A = (x, \infty)$, $x \geq 0$, the time change formula applied with the function $f(x_0) = \mathbb{1}_{\{x_0 \in A\}}$ was used to design an alternative estimator for $P(\Theta_i \in A)$ given by

$$\hat{p}^b_{n,A} := \frac{1}{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}} \sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}} \|X_{t-i}/X_t\|^\alpha \mathbb{1}_A \left(X_t/\|X_{t-i}\| \right). \tag{5.1.2}$$

Usually the index of regular variation α is unknown. One could replace α by an appropriate estimator $\hat{\alpha}_n$, e.g. as defined in (5.1.5) below. In this case, we denote the estimator by $\hat{\hat{p}}^b_{n,A}$. The estimator $\hat{p}^b_{n,A}$ (or $\hat{\hat{p}}^b_{n,A}$, respectively) is called *backward estimator* and was introduced in a Markovian setting for $i = 1$ by Drees et al. (2015) and in a general univariate setting for arbitrary i by Davis et al. (2018). By applying the time change formula for one single lag i, this estimator only makes use of small part of this structural property of the distribution of $(\Theta_t)_{t\in\mathbb{Z}}$, since the time change formula holds for all $i \in \mathbb{Z}$ (cf. Definition 2.2.2). Simulation studies have shown that the backward estimator can have a smaller root mean square error (RMSE) than the forward estimator for different classes of models and in particular for larger values of x, see Davis et al. (2018) and Drees et al. (2015).

Our aim is to construct an estimator for the distribution $P^{(\Theta_t)_{t\in\mathbb{Z}}}$ of the spectral tail process $(\Theta_t)_{t\in\mathbb{Z}}$ or $P(\Theta_i \in A)$ for some Borel sets A, respectively, which is not only an empirical version of the distribution, but also makes use of the whole special structure of such a distribution formalized by the time change formula.

Under the summability condition (SC), a time series $\Theta = (\Theta_t)_{t\in\mathbb{Z}}$ satisfies the TCF if and only if the distribution P^Θ of the time series is invariant under the RS transformation (cf. Theorem 2.2.5). Exactly this RS transformation shall be used in the estimation approach considered here. Therefore, the condition (SC) is always assumed in this section, below it follows from stronger conditions (namely (PP)). With this RS-transformation the whole structure of the time change formula is used, which should improve the estimation.

In order to apply the RS-transformation here, we first redefine it as the transformation of probability measures. Under the condition (SC), the distribution P^Θ of a spectral tail process is a probability measure on

$$l_\alpha = \Big\{ z \in (\mathbb{R}^d)^{\mathbb{Z}} \mid 0 \leq \|z\|_\alpha^\alpha < \infty \Big\},$$

where $\alpha > 0$ and $\|z\|_\alpha^\alpha = \sum_{t\in\mathbb{Z}} \|z_t\|^\alpha$ as defined in (2.2.4). In particular $(\Theta_t)_{t\in\mathbb{Z}} \in l_\alpha$ a.s. if (SC) holds. In this chapter, we consider some functions defined on l_α equipped with the supremum norm. For arbitrary $n \in \mathbb{N}$ the space $\mathbb{R}^{2n+1}$ is embedded in l_α by the mapping $(\mathbb{R}^d)^{2n+1} \ni (z_t)_{|t|\leq n} \mapsto (z_t)_{t\in\mathbb{Z}} \in l_\alpha$ with $z_t = 0$ for $|t| > n$. Note that (2.1.2) ensures that the realizations of the spectral tail process a.s. belong to l_α (see above or Remark 2.3 of Janßen (2019)).
The RS-transformation was introduced in Definition 2.2.4 as a transformation of the random variable $(\Theta_t)_{t\in\mathbb{Z}}$, where only equality in distribution was required. However, the RS-transformation can also be applied directly to the distribution without changing the meaning. To this end, we define the RS-transformation Q^{RS} of a probability measure Q on l_α by

$$Q^{RS}(A) := \sum_{k\in\mathbb{Z}} \int \frac{\|z_k\|^\alpha}{\|z\|_\alpha^\alpha} \mathbb{1}_A \left(\left(\frac{z_{s+k}}{\|z_k\|} \right)_{s\in\mathbb{Z}} \right) Q(dz)$$

for all $A \in \mathcal{B}(l_\alpha)$, where the argument in the integral is 0 if $\|z\|_\alpha = 0$.
The following lemma shows that RS-transformed measures are invariant under the RS-transformation. The application of the RS-transformation is thus a projection of probability measures on l_α on the subset of RS-invariant probability measures, which is the set of admissible distributions for a spectral tail process. As shown in the next lemma, this is a projection in the algebraic sense $p \circ p = p$ for some map p.

Lemma 5.1.1. *For all probability measures Q on l_α it holds $(Q^{RS})^{RS} = Q^{RS}$.*

The distribution of a spectral tail process satisfies the TCF, which under condition (SC) is equivalent to the fact that the distribution is invariant under the RS-transformation. Conversely, each measure in the set of RS-invariant probability measures on l_α which satisfies the condition (SC) is the distribution of a spectral tail process. This is because each process satisfying the TCF is the spectral tail process of some max stable time series, see Janßen (2019), Theorem 3.2, or Planinić and Soulier (2018), Theorem 5.1. Thus, an estimated distribution which is invariant under the RS-transformation is automatically the distribution of some spectral tail process. Now, a reasonable and desirable goal is that the estimated measure is invariant under the RS-transformation, so that it satisfies the crucial property of a distribution of some spectral tail process.
The estimation idea is to take an empirical version of the probability measure $P^{(\Theta_t)_{t\in\mathbb{Z}}}$ and apply the RS-transformation to it. According to the previous lemma, this ensures that the estimated measure is invariant under the RS-transformation and thereby a random variable with this estimated distribution satisfies the TCF. With this projection we make use of the whole structure given by the TCF and ensure that the estimated object has all essential properties of the distribution of a spectral tail process.

The idea to project some initial estimator on a subset of admissible quantities was already used in the literature. A projection estimator is defined most commonly as best approximation of an initial estimator w.r.t. some norm, see e.g. Fils-Villetard et al. (2008) or Mammen et al. (2001). However, we define the projection based estimator by ensuring that the transformation induces only a random shift in time and scale.
The empirical counterpart of $P^{(\Theta_t)_{|t|\le s_n}}$ given the observations $X_{1-s_n},...,X_{n+s_n}$, results from Definition 2.1.6 and can be defined as

$$\hat{P_n^\Theta} := \frac{1}{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}} \sum_{t=1}^n \delta_{\left((X_{t+s}/\|X_t\|)\mathbb{1}_{\{|s|\le s_n\}}\right)_{s\in\mathbb{Z}}} \mathbb{1}_{\{\|X_t\|>u_n\}},$$

where s_n, $n \in \mathbb{N}$, is an intermediary sequence (i.e $s_n \to \infty$ and $s_n = o(n)$) and the Dirac-measure with point mass 1 in $x \in l_\alpha$ is denoted by δ_x. The standardization in the denominator ensures that the estimator itself is a probability measure. This is more general representation for the forward estimator as considered in in (5.1.1). Note that we trimmed $(\Theta_t)_{t\in\mathbb{Z}}$ to finite length $2s_n+1$, since we only have finitely many observations. The constant s_n determines the length of the time interval for which Θ is estimated. The condition $s_n = o(n)$ is important, since one cannot expect a good estimator for $P^{(\Theta_t)_{|t|\le n+2s_n+1}}$ based on observations $X_{1-s_n},...,X_{n+s_n}$, since e.g. for the distribution of Θ_{n+2s_n+1} the estimation would only be based on a single pair of observations (X_{1-s_n}, X_{n+s_n}).
For all $t \in \{1,...,n\}$ the numerator in the Dirac-measure in $\hat{P_n^\Theta}$ is set to 0 outside of $\{t-s_n,...,t+s_n\}$. At first glance this is a somewhat arbitrary choice, however there are two reasons for this definition: First, we only estimate the distribution $P^{(\Theta_t)_{-s_n\le t\le s_n}}$, i.e. the remaining observations are not important for this estimation problem. Second, this choice ensures that the summands for $t=1$ and $t=2s_n+2$ do not depend on any shared observations. This is important for the technical analysis with sliding blocks methods which will be used below.
Application of the RS-transformation to $\hat{P_n^\Theta}$ yields for all Borel sets A in l_α

$$\begin{aligned}
\hat{P_n^\Theta}^{RS}(A) &= \sum_{k\in\mathbb{Z}} \int \frac{\|z_k\|^\alpha}{\|z\|_\alpha^\alpha} \mathbb{1}_A\Big(\Big(\frac{z_{s+k}}{\|z_k\|}\Big)_{s\in\mathbb{Z}}\Big) \hat{P^\Theta}_n(dz) \\
&= \frac{\sum_{t=1}^n \left(\mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h\in\mathbb{Z}} \frac{\|\frac{X_{t+h}}{\|X_t\|}\mathbb{1}_{\{|h|\le s_n\}}\|^\alpha}{\|\left(\frac{X_{t+k}}{\|X_t\|}\mathbb{1}_{\{|k|\le s_n\}}\right)_{k\in\mathbb{Z}}\|_\alpha^\alpha} \mathbb{1}_A\left(\left(\frac{\frac{X_{t+s+h}}{\|X_t\|}\mathbb{1}_{\{|s+h|\le s_n\}}}{\|\frac{X_{t+h}}{\|X_t\|}\mathbb{1}_{\{|h|\le s_n\}}\|}\right)_{s\in\mathbb{Z}}\right)\right)}{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}} \\
&= \frac{\sum_{t=1}^n \left(\mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{|h|\le s_n} \frac{\|X_{t+h}\|^\alpha}{\sum_{|k|\le s_n}\|X_{t+k}\|^\alpha} \mathbb{1}_A\left(\left(\frac{X_{t+s+h}\mathbb{1}_{\{|s+h|\le s_n\}}}{\|X_{t+h}\|}\right)_{s\in\mathbb{Z}}\right)\right)}{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}}.
\end{aligned}$$

This results in the estimator for the distribution $P^{(\Theta_t)_{|t|\le s_n}}$ as

$$\hat{P_n^\Theta}^{RS} = \frac{1}{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}}$$

$$\times \sum_{t=1}^{n} \left(\mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h=-s_n}^{s_n} \frac{\|X_{t+h}\|^\alpha}{\sum_{k=-s_n}^{s_n} \|X_{t+k}\|^\alpha} \delta_{\left(\frac{X_{t+h+s}}{\|X_{t+h}\|} \mathbb{1}_{\{|h+s|\le s_n\}} \right)_{s\in\mathbb{Z}}} \right). \tag{5.1.3}$$

We will call this estimator a *projection based estimator*, due to its derivation. By construction, this measure fulfills all essential structural properties of a spectral tail process and is itself the distribution of some spectral tail process.

One important feature of this projection based estimator is that non-extreme observations, i.e. observations X_t with $\|X_t\| < u_n$, have an impact on the estimation. This makes the asymptotic analysis a bit more challenging. The idea that non-extreme observations should have an impact on estimators for extreme parts of a time series already occurred in the literature, e.g. in Sun and Samorodnitsky (2019). In the construction here, the non-extreme observations occur naturally by the RS-projection in the derivation.

From the estimator for the whole distribution, an estimator for special probabilities for a marginal distribution of the spectral tail process can be derived. For this purpose, the estimated measure $\hat{P_n^\Theta}^{RS}$ or the appearing Dirac measures, respectively, are evaluated at the corresponding set. This way we obtain the following projection based estimator $\hat{p}_{n,A}$ for $P(\Theta_i \in A)$, for a Borel set $A \in \mathcal{B}(\mathbb{R}^d)$ and for some $i \in \mathbb{Z}$:

$$\begin{aligned} \hat{p}_{n,A} :&= \hat{P_n^\Theta}^{RS}(\Theta_i \in A) \\ &= \frac{1}{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}} \sum_{t=1}^{n} \mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h=-s_n}^{s_n} \frac{\|X_{t+h}\|^\alpha}{\sum_{k=-s_n}^{s_n} \|X_{t+k}\|^\alpha} \\ &\qquad \times \left(\mathbb{1}_A\left(\frac{X_{t+h+i}}{\|X_{t+h}\|} \right) \mathbb{1}_{\{h\in H_{n,i}\}} + \mathbb{1}_A(0) \mathbb{1}_{\{h\in H_{n,i}^C\}} \right), \end{aligned} \tag{5.1.4}$$

where $H_{n,i} := \{(-s_n - i) \vee (-s_n), \ldots, (s_n - i) \wedge s_n\}$ and $H_{n,i}^C := \{-s_n, \ldots, s_n\} \setminus H_{n,i}$. The set $H_{n,i}^C$ equals $\{s_n - i + 1, \ldots, s_n\}$ for $i > 0$, $\{-s_n, \ldots, -s_n - i - 1\}$ for $i < 0$, and is the empty set for $i = 0$. Here, $|i| \le s_n$ is a necessary restriction because the estimator $\hat{P_n^\Theta}^{RS}$ estimates only the distribution of $(\Theta_t)_{|t|\le s_n}$.

For this estimator $\hat{p}_{n,A}$ and the analysis of its asymptotics, we assume the index α of the regular variation to be known. The asymptotic normality of this estimator will be shown in Section 5.2.1 and the uniform asymptotic normality for different sets A in Section 5.2.2. In general α is unknown, so it must be replaced by a suitable estimator. One possible estimator is the Hill type estimator

$$\hat{\alpha}_n := \frac{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}}{\sum_{t=1}^n \log(\|X_t\|/u_n) \mathbb{1}_{\{\|X_t\|>u_n\}}} \tag{5.1.5}$$

and we denote the projection based estimator with estimated α by

$$\hat{\hat{p}}_{n,A} := \frac{1}{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}} \sum_{t=1}^{n} \mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h=-s_n}^{s_n} \frac{\|X_{t+h}\|^{\hat{\alpha}_n}}{\sum_{k=-s_n}^{s_n} \|X_{t+k}\|^{\hat{\alpha}_n}}$$

$$\times \left(\mathbb{1}_A \left(\frac{X_{t+h+i}}{\|X_{t+h}\|} \right) \mathbb{1}_{\{h \in H_{n,i}\}} + \mathbb{1}_A(0) \mathbb{1}_{\{h \in H_{n,i}^C\}} \right). \tag{5.1.6}$$

The asymptotic normality of this estimator is established in Section 5.3. The replacement of α with $\hat{\alpha}_n$ makes the asymptotic analysis much more challenging and has an impact on the asymptotic distribution of the suitably standardized estimation error.
The projection based estimator $\hat{P_n^{\Theta}}^{RS}$ in (5.1.3) for the distribution $P^{(\Theta_t)_{|t|\leq s_n}}$ allows not only the derivation of estimators for $P(\Theta_i \in A)$ for some single $i \in \mathbb{Z}$. It also allows the construction of estimators for probabilities of events affected by multiple time points $i_1, ..., i_M \in \mathbb{Z}$, $M \in \mathbb{N}$. Such a generalization is considered in Section 5.4.
The finite sample performance of this estimator is considered in a simulation study presented in Section 5.6. We will show that the projection based estimator not only has the advantage that the estimated distribution is itself the distribution of some spectral tail process, but we will also demonstrate that this estimator performs reasonably well in the simulations.

5.2 Asymptotic behavior

The aim of the next two sections is the development of the uniform asymptotic normality of $(\hat{p}_{n,A})_{A\in\mathcal{A}}$ and $(\hat{\hat{p}}_{n,A})_{A\in\mathcal{A}}$ for some family $\mathcal{A} \subset \mathcal{B}(\mathbb{R}^d)$, where $\mathcal{B}(\mathbb{R}^d)$ denotes the Borel sets of $\mathbb{R}^d$. We will start with the asymptotic normality of the estimator $(\hat{p}_{n,A})_{A\in\mathcal{A}}$ with known α. For this, we start with the proof of the finite dimensional marginal distributions (fidis) convergence for the empirical process associated to the suitable standardized estimator $\hat{p}_{n,A}$ before we continue with the process convergence.

5.2.1 Asymptotic behavior of the fidis of the estimator

Most common extreme value statistics depend only on extreme observations of a time series. However, due to the construction of the estimator $\hat{p}_{n,A}$, even non-extreme observations, i.e. observations X_t with $\|X_t\| < u_n$, are included in the estimator if an observation in the neighborhood is extreme. One has to take care of this feature for the asymptotic analysis and non extreme observations may not simply be set to 0. Moreover, the numerator of the estimator $\hat{p}_{n,A}$ is a sliding blocks sum. Both properties make the asymptotic analysis challenging and special tools must be used. Here we will apply the theory developed in Section 3.2. The setting of the existing literature is not suitable for this asymptotic analysis.
For the analysis of the asymptotics, a few conditions will be stated in the course of this section which will carry the leading letter (P) for orientation.
Since we want to estimate the spectral tail process, we assume regular variation of the time series $(X_t)_{t\in\mathbb{Z}}$. For the application of the sliding blocks limit theorem (Theorem 3.2.1), some basic assumptions on occurring sequences and on the mixing behavior of the

time series $(X_t)_{t\in\mathbb{Z}}$ are necessary. For the rest of this section we fix some sequence $s_n \in \mathbb{N}$ and $u_n > 0$, $n \in \mathbb{N}$ and thereby $v_n := P(\|X_0\| > u_n)$. The assumptions on the sequences are summarized by condition (P0):

(PR) $(X_t)_{t\in\mathbb{Z}}$ is a regularly varying time series with index α, tail process $Y = (Y_t)_{t\in\mathbb{Z}}$ and spectral tail process $\Theta = (\Theta_t)_{t\in\mathbb{Z}}$.

(P0) (i) There exist sequences $r_n, l_n, s_n \in \mathbb{N}$, $n \in \mathbb{N}$, such that $s_n \le l_n = \mathrm{o}(r_n)$, $s_n \to \infty$, $r_n = \mathrm{o}(\sqrt{nv_n})$, $nv_n \to \infty$, $r_n v_n \to 0$.

(ii) The time series $(X_t)_{t\in\mathbb{Z}}$ satisfies the β-mixing condition $(n/r_n)\beta^X_{n,l_n} \to 0$.

In particular, one has $r_n = \mathrm{o}(n)$. Here, the β-mixing coefficient is defined in (3.1.1) with the triangular scheme $(X_t/u_n)_{1\le t\le n}$ for all $n \in \mathbb{N}$.
Note that the β-mixing condition is satisfied for all sequences l_n' with $l_n \le l_n' = \mathrm{o}(r_n)$ if l_n satisfies the assumptions in (P0). Hence, w.l.o.g. we may assume that, for any fixed $k \in \mathbb{N}$, $\beta_{n,l_n-ks_n} = o(r_n/n)$ holds. To this end, note that if (P0) is satisfied for l_n, then it is satisfied for $l_n' := l_n + ks_n$ for which $\beta_{n,l_n'-ks_n} = \beta_{n,l_n} = \mathrm{o}(r_n/n)$ holds, i.e. we can always switch to the sequence l_n'. The β-mixing condition ensures that in the definition of the numerator of $\hat{p}_{n,A}$ the summands whose indexes differ by at least $l_n - s_n$ are almost independent.
Denote for this whole chapter

$$X_{n,t} := X_t/u_n$$

and simplify the notation by defining the blocks

$$W_{n,t} := ((X_{n,t+h})_{|h|\le s_n}).$$

Define for some Borel set A the function $g_A : l_\alpha \to [0,1]$ by

$$g_A((y_h)_{h\in\mathbb{Z}}) := \mathbb{1}_{\{\|y_0\|>1\}} \sum_{h\in\mathbb{Z}} \frac{\|y_h\|^\alpha}{\sum_{k\in\mathbb{Z}} \|y_k\|^\alpha} \mathbb{1}_A\Big(\frac{y_{h+i}}{\|y_h\|}\Big),$$

which obviously satisfies $0 \le g_A \le 1$. In particular, one has $g_{\mathbb{R}^d}((y_h)_{h\in\mathbb{Z}}) = \mathbb{1}_{\{\|y_0\|>1\}}$. Using this function we define the statistic

$$T_{n,A} := \sum_{t=1}^{n} g_A((X_{n,t+h})_{|h|\le s_n}) = \sum_{t=1}^{n} g_A(W_{n,t}) \tag{5.2.1}$$

for all A in some family of sets $\mathcal{A}$ and for $A = \mathbb{R}^d$. Thus, inserting this function in the definition of $\hat{p}_{n,A}$ in (5.1.4) yields

$$\hat{p}_{n,A} = \frac{\sum_{t=1}^n g_A(W_{n,t})}{\sum_{t=1}^n g_{\mathbb{R}^d}(W_{n,t})} = \frac{T_{n,A}}{T_{n,\mathbb{R}^d}}.$$

(Recall the usual embedding of $(y_t)_{|t|\le s} \in (\mathbb{R}^d)^{2s+1}$ in l_α by defining $y_t := 0$ for $|t| > s$.)

Hence, to determine the asymptotics of the estimator $\hat{p}_{n,A}$, we will first need the asymptotics of $T_{n,A}$ for $A \in \mathcal{A}$. Since the denominator of $\hat{p}_{n,A}$ equals $T_{n,\mathbb{R}^d}$ we will always assume $\mathbb{R}^d \in \mathcal{A}$. Using the asymptotic behavior for the suitable standardized statistic $T_{n,A}$ one can derive the asymptotic behavior of the suitable standardized estimators $\hat{p}_{n,A}$ by a continuous mapping argument.
The asymptotic normality of the standardized statistic $T_{n,A}$ will be derived with the theory from Section 3.2, in particular Theorem 3.2.1. The empirical process associated to the statistic $T_{n,A}$ for which this theorem can make a statement is called $(Z_n(A))_{A\in\mathcal{A}}$ and defined by

$$Z_n(A) := \frac{1}{\sqrt{nv_n}}\left(T_{n,A} - E[T_{n,A}]\right) = \frac{1}{\sqrt{nv_n}}\left(\sum_{t=1}^{n} g_A(W_{n,t}) - E[g_A(W_{n,0})]\right).$$

To establish the convergence of the fidis of $(Z_n(A))_{A\in\mathcal{A}}$ to a centered Gaussian process, three additional conditions will be used. The first condition (PP) is comparable to condition (θP) in Section 4.2 and is of the same type as condition (C) from Drees et al. (2015). It ensures that extreme observations are sufficiently independent for large lags and controls the cluster size of extreme events, i.e. how many extreme observations occur subsequently in time. In particular, we will show that it implies a bound on the second moment of the cluster size. This condition only refers to extremal dependency, in contrast to the β-mixing assumption in (P0), which is weaker (and not sufficient for our results), but restricts the whole time series. Condition (PT) also controls the cluster size, it enables us to truncate clusters to finite lengths in our asymptotic analysis. The condition (PC) is a continuity condition that ensures that functions considered later on are P^Y-a.s. continuous. This condition guarantees that $p_A := P(\Theta_i \in A)$ can be calculated as a limit using the definition of the spectral tail process.

(PP) For all $n \in \mathbb{N}$, for all $k \in \{1, ..., r_n\}$ and for all $c \in (0, 1]$ there exist

$$e_{n,c}(k) \geq P\left(\|X_k\| > u_n c \mid \|X_0\| > u_n c\right),$$

such that $\lim_{n\to\infty} \sum_{k=1}^{r_n} e_{n,c}(k) = \sum_{k=1}^{\infty} \lim_{n\to\infty} e_{n,c}(k) < \infty$. Denote $e_n(k) = e_{n,1}(k)$ as shorthand.

(PT) $\lim_{m\to\infty} \limsup_{n\to\infty} E\left[\frac{\sum_{m<|h|\leq s_n} \|X_{h+j}\|^\alpha \mathbb{1}_{\{\|X_{h+j}\|\leq \xi u_n\}}}{\sum_{|h|\leq s_n} \|X_{h+j}\|^\alpha} \,\Big|\, \|X_0\| > u_n\right] = 0$ for some $\xi > 0$ and all $j \in \mathbb{Z}$.

(PC) $P(\Theta_i \in \partial A) = 0$ for all $A \in \mathcal{A}$.

If condition (PT) is satisfied for some $\xi > 0$, it obviously holds also for all $0 < \tilde{\xi} < \xi$, since the use of $\tilde{\xi}$ instead of ξ only reduces the indicator in the expectation. Therefore, if (PT) holds, one can assume w.l.o.g. that it holds for some $\xi \in (0, 1)$.

These conditions in particular ensure that our statistic $(nv_n)^{-1}T_{n,A}$ is asymptotically unbiased. This is a first desirable feature of the statistic.

Proposition 5.2.1. *Suppose (PR), (P0), (PT) and (PC) hold. Then*

$$\frac{1}{nv_n}E\left[T_{n,A}\right] = E\left[g_A(W_{n,0}) \mid \|X_0\| > u_n\right] \to P(\Theta_i \in A). \tag{5.2.2}$$

Before we prove the asymptotic normality of the fidis of $(Z_n(A))_{A\in\mathcal{A}}$, we state some useful lemmas concerning these conditions. The first lemma states that Condition (PP) implies the anti-clustering condition (AC), see (2.1.2), which will be used in the subsequent proofs.

Lemma 5.2.2. *Suppose conditions (PR) and (PP) are satisfied. Then the anti-clustering condition (AC) is satisfied for all $c \in (0,\infty)$.*

Basrak and Segers (2009), Proposition 4.2, have shown that the anti-clustering condition (AC) (or (2.1.2)) in particular implies the summability condition (SC). There, another sequence u_n was used, but the proof remains unchanged (this holds also true for the proofs of Segers (2005), Section 2, which are used there). Thus, the general assumption (SC), needed to define the RS-transformation, holds if (PR), (P0) (i) and (PP) are satisfied.
While Condition (PC) is originally stated in a way that seems quite natural for estimating the distribution of Θ_i, a reformulation in terms of the tail process will be more useful in the proofs.

Lemma 5.2.3. *Suppose $(\Theta_t)_{t\in\mathbb{Z}}$ is a spectral tail process with corresponding tail process $(Y_t)_{t\in\mathbb{Z}}$. Then, condition (PC) is satisfied if and only if*

$$P\left(\exists t \in \mathbb{Z} : \frac{Y_{t+i}}{\|Y_t\|} \in \partial A, \|Y_t\| > 0\right) = 0 \qquad\qquad \forall\, A \in \mathcal{A}.$$

A crucial part for the application of Theorem 3.2.1, which will be used in the proofs of the asymptotic normality, is the verification of condition (C) - the convergence of the standardized covariance. To start the analysis of the asymptotic behavior of $(Z_n(A))_{A\in\mathcal{A}}$, this limit behavior of the covariance function is considered in the next lemma. In particular, the limit value of the covariances can be expressed by the tail process $(Y_t)_{t\in\mathbb{Z}}$ or the spectral tail process $(\Theta_t)_{t\in\mathbb{Z}}$ and α.

Lemma 5.2.4. *Suppose the conditions (PR), (P0), (PP), (PT) and (PC) are satisfied. Then,*

$$\frac{1}{r_n v_n}Cov\Big(\sum_{t=1}^{r_n} g_A(W_{n,t}), \sum_{t=1}^{r_n} g_B(W_{n,t})\Big) \to c(A,B)$$

for all $A, B \in \mathcal{A}$, with

$$c(A,B) := \sum_{j\in\mathbb{Z}} E\left[\mathbb{1}_{\{\|Y_j\|>1\}}\Big(\sum_{h\in\mathbb{Z}} \frac{\|Y_h\|^\alpha}{\|Y\|_\alpha^\alpha}\mathbb{1}_A\Big(\frac{Y_{h+i}}{\|Y_h\|}\Big)\Big)\Big(\sum_{l\in\mathbb{Z}} \frac{\|Y_l\|^\alpha}{\|Y\|_\alpha^\alpha}\mathbb{1}_B\Big(\frac{Y_{l+i}}{\|Y_l\|}\Big)\Big)\right]$$

$$= \sum_{j\in\mathbb{Z}} E\Big[(\|\Theta_j\|^\alpha \wedge 1)\Big(\sum_{h\in\mathbb{Z}} \frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha}\mathbb{1}_A\Big(\frac{\Theta_{h+i}}{\|\Theta_h\|}\Big)\Big)\Big(\sum_{l\in\mathbb{Z}} \frac{\|\Theta_l\|^\alpha}{\|\Theta\|_\alpha^\alpha}\mathbb{1}_B\Big(\frac{\Theta_{l+i}}{\|\Theta_l\|}\Big)\Big)\Big].$$

This limit is finite.

Remark. The covariance in Lemma 5.2.4 has two further representations:

$$\begin{aligned} c(A,B) &= E\Big[\Big(\sum_{h\in\mathbb{Z}} \frac{\|Y_h\|^\alpha}{\|Y\|_\alpha^\alpha}\mathbb{1}_A\Big(\frac{Y_{h+i}}{\|Y_h\|}\Big)\Big)\Big(\sum_{l\in\mathbb{Z}} \frac{\|Y_l\|^\alpha}{\|Y\|_\alpha^\alpha}\mathbb{1}_B\Big(\frac{Y_{l+i}}{\|Y_l\|}\Big)\Big)\Big] \\ &\quad + 2\sum_{j\in\mathbb{N}} E\Big[\mathbb{1}_{\{\|Y_j\|>1\}}\Big(\sum_{h\in\mathbb{Z}} \frac{\|Y_h\|^\alpha}{\|Y\|_\alpha^\alpha}\mathbb{1}_A\Big(\frac{Y_{h+i}}{\|Y_h\|}\Big)\Big)\Big(\sum_{l\in\mathbb{Z}} \frac{\|Y_l\|^\alpha}{\|Y\|_\alpha^\alpha}\mathbb{1}_B\Big(\frac{Y_{l+i}}{\|Y_l\|}\Big)\Big)\Big] \\ &= E\Big[\mathbb{1}_A(\Theta_i)\Big(\sum_{l\in\mathbb{Z}} \frac{\|\Theta_l\|^\alpha}{\|\Theta\|_\alpha^\alpha}\mathbb{1}_B\Big(\frac{\Theta_{l+i}}{\|\Theta_l\|}\Big)\Big)\Big] \\ &\quad + 2\sum_{j\in\mathbb{N}} E\Big[(\|\Theta_j\|^\alpha \wedge 1)\Big(\sum_{h\in\mathbb{Z}} \frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha}\mathbb{1}_A\Big(\frac{\Theta_{h+i}}{\|\Theta_h\|}\Big)\Big)\Big(\sum_{l\in\mathbb{Z}} \frac{\|\Theta_l\|^\alpha}{\|\Theta\|_\alpha^\alpha}\mathbb{1}_B\Big(\frac{\Theta_{l+i}}{\|\Theta_l\|}\Big)\Big)\Big]. \end{aligned} \tag{5.2.3}$$

The proof of these representations are given in Section 5.7.2. In these representations of $c(A,B)$, the respective sums outside the expectation are only over $\mathbb{N}$, but the sum inside the expectations still depends on $\mathbb{Z}$. However, this representation could be advantageous if the inner sum equals 1 a.s. For example it holds

$$c(\mathbb{R}^d,\mathbb{R}^d) = \sum_{j\in\mathbb{Z}} E\left[(\|\Theta_j\|^\alpha \wedge 1)\right] = 1 + 2\sum_{j\in\mathbb{N}} E\left[(\|\Theta_j\|^\alpha \wedge 1)\right] < \infty.$$

In this case, the second representation is often easier to calculate, since the forward tail process is often easier to calculate than the backward tail process.
The representation (5.2.3) is less compact compared to the representation of $c(A,B)$ in Lemma 5.2.4. Most remarkable is, that direct calculations do not obviously lead to this alternative representation, but the proof of (5.2.3) shows the alternative representation with only a simple change in one argument. ◇

With these preparations, we next establish the fidis convergence of $(Z_n(A))_{A\in\mathcal{A}}$ to the fidis of some centered Gaussian process. With this proposition we establish the asymptotic behavior of the suitable standardized statistic $T_{n,A}$, $A \in \mathcal{A}$, which was defined in (5.2.1).

Proposition 5.2.5. *Suppose the conditions (PR), (P0), (PP), (PT) and (PC) are satisfied and $\mathbb{R}^d \in \mathcal{A}$. Then the fidis of the empirical process $(Z_n(A))_{A\in\mathcal{A}}$ converge weakly to the fidis of a centered Gaussian process $(Z(A))_{A\in\mathcal{A}}$ with covariance function c defined in Lemma 5.2.4.*

So far, we have shown the fidi convergence of the standardized statistic $T_{n,A}$. This result suffices to analyze the estimator $\hat{p}_{n,A}$ for finitely many sets A from $\mathcal{A}$. In the next theorem we state the asymptotic normality of finite families of the standardized estimator $\hat{p}_{n,A}$.

The proof is basically a continuous mapping argument, using the representation $\hat{p}_{n,A} = T_{n,A}/T_{n,\mathbb{R}^d}$. We define the empirical process $(Z_n^{(p)}(A))_{A\in\mathcal{A}}$ associated with $\hat{p}_{n,A}$ by

$$Z_n^{(p)}(A) := \sqrt{nv_n}\left(\hat{p}_{n,A} - E\left[g_A(W_{n,0}) \mid \|X_0\| > u_n\right]\right).$$

Theorem 5.2.6. *Suppose the conditions (PR), (P0), (PP), (PT) and (PC) are satisfied. Denote $p_A := P(\Theta_i \in A)$ for $A \in \mathcal{A}$. Then, the fidis of $Z_n^{(p)}$ converge weakly to the fidis of $Z^{pb}(A) := Z(A) - p_A Z(\mathbb{R}^d)$, where $(Z(A))_{A\in\mathcal{A}}$ denotes the Gaussian process defined in Proposition 5.2.5. For $A, B \in \mathcal{A}$ the covariance of Z^{pb} is given by*

$$\begin{aligned} &c^{pb}(A,B) \\ &= \sum_{j\in\mathbb{Z}} E\left[(\|\Theta_j\|^\alpha \wedge 1)\left(p_B - \sum_{h\in\mathbb{Z}} \frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbb{1}_B\left(\frac{\Theta_{h+i}}{\|\Theta_h\|}\right)\right)\left(p_A - \sum_{l\in\mathbb{Z}} \frac{\|\Theta_l\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbb{1}_A\left(\frac{\Theta_{l+i}}{\|\Theta_l\|}\right)\right)\right]. \end{aligned} \tag{5.2.4}$$

If in addition

$$E\left[g_A(W_{n,0}) \mid \|X_0\| > u_n\right] - P(\Theta_i \in A) = o((nv_n)^{-1/2}) \tag{5.2.5}$$

holds for all $A \in \mathcal{A}$, then

$$(Z_n^{pb}(A))_{A\in\tilde{\mathcal{A}}} := \sqrt{nv_n}\left(\hat{p}_{n,A} - P(\Theta_i \in A)\right)_{A\in\tilde{\mathcal{A}}} \to (Z^{pb}(A))_{A\in\tilde{\mathcal{A}}}$$

weakly for all finite subsets $\tilde{\mathcal{A}} \subset \mathcal{A}$.

Note that due to Proposition 5.2.1, condition (5.2.5) is only a condition on the rate of convergence of the bias.

Remark. The calculated covariance has a second representation, similar to the second representation of $c(A,B)$ as stated in a remark after Lemma 5.2.4. Using the representation (5.2.3) of $c(A,B)$, the same calculations as in the proof of Theorem 5.2.6 yield the following representation of $c^{pb}(A,B)$:

$$\begin{aligned} c^{pb}(A,B) &= E\left[\mathbb{1}_{\{\Theta_i\in A\}}\left(\sum_{l\in\mathbb{Z}} \frac{\|\Theta_l\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbb{1}_B\left(\frac{\Theta_{l+i}}{\|\Theta_l\|}\right)\right)\right] - p_A p_B \\ &\quad + 2\sum_{j\in\mathbb{N}} E\left[(\|\Theta_j\|^\alpha \wedge 1)\left(p_B - \sum_{h\in\mathbb{Z}} \frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbb{1}_B\left(\frac{\Theta_{h+i}}{\|\Theta_h\|}\right)\right)\left(p_A - \sum_{l\in\mathbb{Z}} \frac{\|\Theta_l\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbb{1}_A\left(\frac{\Theta_{l+i}}{\|\Theta_l\|}\right)\right)\right]. \end{aligned}$$

◇

Thus, we have proven that the estimator $\hat{p}_{n,A}$ centered with $P(\Theta_i \in A)$ converges to a Gaussian distribution with rate $\sqrt{nv_n}$, for finite many sets $A \in \mathcal{A}$. The convergence rate $\sqrt{nv_n}$ is the typical rate for extreme value statistics, it is the square root of the expected number of extreme observations included in the statistics. In particular, this result implies

that the new projection based estimator is consistent for $P(\Theta_i \in A)$.
However, to estimate $P(\Theta_i \in A)$ for a larger family of sets $\mathcal{A}$, we need uniform convergence over a larger (infinite) index set $\mathcal{A}$. In particular, if one wants to estimate the whole distribution of Θ_i, such a larger index set is necessary. The process convergence is considered in the next section.

5.2.2 Uniform convergence of the standardized estimator

In the following, the uniform asymptotic behavior of $(\hat{p}_{n,A})_{A\in\mathcal{A}}$ over some family of Borel sets $\mathcal{A} \subset \mathcal{B}(\mathbb{R}^d)$ in $\mathbb{R}^d$ shall be determined. To this end, we want to consider the uniform process convergence of the empirical processes $(Z_n(A))_{A\in\mathcal{A}}$ associated to $T_{n,A}$ and $(Z_n^{pb}(A))_{A\in\mathcal{A}}$ associated to $\hat{p}_{n,A}$, respectively. Again, we start with the analysis for $(Z_n(A))_{A\in\mathcal{A}}$ and conclude the uniform process convergence of $(Z_n^{pb}(A))_{A\in\mathcal{A}}$ using the same continuous mapping arguments as before.
For uniform process convergence, we have to show fidis convergence and either asymptotic tightness or asymptotic equicontinuity of $(Z_n(A))_{A\in\mathcal{A}}$. The fidis convergence was already established in the previous section. We will consider asymptotic tightness and, basically, the uniform convergence holds, if the family $\mathcal{A}$ is not too complex, i.e. we have to restrict the family $\mathcal{A}$.
Here, we assume that $\mathcal{A}$ can be indexed by a unit cube of arbitrary dimension $q \in \mathbb{N}$ in a suitable way. These assumptions are summarized in the following conditions (PA) (i)-(vii). They will enable us to apply suitable brackets for the bracketing entropy (cf. condition (D2)). Note that we use the vector notation $\bar{t} = (\bar{t}_1, ..., \bar{t}_d) \leq (t_1, ..., t_d) = t \in \mathbb{R}^d$ as componentwise inequality, i.e. $\bar{t}_j \leq t_j$ for all $1 \leq j \leq d$ and $\bar{t} = (\bar{t}_1, ..., \bar{t}_l) < (t_1, ..., t_l) = t$ if and only if $t_j \leq \bar{t}_j$ for all $j \in \{1, ..., l\}$ and $t_k < \bar{t}_k$ for some $k \in \{1, ..., l\}$, i.e. $\bar{t} < t$ if and only if $\bar{t} \leq t$ and $\bar{t} \neq t$.

(PA) For some $q \in \mathbb{N}$, there exists a map $[0,1]^q \to \mathcal{A}, t \mapsto A_t$ such that

- (i) $\mathcal{A} = \{A_t | t \in [0,1]^q\}$, $A_{(1,...,1)} = \mathbb{R}^d$ and $A_{(t_1,...,t_q)} = \emptyset$ if $t_j = 0$ for some $1 \leq j \leq q$;
- (ii) for all $1 \leq j, k \leq q$, and all $s_j, t_l \in [0,1]$, ($l \in \{1, \ldots, q\} \setminus \{j\}$) the mapping $t_k \mapsto A_{(t_1,...,t_q)} \setminus A_{(t_1,...,t_{j-1},s_j,t_{j+1},...,t_q)}$ is non-decreasing on $[0,1]$ w.r.t. inclusion;
- (iii) the processes $\left(\sum_{i=1}^{r_n} g_{A_t}(W_{n,i})\right)_{t\in[0,1]^q}$ are separable;
- (iv) $P(\Theta_i \in \partial A_t^-) = 0$ for all $t \in [0, 1+\iota]^q$ for some $\iota > 0$ where $A_t^- := \bigcup_{s\in[0,t)} A_s$ and $A_t := A_{t\wedge 1}$ for $t \notin [0,1]^q$;
- (v) $P(\Theta_i \in \bigcap_{s\in(t,1]} A_{s^{(k)}} \setminus A_{t^{(k)}}) = 0$ for all $t \in [0,1)$ and $1 \leq k \leq q$ where $t^{(k)} := (1, \ldots, 1, t, 1, \ldots, 1)$ with t in the k-th coordinate;
- (vi) $P(\|X_0\| > 0,\, X_i/\|X_0\| \in \bigcap_{s\in(t,1]} A_{s^{(k)}} \setminus A_{t^{(k)}}) = 0$ for all $t \in [0,1)$ and $1 \leq k \leq q$;
- (vii) there exists $w \in [0,1]^q$ such that $0 \in A_w \setminus \bigcup_{s<w} A_s$.

In particular, condition (PA) (ii) implies that the map $t \mapsto A_t$ is non-decreasing w.r.t. inclusion in each coordinate. This will be an important argument below.

Example. The archetypical example for which condition (PA) is satisfied, provided the marginal distributions of Θ_i are continuous, is $\mathcal{A} = \{(-\infty, t] \mid t \in \mathbb{R}^d\} \cup \{\emptyset, \mathbb{R}^d\}$. To see this, define $A_{(t_1,\ldots,t_d)} := \times_{i=1}^d (-\infty, w(t_i)] \cap \mathbb{R}^d$ for some continuous, increasing mapping $w : [0,1] \to [-\infty, \infty]$ with $w(0) = -\infty$ and $w(1) = \infty$. Then, (i), (ii), (iii) and (vii) are obviously satisfied. Moreover, $\bigcap_{s\in(t,1]} A_{s^{(k)}} \setminus A_{t^{(k)}} = \emptyset$ for all $1 \le k \le q$ and $t \in [0,1]$, which is why (v) and (vi) are trivially fulfilled. Part (iv) corresponds to $P(\exists 1 \le j \le q : \Theta_{i,j} = t_j) = 0$ for all $t \in \mathbb{R}^d$, which follows, if Θ_i has continuous marginal distributions. In fact, by this choice of $\mathcal{A}$, condition (PA) (iv) is equivalent to the assumption that the marginal distribution of Θ_i are continuous. The same holds if $\mathcal{A}$ is a continuum with respect to the Lebesgue measure $\boldsymbol{\lambda}$, i.e. if $\boldsymbol{\lambda}(A_t \setminus A_s) \to 0$ for $s \uparrow t$.
Indeed, condition (PA) covers almost all natural finite-dimensional families of sets if the distributions of Θ_i and $X_i/\|X_0\|$ are sufficiently smooth. $\diamond$

Using the condition (PA), in the next proposition we show the process convergence of $(Z_n(A))_{A\in\mathcal{A}}$.

Proposition 5.2.7. *Suppose the conditions (PR), (P0), (PP), (PT), (PC) and (PA) are satisfied. Then the process $(Z_n(A))_{A\in\mathcal{A}}$ converges weakly to a centered Gaussian process $(Z(A))_{A\in\mathcal{A}}$ with covariance function c as given in Lemma 5.2.4.*

The proof of this proposition establishes asymptotic tightness using some bracketing conditions. Alternatively to this approach, one could establish the process convergence of $(Z_n(A))_{A\in\mathcal{A}}$ in the space of bounded functions indexed by $\mathcal{A}$ (equipped with the supremum norm) relatively easily by using Vapnik-Chervonenkis (VC) theory if the family of sets $\mathcal{A}$ under consideration is linearly ordered w.r.t. inclusion. In this case one can weaken condition (PA) considerably and the proof of process convergence simplifies. For $d = 1$ one can determine the whole distribution of Θ_i by some linearly ordered index set. However, this assumption on $\mathcal{A}$ is too restrictive if multivariate data is observed, i.e. if $d > 1$, which is why we introduced the more complex but also more general condition (PA). Still, in Corollary 5.2.8 the special case of linearly ordered sets is considered.
In case that $\mathcal{A}$ is linearly ordered, one has $q = 1$ in condition (PA) and part (ii) is trivially fulfilled and the statement of the previous theorem remains true, if one omits parts (vi) and (vii) of (PA). In fact, with $q = 1$ condition (PA) (ii) is equivalent to the assumption of linearly ordered $\mathcal{A}$.

Example. For $d = 1$, $\mathcal{A} = \{(-\infty, x] \mid x \in \mathbb{R}\}$ is one example of a linearly ordered family. In this special case $(P(\Theta_i \in A))_{A\in\mathcal{A}}$ determines the whole distribution of Θ_i.
Another example for linearly ordered sets in $\mathbb{R}^d$ is $\mathcal{A} = \{B_r(y) | r \in \mathbb{R}^+\}$, where we define the ball with radius r around y with respect to the norm $\|\cdot\|$ as $B_r(y) = \left\{x \in \mathbb{R}^d : \|x - y\| \le r\right\}$. However, for $d > 1$ the assumption that $\mathcal{A}$ is linearly ordered is

quite restrictive. ◇

Corollary 5.2.8. *Suppose the conditions (PR), (P0), (PP), (PT) and (PC) are satisfied. Suppose $\mathcal{A}$ is linearly ordered and includes $\emptyset$ and $\mathbb{R}^d$. In addition, assume conditions (PA) (iii),(iv) and (v). Then the process $(Z_n(A))_{A\in\mathcal{A}}$ converges weakly to centered Gaussian process $(Z(A))_{A\in\mathcal{A}}$ with covariance function c as given in Lemma 5.2.4.*

So far, we have completed the investigation of the asymptotic behavior of the statistics $(T_{n,A})_{A\in\mathcal{A}}$. Now we derive the uniform asymptotic normality of $\hat{p}_{n,A}$, i.e. the process convergence of $(Z_n^{pb}(A))_{A\in\mathcal{A}}$. Only one additional assumption is needed in order to control the bias of $\hat{p}_{n,A}$:

(PB$_T$)

$$\sup_{A\in\mathcal{A}}\left|\frac{1}{nv_n}E\left[T_{n,A}\right]-P(\Theta_i\in A)\right|=\mathrm{o}\Big((nv_n)^{-1/2}\Big).$$

Due to Proposition 5.2.1, the difference on the left hand side of (PB$_T$) converges to 0 as $n\to\infty$ and $(nv_n)^{-1}T_{n,A}$ is asymptotically unbiased for $P(\Theta_i\in A)$. The condition (PB$_T$) imposes only a condition on the rate and uniformness of this convergence. In other words: nv_n has to increase sufficiently slowly, i.e. by this condition v_n is not allowed to decrease too slowly, i.e. u_n has to increase fast enough.
The next theorem states the asymptotic behavior of the projection based estimator $\hat{p}_{n,A}$. The covariance of the limit distribution depends on the spectral tail process and on the index of the regular variation α.

Theorem 5.2.9. *Suppose the conditions (PR), (P0), (PP), (PT), (PC), (PA) and (PB$_T$) are satisfied. Then, with $p_A:=P(\Theta_i\in A)$, $A\in\mathcal{A}$, the weak convergence*

$$(Z_n^{pb}(A))_{A\in\mathcal{A}}=(\sqrt{nv_n}\,(\hat{p}_{n,A}-p_A))_{A\in\mathcal{A}}\xrightarrow{w}\Big(Z^{pb}(A)\Big)_{A\in\mathcal{A}}$$

holds for the centered Gaussian process Z^{pb} defined by $Z^{pb}(A):=Z(A)-p_AZ(\mathbb{R}^d)$, where $(Z(A))_{A\in\mathcal{A}}$ is the limit process from Proposition 5.2.7, the covariance structure of Z^{pb} is given in (5.2.4).

Again we have proved the uniform asymptotic normality of $\hat{p}_{n,A}$ with the typical (extreme value) convergence rate $\sqrt{nv_n}$.
One drawback of this central limit theorem for practical applications with unknown spectral tail process is that the covariance is a complex expression including infinite sums and depending on the unknown spectral tail process. This makes the practical construction of confidence intervals challenging. One way out would be the estimation of the asymptotic variance with bootstrap techniques. Two techniques were proposed by Davis et al. (2018) for a similar problem, a stationary bootstrap as in Davis et al. (2012) and a multiplier block bootstrap as in Drees (2015). However, these methods are not directly applicable

and have to be adapted for the sliding blocks setting used here. In this thesis we will not go further into this problem - it remains an open research question.
In practice, when calculating $\hat{p}_{n,A}$, one often has to use data driven thresholds $\hat{u}_n$ instead of a deterministic threshold u_n. This does not fit to the limit theory developed in the previous theorem. However, with similar methods as in Section 4.2.3 one can prove that the limit distribution of such a modified estimator with suitable random thresholds $\hat{u}_n$ is the same as in the previous theorem. This is shown in detail in the Supplemental material, Section 15, of Drees et al. (2021).
At this point, we have completed the asymptotic analysis of the projection based estimator $\hat{p}_{n,A}$ with known α. Before we continue with the estimator $\hat{\hat{p}}_{n,A}$ with unknown and estimated α, we want to consider one special case where the asymptotic variance equals 0.

Covariances for a deterministic shape

We consider the special case of a *deterministic shape* of the spectral tail process $(\Theta_t)_{t\in\mathbb{Z}}$. The meaning of the shape can be explained as follows: denote by $(\Theta_t^*)_{t\in\mathbb{Z}}$ the time-shifted and rescaled process $(\Theta_t)_{t\in\mathbb{Z}}$ such that the maximal norm 1 occurs the first time at time point 0, i.e. $1 = \|\Theta_0^*\| \geq \|\Theta_t^*\|$ for all $t \in \mathbb{Z}$ and $\|\Theta_t^*\| < 1$ for $t < 0$. With the random variable $T^* := \inf\{t \in \mathbb{Z} : \|\Theta_t\| = \sup_{l\in\mathbb{Z}} \|\Theta_l\|\}$ we have

$$\Theta_t^* := \frac{\Theta_{T^*+t}}{\|\Theta_{T^*}\|}$$

for all $t \in \mathbb{Z}$. This process $(\Theta_t^*)_{t\in\mathbb{Z}}$ describes the *shape* of $(\Theta_t)_{t\in\mathbb{Z}}$ (and, therefore, is called the shape of $(\Theta_t)_{t\in\mathbb{Z}}$), in the sense that modulo some random time-shift and rescaling it has the same form as $(\Theta_t)_{t\in\mathbb{Z}}$. This means, that given T^* and (Θ_t^*, Θ_t) are known for one $t \in \mathbb{Z}$, then the whole process $(\Theta_t^*)_{t\in\mathbb{Z}}$ is determined by $(\Theta_t)_{t\in\mathbb{Z}}$ and vice versa, i.e. if the random shift and rescaling is known, then one process determines the other process. From Definition 2.2.4 of the RS-transformation and by the definition of $(\Theta_t^*)_{t\in\mathbb{Z}}$ it directly follows, that $(\Theta_t)_{t\in\mathbb{Z}}$ and $(\Theta_t^*)_{t\in\mathbb{Z}}$ have the same RS-transformation (since the normalization and shift in Θ_t^* cancel out).
The expression $\sum_{h\in\mathbb{Z}}(\|\Theta_h\|^\alpha/\|\Theta\|_\alpha^\alpha)\mathbb{1}_B(\Theta_{h+i}/\|\Theta_h\|)$ in c^{pb} in (5.2.4) obviously only depends on the shape $(\Theta_t^*)_{t\in\mathbb{Z}}$, i.e.

$$\sum_{h\in\mathbb{Z}}(\|\Theta_h\|^\alpha/\|\Theta\|_\alpha^\alpha)\mathbb{1}_B(\Theta_{h+i}/\|\Theta_h\|) = \sum_{h\in\mathbb{Z}}(\|\Theta_h^*\|^\alpha/\|\Theta^*\|_\alpha^\alpha)\mathbb{1}_B(\Theta_{h+i}^*/\|\Theta_h^*\|).$$

Due to the RS-transformation and Theorem 2.2.5 we obtain

$$E\Big[\sum_{h\in\mathbb{Z}}\frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha}\mathbb{1}_B\Big(\frac{\Theta_{h+i}}{\|\Theta_h\|}\Big)\Big] = E[\mathbb{1}_B(\Theta_i)] = P(\Theta_i \in B) = p_B.$$

A deterministic shape means, that $(\Theta_t^*)_{t\in\mathbb{Z}}$ is deterministic. In particular if $(\Theta_t^*)_{t\in\mathbb{Z}}$ is deterministic, then $\sum_{h\in\mathbb{Z}}(\|\Theta_h\|^\alpha/\|\Theta\|_\alpha^\alpha)\mathbb{1}_B(\Theta_{h+i}/\|\Theta_h\|) = p_B$ a.s. for all $B \subset \mathbb{R}^d$ and therefore

$$c^{pb}(A,B) = 0$$

for all sets $A, B \in \mathcal{A}$. For deterministic shapes $(\Theta_t^*)_{t\in\mathbb{Z}}$ of the spectral tail process the projection based estimator $\hat{p}_{n,A}$ has the asymptotic variance 0. In this case the new estimator considered here has a smaller asymptotic variance and is more efficient than the known forward and backward estimators $\hat{p}_{n,A}^f$ and $\hat{p}_{n,A}^b$ from the literature.
One example for such a deterministic shape is the spectral tail process of an AR(1) time series $(X_t)_{t\in\mathbb{Z}}$, with $X_{t+1} = aX_t + \varepsilon_t$, $a \in (0,1)$ and iid innovations ε_t with regularly varying distribution with index $\alpha > 0$, e.g. $|\varepsilon_t| \sim Par(\alpha)$. In this case the spectral tail process is given by $\Theta_t^* = a^t\Theta_0^*$, $t \geq 0$ and $\Theta_t^* = 0$, $t < 0$, which obviously has a deterministic shape (cf. Janssen and Segers (2014), Example 6.1).
Further consideration of the asymptotic variance and a comparison with the asymptotic variances of $\hat{p}_{n,A}^f$ and $\hat{p}_{n,A}^b$ will follow in Section 5.5.1. Next, we consider the asymptotic behavior of $\hat{\hat{p}}_{n,A}$ with estimated α.

5.3 Estimator with unknown index of regular variation

So far we considered the projection based estimator $\hat{p}_{n,A}$ assuming that we know the index α of regular variation. Usually this index α is unknown and has to be estimated itself. One possible choice for the estimator is the Hill-type estimator in (5.1.5)

$$\hat{\alpha}_n := \frac{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}}{\sum_{t=1}^n \log(\|X_t\|/u_n)\mathbb{1}_{\{\|X_t\|>u_n\}}}$$

as introduced before. Recall the definition of $\hat{\hat{p}}_{n,A}$ in (5.1.6)

$$\hat{\hat{p}}_{n,A} = \frac{1}{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}} \sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h=-s_n}^{s_n} \frac{\|X_{t+h}\|^{\hat{\alpha}_n}}{\sum_{k=-s_n}^{s_n} \|X_{t+k}\|^{\hat{\alpha}_n}} \times \left(\mathbb{1}_A\left(\frac{X_{t+h+i}}{\|X_{t+h}\|}\right)\mathbb{1}_{\{h\in H_{n,i}\}} + \mathbb{1}_A(0)\mathbb{1}_{\{h\in H_{n,i}^C\}}\right).$$

With this estimator $\hat{\alpha}_n$ we will analyze the asymptotic behavior of the projection based estimator $\hat{\hat{p}}_{n,A}$. We will conclude the asymptotic behavior of $\hat{\hat{p}}_{n,A}$ from that of $\hat{p}_{n,A}$ and $\hat{\alpha}_n$ using a Taylor argument. For this purpose the asymptotic behavior of $\hat{\alpha}_n$ as well as the joint asymptotic behavior of the denominator of $\hat{\alpha}_n$ and $T_{n,A}$ are important ingredients. For the analysis of the denominator of $\hat{\alpha}_n$ we define the function

$\phi : l_\alpha \to \mathbb{R}$ by $\phi((x_h)_{h\in\mathbb{Z}}) = \log(\|x_0\|)\mathbb{1}_{\{\|x_0\|>1\}} = \log^+(\|x_0\|)$ and a corresponding statistic $T_{n,\phi} = \sum_{t=1}^n \phi(W_{n,t})$ analogously to $T_{n,A}$. Moreover, we define the corresponding empirical process $Z_n(\phi)$ by

$$Z_n(\phi) := \frac{1}{\sqrt{nv_n}}(T_{n,\phi} - E[T_{n,\phi}]).$$

Then, we have $\hat{\alpha}_n = T_{n,\mathbb{R}^d}/T_{n,\phi}$, similar to the representation of $\hat{p}_{n,A} = T_{n,A}/T_{n,\mathbb{R}^d}$.
To analyze the joint asymptotic behavior of $T_{n,\phi}$ and the statistic $T_{n,A}$, some additional conditions are required. First, we amend the cluster size condition (PP) to condition (PP1) which is necessary since ϕ occurring in the definition of $\hat{\alpha}_n$ is unbounded. For the same reason, a further moment condition is needed to ensure the asymptotic normality of $\hat{\alpha}_n$. This moment condition is given in (PM) (i). Moreover, we require the bias of the Hill type estimator $\hat{\alpha}_n$ to be negligible, this is formalized in (PB$_\alpha$). This bias condition imposes a rate for the convergence of the bias, which ensures that the bias converges faster to 0 than the stochastic error (compare this with the previous bias condition (PB$_T$) for $T_{n,A}$).

(PP1) For all $n \in \mathbb{N}$ and for all $1 \le k \le r_n$ there exists

$$\begin{aligned} e_n'(k) \ge E\Bigg[&\max\Big(\log^+\Big(\frac{\|X_0\|}{u_n}\Big), 1\Big) \\ &\times \max\Big(\log^+\Big(\frac{\|X_k\|}{u_n}\Big), \mathbb{1}_{\{\|X_k\|>u_n\}}\Big) \,\Big|\, \|X_0\| > u_n\Bigg], \end{aligned}$$

such that $e_\infty'(k) = \lim_{n\to\infty} e_n'(k)$ exists for all $k \in \mathbb{N}$ and $\lim_{n\to\infty} \sum_{k=1}^{r_n} e_n'(k) = \sum_{k=1}^\infty e_\infty'(k) < \infty$ holds.

(PB$_\alpha$) $\left| E\Big[\log\Big(\frac{\|X_0\|}{u_n}\Big) \,\Big|\, \|X_0\| > u_n\Big] - \frac{1}{\alpha}\right| = \mathrm{o}\Big((nv_n)^{-1/2}\Big).$

(PM) There exists a $\delta > 0$ such that the following moment bounds hold for $n \to \infty$:

(i) $$\sum_{k=1}^{r_n}\Bigg(E\Bigg[\Big(\log^+\Big(\frac{\|X_0\|}{u_n}\Big)\log^+\Big(\frac{\|X_k\|}{u_n}\Big)\Big)^{1+\delta} \,\Big|\, \|X_0\| > u_n\Bigg]\Bigg)^{1/(1+\delta)} = \mathrm{O}(1);$$

(ii) $$\limsup_{m\to\infty} E\left[\frac{\sum_{|h|\le m} |\log(\|\Theta_h\|)|^{1+\delta}\|\Theta_h\|^\alpha}{\sum_{|k|\le m}\|\Theta_k\|^\alpha}\right] < \infty;$$

(iii) $$\lim_{m\to\infty}\limsup_{n\to\infty} E\left[\frac{\sum_{m<|h|\le s_n}\log^-(\|X_h\|/u_n)\|X_h\|^\alpha}{\sum_{|k|\le s_n}\|X_k\|^\alpha} \,\Bigg|\, \|X_0\| > u_n\right] = 0.$$

Here we denote the positive part of any function f by $f^+ := \max(f, 0) = f\mathbb{1}_{\{f>0\}}$ and the negative part by $f^- = \max(-f, 0) = -f\mathbb{1}_{\{f<0\}}$.
Drees et al. (2015), Lemma 4.4, showed that under conditions (PR), (P0), (PP1), (PB$_\alpha$) and (PM) (i) the estimator $\hat{\alpha}_n$ is asymptotically normal:

$$\sqrt{nv_n}(\hat{\alpha}_n - \alpha) \xrightarrow{w} Z_\alpha,$$

where Z_α is a centered normal distributed random variable with variance specified below in the proofs.
The moment conditions in (PM) (ii) - (iii) are required to ensure the asymptotic normality of $\hat{\hat{p}}_{n,A}$. Both conditions are used for a truncation technique which will be applied in the proofs below to cut off some infinite sums. Condition (PM) (ii) is needed because of the unboundedness of the logarithm, while (PM) (iii) would be an additional assumption even if the denominator of $\hat{\alpha}_n$ would be the sum over some bounded function. Condition (PM) (iii) basically restricts the summed norm of non-extreme observations in the surrounding of some extreme event and it allows for an interchange of the limit for $m \to \infty$ and $n \to \infty$.
The uniform asymptotic normality of the projection based estimator $\hat{\hat{p}}_{n,A}$ is stated in the next theorem.

Theorem 5.3.1. *Suppose the conditions (PR), (P0), (PP), (PT), (PC), (PA), (PB_T), (PP1), (PB_α) and (PM) are satisfied and, in addition, $\log(n)^4 = o(nv_n)$. Then the joint convergence*

$$((Z_n(A))_{A\in\mathcal{A}}, Z_n(\phi)) \xrightarrow[n\to\infty]{w} ((Z(A))_{A\in\mathcal{A}}, Z(\phi))$$

holds weakly for a centered Gaussian process $((Z(A))_{A\in\mathcal{A}}, Z(\phi))$ with covariances given by $Cov(Z(A), Z(B)) = c(A, B)$, $A, B \in \mathcal{A}$, given in Lemma 5.2.4 and $Cov(Z(A), Z(\phi))$ and $Var(Z(\phi))$ given below in Lemma 5.7.7 part (i) and (ii), respectively.
Moreover, the weak convergence

$$\left(\sqrt{nv_n}\left(\hat{\hat{p}}_{n,A} - p_A\right)\right)_{A\in\mathcal{A}} \xrightarrow{w} \left(Z^{pb,\alpha}(A)\right)_{A\in\mathcal{A}}$$

holds for the centered Gaussian process $Z^{pb,\alpha}$ defined by

$$Z^{pb,\alpha}(A) := Z(A) - (p_A - \alpha d_A)Z(\mathbb{R}^d) - d_A\alpha^2 Z(\phi)$$

for all $A \in \mathcal{A}$, with $p_A := P(\Theta_i \in A)$ and

$$d_A := -E\Big[\sum_{k\in\mathbb{Z}} \log(\|\Theta_k\|)\|\Theta_k\|^\alpha \|\Theta\|_\alpha^{-\alpha} \mathbb{1}_A(\Theta_i)\Big].$$

Remark. The covariance of the limit process $(Z^{pb,\alpha}(A))_{A\in\mathcal{A}}$ in Theorem 5.3.1 can be calculated more explicitly. The single covariances $Cov(Z(A), Z(B))$, $Cov(Z(A), Z(\phi))$ and $Var(Z(\phi))$, $A, B \in \mathcal{A}$, of the components in $Z^{pb,\alpha}(A)$ are given in Lemma 5.2.4 and Lemma 5.7.7. This leads to the covariance of $Z^{pb,\alpha}$:

$$\begin{aligned} &Cov(Z^{pb,\alpha}(A), Z^{pb,\alpha}(B)) \qquad (5.3.1)\\ &= Cov(Z(A) - (p_A - \alpha d_A)Z(\mathbb{R}^d) - \alpha^2 d_A Z(\phi), \end{aligned}$$

$$
\begin{aligned}
&Z(B) - (p_B - \alpha d_B) Z(\mathbb{R}^d) - \alpha^2 d_B Z(\phi)) \\
= {} & Cov(Z(A), Z(B)) - (p_B - \alpha d_B) Cov(Z(A), Z(\mathbb{R}^d)) \\
& - \alpha^2 d_B Cov(Z(A), Z(\phi)) - (p_A - \alpha d_A) Cov(Z(B), Z(\mathbb{R}^d)) \\
& + (p_A - \alpha d_A)(p_B - \alpha d_B) Cov(Z(\mathbb{R}^d), Z(\mathbb{R}^d)) \\
& + (p_A - \alpha d_A) \alpha^2 d_B Cov(Z(\mathbb{R}^d), Z(\phi)) \\
& - \alpha^2 d_A Cov(Z(B), Z(\phi)) + (p_B - \alpha d_B) \alpha^2 d_A Cov(Z(\mathbb{R}^d), Z(\phi)) \\
& + \alpha^4 d_A d_B Var(Z(\phi)).
\end{aligned}
$$

◇

Condition (PM) (ii) is equivalent to the same condition with the tail process instead of $(\Theta_t)_{t\in\mathbb{Z}}$ (see next lemma). This is also the way how it is used in the proofs. However, it seems natural to state the conditions for an estimation problem of $P(\Theta_i \in A)$ in terms of the spectral tail process.

The condition (PM) (ii) is just as needed in the proof and cannot be weakened significantly. The verification of condition (PM) (ii) could be challenging and sometimes it could be much easier to check a stronger condition without fractions. One such stronger moment condition is given by (PM1), for which the denominator is basically eliminated by bounding it from below with 1. We will verify this condition instead of (PM) (ii) in Section 5.5 for solutions to stochastic recurrence equations.

(PM1) There exists an $\delta > 0$ such that the following moment bounds hold for $n \to \infty$:

(i) $E\left[\sup_{-s_n \le h \le s_n} \left(\log^+\left(\frac{\|X_h\|}{u_n}\right)\right)^{1+\delta} \,\middle|\, \|X_0\| > u_n\right] = \mathrm{O}(1)$,

(ii) $\sum_{h=-s_n}^{s_n} E\left[\left(\log^-\left(\frac{\|X_h\|}{u_n}\right)\right)^{1+\delta} \left(\frac{\|X_h\|}{u_n}\right)^{\alpha} \,\middle|\, \|X_0\| > u_n\right] = \mathrm{O}(1)$.

Lemma 5.3.2. *(i) Condition (PM) (ii) is equivalent to*

$$
\limsup_{m\to\infty} E\left[\frac{\sum_{|h|\le m} |\log(\|Y_h\|)|^{1+\delta} \|Y_h\|^{\alpha}}{\sum_{|k|\le m} \|Y_k\|^{\alpha}}\right] < \infty. \tag{5.3.2}
$$

(ii) Conditions (PM1) (i) and (ii) imply Condition (PM) (ii).

Condition (PM1) (ii) is in particular satisfied if

$$
\sum_{h=-s_n}^{s_n} E\left[\left(\frac{\|X_h\|}{u_n}\right)^{\eta_0} \mathbb{1}_{\{\|X_h\| < u_n\}} \,\middle|\, \|X_0\| > u_n\right] = \mathrm{O}(1)
$$

for some $\eta_0 \in (0, \alpha)$. This holds true, since the function $x \mapsto \log^-(x) x^q$ is bounded on $[0, 1]$ for all $q > 0$.

Note that condition (PM1) (i) includes a supremum over $h \in \{-s_n, ..., s_n\}$, while (PM) (i) sums only over positive $h \in \{1, ..., r_n\}$, which is why (PM1) (i) is a different type of

assumption. It is often harder to check such moment bound conditions for the backward process (i.e. for $h < 0$) than for the forward process as in (PM) (i). The condition (PM1) (i) with a supremum over $h \in \{1, ..., s_n\}$ would be implied by (PM) (i).
This concludes the consideration of asymptotic normality of $\hat{p}_{n,A}$ and $\hat{\hat{p}}_{n,A}$ and the discussion of the conditions. In particular, under these conditions estimators are consistent with the true probabilities. In the next part we generalize the method of the projection based estimator for multiple time points.

5.4 Estimator for multiple time points

In Equation (5.1.4) the projection based estimator $\hat{p}_{n,A}$ for $P(\Theta_i \in A)$ for some fixed $i \in \mathbb{Z}$ and some Borel set $A \subset \mathbb{R}^d$ was introduced. In the previous sections this estimator was analyzed for a single time point $i \in \mathbb{Z}$. However, the motivation in Section 5.1 allows for the construction of a projection based estimator for the whole distribution of $(\Theta_t)_{|t| \le s_n}$. This estimator for the measure is stated in (5.1.3). In particular, this projection based estimator can be used to define an estimator for multiple time points similar to the estimator of $P(\Theta_i \in A)$ considered before. Denote a finite number of time points by $i_1 < i_2 < ... < i_M \in \mathbb{Z}$, $M \in \mathbb{N}$, with $|i_j| \le s_n$ for $j = 1, ..., M$.
The projection based estimator for the probability $P((\Theta_{i_1}, ..., \Theta_{i_M}) \in A)$, with some Borel set $A \subset (\mathbb{R}^d)^M$, is given by inserting the specific set $\{z \in l_\alpha | (z_{i_1}, ..., z_{i_M}) \in A\}$ in (5.1.3). This leads to

$$
\begin{aligned}
\hat{p}^M_{n,A} :&= \hat{P^{\Theta}_n}^{RS}\Big(\{z \in l_\alpha | (z_{i_1}, ..., z_{i_M}) \in A\}\Big) \\
&= \frac{1}{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\| > u_n\}}} \\
&\quad \times \sum_{t=1}^n \Bigg(\mathbb{1}_{\{\|X_t\| > u_n\}} \Bigg(\sum_{h=(-s_n-i_1)\vee -s_n}^{(s_n-i_M)\wedge s_n} \frac{\|X_{t+h}\|^\alpha}{\sum_{k=-s_n}^{s_n} \|X_{t+k}\|^\alpha} \mathbb{1}_A\Big(\frac{X_{t+h+i_1}}{\|X_{t+h}\|}, ..., \frac{X_{t+h+i_M}}{\|X_{t+h}\|}\Big) \\
&\quad + \sum_{\substack{h \in \{-s_n, ..., -s_n-i_M-1\} \\ \cup \{s_n-i_1+1, ..., s_n\}}} \frac{\|X_{t+h}\|^\alpha}{\sum_{k=-s_n}^{s_n} \|X_{t+k}\|^\alpha} \mathbb{1}_A(0, ..., 0) \\
&\quad + \sum_{j=1}^{M-1} \sum_{h=(s_n-i_{j+1}+1)\wedge s_n}^{(s_n-i_j)\wedge s_n} \frac{\|X_{t+h}\|^\alpha}{\sum_{k=-s_n}^{s_n} \|X_{t+k}\|^\alpha} \mathbb{1}_A\Big(\frac{X_{t+h+i_1}}{\|X_{t+h}\|}, ..., \frac{X_{t+h+i_j}}{\|X_{t+h}\|}, 0, ..., 0\Big) \\
&\quad + \sum_{j=1}^{M-1} \sum_{h=(-s_n-i_{j+1})\vee -s_n}^{(-s_n-i_j)\vee -s_n} \frac{\|X_{t+h}\|^\alpha}{\sum_{k=-s_n}^{s_n} \|X_{t+k}\|^\alpha} \mathbb{1}_A\Big(0, ..., 0, \frac{X_{t+h+i_{j+1}}}{\|X_{t+h}\|}, ..., \frac{X_{t+h+i_M}}{\|X_{t+h}\|}\Big)\Bigg)\Bigg),
\end{aligned}
$$

where $\Big(0, ..., 0, X_{t+h+i_{j+1}}/\|X_{t+h}\|, ..., X_{t+h+i_M}/\|X_{t+h}\|\Big)$ is a short notation for the vector $(z_1, ..., z_M) \in (\mathbb{R}^d)^m$ with $z_k = X_{t+h+i_k}/\|X_{t+h}\|$ for $k \ge j+1$ and $z_k = 0$ for $k \le j$. The vector $\Big(X_{t+h+i_1}/\|X_{t+h}\|, ..., X_{t+h+i_j}/\|X_{t+h}\|, 0, ..., 0\Big)$ is interpreted analogously. Define the function $g_{A,M}$ by

$$g_{A,M}\left((y_h)_{h\in\mathbb{Z}}\right) := \mathbb{1}_{\{\|y_0\|>1\}} \sum_{h\in\mathbb{Z}} \frac{\|y_h\|^\alpha}{\sum_{k\in\mathbb{Z}}\|y_k\|^\alpha} \mathbb{1}_A\left(\frac{y_{h+i_1}}{\|y_h\|}, ..., \frac{y_{h+i_M}}{\|y_h\|}\right)$$

and $T^M_{n,A} = \sum_{t=1}^n g_{A,M}(W_{n,t})$ with $W_{n,t} = (X_t/u_n)_{-s_n\le t\le s_n}$. Then

$$\hat{p}^M_{n,A} = \frac{T^M_{n,A}}{T^M_{n,(\mathbb{R}^d)^M}}.$$

The principal idea of this estimator is the same as of the estimator for a single time point in (5.1.4), the notation is only a bit more complicated due to the multiple time points and the summands which include some zeros in the indicators. However, with the same techniques as in the previous sections one can prove the asymptotic normality of this estimator. Similarly as for a single time point, the summands for $h \notin \{(-s_n - i_1) \vee -s_n, ..., (s_n - i_M) \wedge s_n\}$, which are the last three lines in the definition of $\hat{p}^M_{n,A}$, are asymptotically negligible for fixed time points $i_1, ..., i_M$ and increasing $s_n \to \infty$. Asymptotically, only the first two lines of the estimator $\hat{p}^M_{n,A}$ are relevant.
As before, in practice α is often unknown and has to be estimated itself. We denote the corresponding projection based estimator for multiple time points with estimated α by $\hat{\hat{p}}^M_{n,A}$. This estimator is defined as $\hat{p}^M_{n,A}$, just replace α by the Hill type estimator $\hat{\alpha}_n$ from (5.1.5).
For the statement of asymptotic normality some modified conditions are necessary, which take into account the multiple time points. The obvious change is that we now consider Borel sets $A \subset (\mathbb{R}^d)^M$.

(PCM) $P((\Theta_{i_1}, ..., \Theta_{i_M}) \in \partial A) = 0$ for all $A \in \mathcal{A}$.

(PAM) For some $q \in \mathbb{N}$, there exists a map $[0,1]^q \to \mathcal{A} \subset \mathcal{B}((\mathbb{R}^d)^M), t \mapsto A_t$, such that

- (i) $\mathcal{A} = \{A_t | t \in [0,1]^q\}$, $A_{(1,...,1)} = (\mathbb{R}^d)^M$ and $A_{(t_1,...,t_q)} = \emptyset$ if $t_j = 0$ for some $1 \le j \le q$;
- (ii) for all $1 \le j,k \le q$, and all $s_j, t_l \in [0,1]$, ($l \in \{1,\dots,q\} \setminus \{j\}$) the mapping $t_k \mapsto A_{(t_1,...,t_q)} \setminus A_{(t_1,...,t_{j-1},s_j,t_{j+1},...,t_q)}$ is non-decreasing w.r.t. inclusion on $[0,1]$;
- (iii) The processes $\left(\sum_{i=1}^{r_n} g_{A_t,M}(W_{n,i})\right)_{t\in[0,1]^q}$ are separable;
- (iv) $P((\Theta_{i_1}, ..., \Theta_{i_M}) \in \partial A_t^-) = 0$ for all $t \in [0, 1+\iota]^q$ for some $\iota > 0$ where $A_t^- := \bigcup_{s\in[0,t)} A_s$ and $A_t := A_{t\wedge 1}$ for $t \notin [0,1]^q$;
- (v) $P((\Theta_{i_1}, ..., \Theta_{i_M}) \in \bigcap_{s\in(t,1]} A_{s^{(k)}} \setminus A_{t^{(k)}}) = 0$ for all $t \in [0,1)$ and $1 \le k \le q$;
- (vi) $P((X_{i_1}/\|X_0\|, ..., X_{i_M}/\|X_0\|) \in \bigcap_{s\in(t,1]} A_{s^{(k)}} \setminus A_{t^{(k)}}) = 0$ for all $t \in [0,1)$ and $1 \le k \le q$;
- (vii) there exists $w \in [0,1]^q$ such that $0 \in A_w \setminus \bigcup_{s<w} A_s$.

(PB$_T$M)

$$\sup_{A\in\mathcal{A}}\left|\frac{1}{nv_n}E\left[T_{n,A}^M\right]-P((\Theta_{i_1},...,\Theta_{i_M})\in A)\right|=\mathrm{o}\left((nv_n)^{-1/2}\right).$$

Again, the bias assumption is only an assumption on the rate and uniformness of the convergence, which ensures that the bias converges faster than the stochastic error.
If all sets $A\in\mathcal{A}$ are Cartesian products of the form $A=A_1\times...\times A_M$, condition (PCM) is implied by $P(\Theta_{i_j}\in\partial A_j)=0$ for all $j=1,...,M$.
All other conditions used to establish asymptotic normality of $\hat{\hat{p}}_{n,A}$ do not depend on the time point i or the sets $A\in\mathcal{A}$ and, therefore, remain unchanged. With these new conditions we can state asymptotic normality of the projection based estimator $\hat{p}_{n,A}^M$ and $\hat{\hat{p}}_{n,A}^M$ for multiple time points with known and estimated α, respectively.

Theorem 5.4.1. *Suppose the conditions (PR), (P0), (PP), (PT), (PCM), (PAM) and (PB$_T$M) are satisfied. Then the weak convergence*

$$\left(\sqrt{nv_n}\left(\hat{p}_{n,A}^M-P((\Theta_{i_1},...,\Theta_{i_M})\in A)\right)\right)_{A\in\mathcal{A}}\xrightarrow{w}\left(Z^{pb,M}(A)\right)_{A\in\mathcal{A}}$$

holds for a centered Gaussian process $Z^{pb,M}$.
If, in addition, (PP1), (PB$_\alpha$) and (PM) are met, then the weak convergence

$$\left(\sqrt{nv_n}\left(\hat{\hat{p}}_{n,A}^M-P((\Theta_{i_1},...,\Theta_{i_M})\in A)\right)\right)_{A\in\mathcal{A}}\xrightarrow{w}\left(Z^{pb,M,\alpha}(A)\right)_{A\in\mathcal{A}}$$

holds for a centered Gaussian process $Z^{pb,M,\alpha}$.

The covariance of $Z^{pb,M}$ can be calculated analogously to the covariance of Z^{pb} in (5.2.4). For $A,B\in\mathcal{A}$ it is given by

$$\begin{aligned}Cov(Z^{pb,M}(A),Z^{pb,M}(B))=\sum_{j\in\mathbb{Z}}E\Bigg[(\|\Theta_j\|^\alpha\wedge 1)\Big(p_B-\sum_{h\in\mathbb{Z}}\frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha}\mathbb{1}_B\Big(\frac{\Theta_{h+i_1}}{\|\Theta_h\|},...,\frac{\Theta_{h+i_M}}{\|\Theta_l\|}\Big)\Big)\\ \times\Big(p_A-\sum_{l\in\mathbb{Z}}\frac{\|\Theta_l\|^\alpha}{\|\Theta\|_\alpha^\alpha}\mathbb{1}_A\Big(\frac{\Theta_{l+i_1}}{\|\Theta_l\|},...,\frac{\Theta_{l+i_M}}{\|\Theta_l\|}\Big)\Big)\Bigg]\end{aligned}$$

with $p_A:=P((\Theta_{i_1},...,\Theta_{i_M})\in A)$. The covariance of $Z^{pb,M,\alpha}$ can be calculated analogously as for $Z^{pb,\alpha}$ in (5.3.1).
As a special case, the previous theorem about $\hat{\hat{p}}_{n,A}^M$ also shows the asymptotic normality of the estimator for $(P(\Theta_t\in\tilde{A}))_{t\leq|T|}$ for any fixed Borel set $\tilde{A}\subset\mathbb{R}^d$ and $T\in\mathbb{N}$. To this end, consider the time points $i_j=-T+j-1$, $j=1,...,M:=2T+1$, and the estimators $(\hat{\hat{p}}_{n,A}^M)_{A\in\tilde{\mathcal{A}}}$ for $\tilde{\mathcal{A}}:=\{(\mathbb{R}^d)^{-T+|t|}\times\tilde{A}\times(\mathbb{R}^d)^{T-|t|},|t|\leq T\}$. Then the rest follows from Theorem 5.4.1. Note that $\tilde{\mathcal{A}}$ is a finite set such that the condition (PAM) for process convergence is not needed. (One easily checks in the proofs that (PAM) is only needed if $\mathcal{A}$ contains infinitely many sets.)
This rather simple construction of the estimator for multiple time points is an advantage of the motivation of the projection-based estimator compared to previous methods. For

the backward estimator in Davis et al. (2018), which uses the structure of the TCF for one shift i, the shift i is chosen as the lag for which $P(\Theta_i \le x)$ should be estimated. This is an ad-hoc procedure, a priori it is not clear if this is optimal or if it might be more efficient to use another shift for the application of the TCF to construct the backward estimator, e.g. a shift smaller than $|i|$. The applied shift is an implicit parameter in the construction of the backward estimator. If we consider multiple points in time, it is not clear which shift should be chosen for the construction of a backward estimator; the smallest, the largest or something in between or even outside the range of considered lags. Here, the shift is an additional parameter. For the projection based estimator $\hat{\hat{p}}^M_{n,A}$, this problem does not exist, but the parameter s_n appears. However, for the backward estimator the shift has an influence on the asymptotic variance, s_n has no influence on the asymptotic variance of the projection based estimator. All in all, the concept of the projection-based estimator is much easier to generalize to multiple time points than the ad-hoc method of the backward estimator.

This short section demonstrates that the RS-projection method can be used not only to estimate $P(\Theta_i \in A)$ for a single time point $i \in \mathbb{Z}$, but rather could be easily used to construct estimators for other probabilities as $P((\Theta_{i_1}, ..., \Theta_{i_M}) \in A)$ for finitely many multiple time points. The consideration of infinitely many time points and a large family $\mathcal{A}$ would be more involved.

So far we introduced the projection based estimator and proved asymptotic normality of this estimator under certain conditions. In the next section, we consider an example for which these conditions could be satisfied for a reasonable class of time series.

5.5 Examples

In this section, we present two examples. The first one deals with the asymptotic covariances and how they could be calculated in a discrete case. With this example we show that neither our new projection based estimator $\hat{\hat{p}}_{n,A}$ (or $\hat{p}_{n,A}$ in case of known α) nor the backward estimator $\hat{p}^b_{n,A}$ nor the forward estimator $\hat{p}^f_{n,A}$ have the uniformly smallest asymptotic variance. In the second example we consider stationary solutions to stochastic recurrence equations and we verify the conditions from the previous sections. In particular, this example demonstrates that all the conditions can be satisfied for a reasonable class of stochastic processes.

5.5.1 Comparison of asymptotic covariances

We want to compare the efficiency of our new projection based estimator $\hat{\hat{p}}_{n,A}$ in terms of asymptotic variance with two known estimators for $P(\Theta_i \in A)$ from the literature, the above introduced forward estimator $\hat{p}^f_{n,A}$ (cf. (5.1.1)) and the backward estimator $\hat{\hat{p}}^b_{n,A}$ (cf. (5.1.2)). We denote the backward estimator with estimated α by $\hat{\hat{p}}^b_{n,A}$. Due to the

complex formulas, a comparison of the asymptotic variances in general is not possible. However, we consider a specific example for the comparison of the covariances, which will show that neither $\hat{\hat{p}}_{n,A}$ nor $\hat{p}^b_{n,A}$ nor $\hat{p}^f_{n,A}$ have the uniformly smallest variance.
We consider a relatively simple model, where the shape of the spectral tail process $(\Theta^*_t)_{t\in\mathbb{Z}}$ is a mixture of two deterministic shapes, i.e. $(\Theta^*_t)_{t\in\mathbb{Z}}$ can take two different values with positive probability. For the definition of the shape, see the example on page 152.
Define a real valued time series $(U_t)_{t\in\mathbb{Z}}$ by

$$P\left(U_0 = a^{-1}, U_1 = -1, U_t = 0\,\forall t \notin \{0,1\}\right) = p,$$
$$P\left(U_0 = b, U_1 = 1, U_t = 0\,\forall t \notin \{0,1\}\right) = 1-p$$

for some $(a,b) \in (1,\infty)^2$ and $p \in [0,1]$, the index of regular variation as $\alpha = 1$ and $(\Theta_t)_{t\in\mathbb{Z}} = (U^{RS}_t)_{t\in\mathbb{Z}}$ as the RS-transformation of $(U_t)_{t\in\mathbb{Z}}$ (or define the distribution of $(\Theta_t)_{t\in\mathbb{Z}}$ as the RS-transformation of $(U_t)_{t\in\mathbb{Z}}$, i.e. $P^{(\Theta_t)_{t\in\mathbb{Z}}} = (P^{(U_t)_{t\in\mathbb{Z}}})^{RS}$, respectively). Then, $(\Theta_t)_{t\in\mathbb{Z}}$ is invariant under the RS-transformation (cf. Lemma 5.1.1) and therefore a spectral tail process (cf. Theorem 2.2.5). By the definition of the RS-transformation (cf. (2.2.5)) one has

$$\begin{aligned}
P((\Theta_t)_{t\in\mathbb{Z}} \in D) &= E\Big[\mathbb{1}_D((\Theta_t)_{t\in\mathbb{Z}})\Big] = E\Big[\mathbb{1}_D((U^{RS}_t)_{t\in\mathbb{Z}})\Big] \\
&= E\Bigg[\sum_{k\in\mathbb{Z}} \frac{\|U_k\|}{\|U\|_1^1}\mathbb{1}_D\Bigg(\Bigg(\frac{U_{t+k}}{\|U_k\|}\Bigg)_{t\in\mathbb{Z}}\Bigg)\Bigg] \\
&= E\Bigg[\frac{\|U_0\|}{\|U_0\|+\|U_1\|}\mathbb{1}_D\Bigg(\Bigg(\frac{U_t}{\|U_0\|}\Bigg)_{t\in\mathbb{Z}}\Bigg) + \frac{\|U_1\|}{\|U_0\|+\|U_1\|}\mathbb{1}_D\Bigg(\Bigg(\frac{U_{t+1}}{\|U_1\|}\Bigg)_{t\in\mathbb{Z}}\Bigg)\Bigg]
\end{aligned}$$

for all Borel sets D in $\mathbb{R}^d$. For the specific choice $D = \{(y_t)_{t\in\mathbb{Z}} | y_{-1} = a^{-1}, y_0 = -1, y_t = 0\,\forall t \notin \{-1,0\}\}$, a direct evaluation of the above discrete expectation yields

$$\begin{aligned}
P\left(\Theta_{-1} = a^{-1}, \Theta_0 = -1, \Theta_t = 0\,\forall t \notin \{-1,0\}\right) &= (1-p)\cdot(0+0) + p\cdot\left(0 + \frac{1}{1+a^{-1}}\right) \\
&= p\frac{a}{a+1} =: p_1.
\end{aligned}$$

Analogously,

$$\begin{aligned}
&P\left(\Theta_0 = 1, \Theta_1 = -a, \Theta_t = 0\,\forall t \notin \{0,1\}\right) = p\frac{1}{a+1} =: p_2, \\
&P\left(\Theta_{-1} = b, \Theta_0 = 1, \Theta_t = 0\,\forall t \notin \{-1,0\}\right) = (1-p)\frac{1}{1+b} =: p_3, \\
&P\left(\Theta_0 = 1, \Theta_1 = b^{-1}, \Theta_t = 0\,\forall t \notin \{0,1\}\right) = (1-p)\frac{b}{1+b} =: p_4.
\end{aligned}$$

In this example, we want to estimate $P(\Theta_i \in A)$ for lags $i \in \{-1,0,1\}$ and some sets $A = (x,\infty)$ such that $P(\Theta_i = x) = 0$. More precisely, we consider the half lines (ε,∞) and $(1-\varepsilon,\infty)$ for some $\varepsilon > 0$ such that $\varepsilon < a^{-1} \wedge b^{-1}$ and $1-\varepsilon > a^{-1} \vee b^{-1}$. We

consider the probabilities $p_A := P(\Theta_0 > \varepsilon) = 1 - p_1$, $p_B := P(\Theta_1 > \varepsilon) = p_4$ and $p_C := P(\Theta_{-1} > 1 - \varepsilon) = p_3$ and the resulting estimators. In slight abuse of notation, we set $A = B = (\varepsilon, \infty)$ and $C = (1 - \varepsilon, \infty)$, where the use of A indicates that $i = 0$, the use of B indicates $i = 1$ and the use of C indicates $i = -1$.
The asymptotic variance of $\hat{\hat{p}}_{n,A}$ can be calculated using the representation (5.3.1), which can be slightly simplified for variances to

$$\begin{aligned} Var(Z^{pb,\alpha}(A)) =& Var(Z(A)) - 2(p_A - \alpha d_A) Cov(Z(A), Z(\mathbb{R})) \\ &- 2\alpha^2 d_A Cov(Z(A), Z(\phi)) + (p_A - \alpha d_A)^2 Var(Z(\mathbb{R})) \\ &+ 2(p_A - \alpha d_A)\alpha^2 d_A Cov(Z(\mathbb{R}), Z(\phi)) + \alpha^4 d_A^2 Var(Z(\phi)), \end{aligned}$$

with $((Z(A))_{A\in\mathcal{A}}, Z(\phi))$ and d_A given in Theorem 5.3.1; note that we are in the real valued case $d = 1$. Applying Lemma 5.2.4 and Lemma 5.7.7, the single terms in this representation can be directly calculated for our example as discrete expectations, since $(\Theta_t)_{t\in\mathbb{Z}}$ vanishes for lag $|t| \geq 1$ and the distribution is discrete with just four points of mass:

$$\begin{aligned}
Var(Z(A)) &= E\Big[\sum_{j\in\mathbb{Z}}(\|\Theta_j\|^\alpha \wedge 1)\Big(\sum_{h\in\mathbb{Z}}\frac{\|\Theta_h\|}{\|\Theta\|_\alpha^\alpha}\mathbb{1}_A\Big(\frac{\Theta_{h+i}}{\|\Theta_h\|}\Big)\Big)\Big(\sum_{l\in\mathbb{Z}}\frac{\|\Theta_l\|}{\|\Theta\|_\alpha^\alpha}\mathbb{1}_A\Big(\frac{\Theta_{l+i}}{\|\Theta_l\|}\Big)\Big)\Big] \\
&= p_1\Big(1+\frac{1}{a}\Big)\Big(\frac{1}{1+a}\Big)^2 + 2p_2\Big(\frac{1}{1+a}\Big)^2 + 2p_3 + p_4\Big(1+\frac{1}{b}\Big), \\
Var(Z(B)) &= 2p_3\Big(\frac{b}{1+b}\Big)^2 + p_4\Big(1+\frac{1}{b}\Big)\Big(\frac{b}{1+b}\Big)^2, \\
Var(Z(C)) &= 2p_3\Big(\frac{1}{1+b}\Big)^2 + p_4\Big(1+\frac{1}{b}\Big)\Big(\frac{1}{1+b}\Big)^2, \\
Var(Z(\mathbb{R})) &= p_1\Big(1+\frac{1}{a}\Big) + 2p_2 + 2p_3 + p_4\Big(1+\frac{1}{b}\Big), \\
Cov(Z(A), Z(\mathbb{R})) &= p_1\Big(1+\frac{1}{a}\Big)\frac{1}{1+a} + 2p_2\frac{1}{1+a} + 2p_3 + p_4\Big(1+\frac{1}{b}\Big), \\
Cov(Z(B), Z(\mathbb{R})) &= 2p_3\frac{b}{1+b} + p_4\Big(1+\frac{1}{b}\Big)\frac{b}{1+b}, \\
Cov(Z(C), Z(\mathbb{R})) &= 2p_3\frac{1}{1+b} + p_4\Big(1+\frac{1}{b}\Big)\frac{1}{1+b}, \\
Cov(Z(A), Z(\Phi)) &= E\Big[\sum_{j\in\mathbb{Z}}(\|\Theta_j\|^\alpha \wedge 1)(\log(\|\Theta_j\| \vee 1) + \alpha^{-1})\Big(\sum_{h\in\mathbb{Z}}\frac{\|\Theta_h\|}{\|\Theta\|_\alpha^\alpha}\mathbb{1}_A\Big(\frac{\Theta_{h+i}}{\|\Theta_h\|}\Big)\Big)\Big] \\
&= p_1\Big(\frac{1}{a}+1\Big)\frac{1}{a+1} + p_2(2+\log(a))\frac{1}{1+a} + p_3(2+\log(b)) + p_4\Big(1+\frac{1}{b}\Big), \\
Cov(Z(B), Z(\phi)) &= p_3(2+log(b))\frac{b}{b+1} + p_4\Big(1+\frac{1}{b}\Big)\frac{b}{b+1}, \\
Cov(Z(C), Z(\phi)) &= p_3(\log(b)+2)\frac{1}{b+1} + p_4\Big(1+\frac{1}{b}\Big)\frac{1}{b+1}, \\
Cov(Z(\mathbb{R}), Z(\phi)) &= p_1\Big(\frac{1}{a}+1\Big) + p_2(\log(a)+2) + p_3(\log(b)+2) + p_4\Big(1+\frac{1}{b}\Big), \\
Var(Z(\Phi)) &= \alpha^{-1}\sum_{k\in\mathbb{Z}} E\left[(1 \wedge \|\Theta_k\|^\alpha)(|\log(\|\Theta_k\|)| + 2\alpha^{-1})\right]
\end{aligned}$$

$$= p_1\Big(\frac{1}{a}(\log(a)+2)+2\Big) + p_2(4+\log(a)) + p_3(\log(b)+4) + p_4\Big(2+\frac{1}{b}(\log(b)+2)\Big),$$

$$d_A = -E\Big[\sum_{j\in\mathbb{Z}} \log(\|\Theta_j\|)\|\Theta_j\|^\alpha \|\Theta_j\|_\alpha^{-\alpha} \mathbb{1}_A(\Theta_i)\Big]$$
$$= -\Big(p_1\cdot 0 + p_2\frac{\log(a)a}{1+a} + p_3\frac{\log(b)b}{1+b} - p_4\frac{\log(b)}{1+b}\Big) = -p_2\frac{\log(a)a}{1+a},$$
$$d_B = p_4\frac{\log(b)}{b+1},$$
$$d_C = -p_3\frac{\log(b)b}{1+b},$$

where $d_A = -p_2\log(a)a/(1+a)$ holds, since $p_4 = b\cdot p_3$. From this, one can directly calculate the asymptotic variance of $\hat{\hat{p}}_{n,A}$ for all three sets A, B, C.
For the calculation of the covariance of the backward estimator $\hat{\hat{p}}^b_{n,A}$ with estimated α, we denote $Z(\phi^t_{2,x})$ from Theorem 3.1 of Davis et al. (2018) by $\tilde{Z}(A)$. Then, using the results of Theorem 3.1 and Proposition 6.1 from Davis et al. (2018), we can calculate the variance $c^{b,\alpha}(A)$ of the backward estimator $\hat{\hat{p}}^b_{n,A}$ by

$$\begin{aligned} c^{b,\alpha}(A) &:= Var(\tilde{Z}(A) - p_A Z(\mathbb{R}) + (\alpha^2 Z(\phi) - \alpha Z(\mathbb{R}))e_A) \\ &= Var(\tilde{Z}(A)) + (e_A+p_A)^2 Var(Z(\mathbb{R})) + e_A^2 Var(Z(\Phi)) \\ &\quad + 2e_A Cov(\tilde{Z}(A), Z(\Phi)) - 2e_A(e_A+p_A)Cov(Z(\Phi), Z(\mathbb{R})) \\ &\quad - 2(e_A+p_A)Cov(\tilde{Z}(A), Z(\mathbb{R})) \end{aligned}$$

with $e_A := E[\log(\|\Theta_i\|)\mathbb{1}_{\{\Theta_i\in A\}}]$ and where $(\tilde{Z}, Z(\phi), Z(\mathbb{R}))$ is a centered Gaussian process with covariance given below and by Lemma 5.7.7. For our sets A, B and C one has

$$\begin{aligned} e_A &= E[\log(\|\Theta_0\|)\mathbb{1}_A(\Theta_0)] = (p_1+p_2+p_3+p_4)\cdot 0 = 0, \\ e_B &= E[\log(\|\Theta_1\|)\mathbb{1}_B(\Theta_1)] = (p_1+p_2+p_3)\cdot 0 + p_4\log\Big(\frac{1}{b}\Big) = -p_4\log(b), \\ e_C &= E[\log(\|\Theta_{-1}\|)\mathbb{1}_C(\Theta_{-1})] = p_3\log(b). \end{aligned}$$

Since $e_A = 0$, we do not need to calculate $Cov(\tilde{Z}(A), Z(\phi))$. All the remaining single covariances in the representation of the variance of $\hat{\hat{p}}^b_{n,A}$ can be calculated by equation (6.3) of Davis et al. (2018) as follows:

$$\begin{aligned} Var(\tilde{Z}(B)) &= E\Big[\sum_{j\in\mathbb{Z}}(\|\Theta_j\|^\alpha\wedge 1)\Big(\frac{\|\Theta_{j-i}\|^\alpha}{\|\Theta_j\|^\alpha}\mathbb{1}_B\Big(\frac{\Theta_j}{\|\Theta_{j-i}\|}\Big)\Big)\Big(\|\Theta_{-i}\|^\alpha\mathbb{1}_B\Big(\frac{\Theta_0}{\|\Theta_{-i}\|}\Big)\Big)\Big] \\ &= p_3 b^2, \\ Var(\tilde{Z}(A)) &= p_2 + 2p_3 + p_4\Big(1+\frac{1}{b}\Big), \\ Var(\tilde{Z}(C)) &= p_4\Big(\frac{1}{b}\Big)^2, \end{aligned}$$

$$
\begin{aligned}
Cov(\tilde{Z}(A), Z(\mathbb{R})) &= E\Big[\sum_{j\in\mathbb{Z}}(\|\Theta_j\|^\alpha \wedge 1)\Big(\|\Theta_{-i}\|^\alpha \mathbb{1}_A\Big(\frac{\Theta_0}{\|\Theta_{-i}\|}\Big)\Big)\Big]\\
&= 2p_2 + 2p_3 + p_4\Big(1+\frac{1}{b}\Big),\\
Cov(\tilde{Z}(B), Z(\mathbb{R})) &= 2p_3 b,\\
Cov(\tilde{Z}(C), Z(\mathbb{R})) &= p_4\Big(1+\frac{1}{b}\Big)\frac{1}{b},\\
Cov(\tilde{Z}(B), Z(\phi)) &= E\Big[\sum_{j\in\mathbb{Z}}(\|\Theta_j\|^\alpha \wedge 1)(\log(\|\Theta_j\| \vee 1)+\alpha^{-1})\|\Theta_{-i}\|^\alpha \mathbb{1}_B\Big(\frac{\Theta_0}{\|\Theta_{-i}\|}\Big)\Big]\\
&= p_3(\log(b)+2)b,\\
Cov(\tilde{Z}(C), Z(\phi)) &= p_4\Big(1+\frac{1}{b}\Big)\frac{1}{b}.
\end{aligned}
$$

From this, one can directly calculate the asymptotic variance $c^{b,\alpha}(A)$ of $\hat{\hat{p}}^b_{n,A}$.
The asymptotic variance $c^f(A)$ of the forward estimator $\hat{p}^f_{n,A}$ can be obtained by Davis et al. (2018), Theorem 3.1. (There, the variance was calculated for sets $\tilde{A} = (-\infty, x]$ with $x \in \mathbb{R}$ and under suitable conditions,which are all satisfied under our conditions (PR), (P0), (PP) and (PC). However, the principle is the same for the sets considered here.) Hence, we can state the variance as

$$
c^f(A) = \sum_{j\in\mathbb{Z}} E\Big[(\|\Theta_j\|^\alpha \wedge 1)\Big(p_A - \mathbb{1}_A\Big(\frac{\Theta_{j+i}}{\|\Theta_j\|}\Big)\Big)(p_A - \mathbb{1}_A(\Theta_i))\Big].
$$

For our concrete sets A, B, C this can be specified as

$$
\begin{aligned}
c^f(A) &= p_1\Big(\frac{1}{a}(p_A-1)p_A + p_A^2\Big) + p_2((p_A-1)^2 + p_A(p_A-1)) + 2p_3(p_A-1)^2\\
&\quad + p_4\Big((p_A-1)^2 + \frac{1}{b}(p_A-1)^2\Big),\\
c^f(B) &= p_1\Big(\frac{1}{a}p_B^2 + p_B^2\Big) + p_2(p_B^2+p_B^2) + p_3((p_B-1)p_B + p_B^2),\\
c^f(C) &= p_1\Big(\frac{1}{a}+1\Big)p_C^2 + 2p_2p_C^2 + p_3(p_C(p_C-1) + (p_C-1)^2) + p_4\Big(p_C^2 + \frac{1}{b}(p_C-1)p_C\Big).
\end{aligned}
$$

Figure 5.1 shows the variances of $\hat{\hat{p}}_{n,A}$, $\hat{p}^f_{n,A}$ and $\hat{\hat{p}}^b_{n,A}$ for our sets A, B, and C with corresponding lags $i \in \{-1,0,1\}$ and the fixed parameters $a = 10$, $b = 2$ as a function of the parameter $p \in [0,1]$. One can observe that there is no uniform smallest variance for all possible choices of the parameter. In fact, this figure shows that each of the three estimators can have the largest or smallest asymptotic variance in this model, depending on the model parameter p (and also a, b) and the probability we want to estimate (i.e. depending on i and x). For the middle plot, observe that the backward estimator is equal to the forward estimator for lag 0. Thus, there is no winner who has uniformly smallest variance and, thereby, non of the three estimators is uniformly most efficient.
If one considers the estimators in the same model with α assumed known as considered

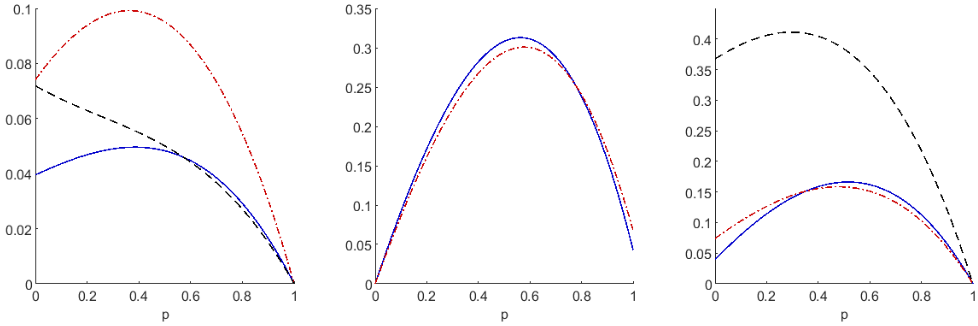

***Figure 5.1:** Asymptotic variances of $\hat{\hat{p}}_{n,A}$, $\hat{p}^f_{n,A}$ and $\hat{\hat{p}}^b_{n,A}$ with unknown α*

Variances of $\hat{\hat{p}}_{n,A}$ (blue solid line), $\hat{p}^f_{n,A}$ (red dashed-dotted line) and $\hat{\hat{p}}^b_{n,A}$ (black dashed line) with parameters $a = 10$, $b = 2$ plotted as a function in $p \in [0, 1]$ for lag $i = -1$ and the set $C = (1 - \varepsilon, \infty)$ (left), $i = 0$, $A = (\varepsilon, \infty)$ (middle) and $i = 1$, $B = (\varepsilon, \infty)$ (right).

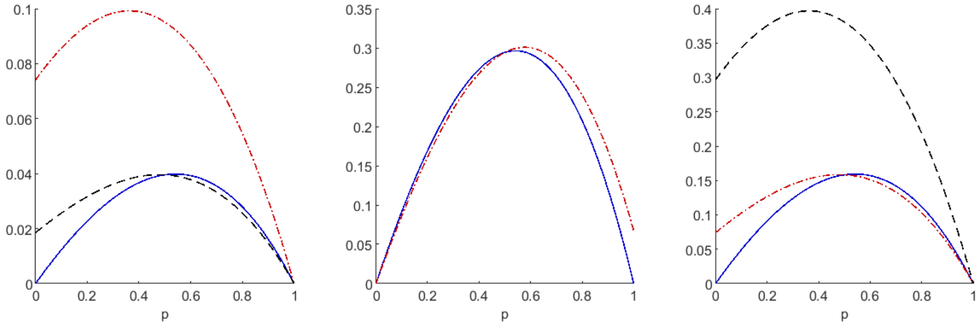

***Figure 5.2:** Asymptotic variances of $\hat{p}_{n,A}$, $\hat{p}^f_{n,A}$ and $\hat{p}^b_{n,A}$ with known α*

Variances of $\hat{p}_{n,A}$ (blue solid line), $\hat{p}^f_{n,A}$ (red dashed-dotted line) and $\hat{p}^b_{n,A}$ (black dashed line) with parameters $a = 10$, $b = 2$ plotted as a function in $p \in [0, 1]$ for lag $i = -1$ and the set $C = (1 - \varepsilon, \infty)$ (left), $i = 0$, $A = (\varepsilon, \infty)$ (middle) and $i = 1$, $B = (\varepsilon, \infty)$ (right).

in Section 5.2, then the results remain qualitively unchanged. For the short comparison between the variance $c^{pb}(A)$ of $\hat{p}_{n,A}$, the variance c^f of $\hat{p}^f_{n,A}$ and the variance c^b of $\hat{p}^b_{n,A}$ with known α, note that the forward estimator is unaffected, since it does not depend on α. Thus, c^f remains unchanged as in the case of the estimated α as discussed before. For the calculation of the variances of $\hat{p}_{n,A}$ and $\hat{p}^b_{n,A}$ one has simply to replace d_A and e_A by 0 (and likewise set d_B, e_B, d_C and e_C to 0) in the formulas above. This is in accordance with Theorem 5.2.9 and Theorem 3.1 of Davis et al. (2018).
Figure 5.2 shows the variances of the estimators $\hat{p}_{n,A}$, $\hat{p}^f_{n,A}$ and $\hat{p}^b_{n,A}$ with known α for the selected parameters $a = 10$ and $b = 2$ as functions of $p \in [0, 1]$ and for our sets (x, ∞) and lag $i \in \{-1, 0, 1\}$. Again, there is no estimator with uniformly smallest variance for all possible choices of the parameter. In fact, depending on the parameter p each of the three estimators $\hat{p}_{n,A}$, $\hat{p}^f_{n,A}$ and $\hat{p}^b_{n,A}$ can have the smallest or the largest asymptotic variance.
One can observe in Figure 5.2 that for $p \to 0$ or $p \to 1$, the asymptotic variance of the estimator $\hat{p}_{n,A}$ converges to 0. This is because $\hat{p}_{n,A}$ has an advantage over the other estimators

when the shape of $(\Theta_t^*)_{t\in\mathbb{Z}}$ has little variation and the shape $(U_t)_{t\in\mathbb{Z}}$ is deterministic for $p \in \{0,1\}$, see also the example on page 152. However, the asymptotic variance of $\hat{\hat{p}}_{n,A}$ needs not vanish due to the remaining variability of the Hill estimator (cf. Figure 5.1).
Even in this simple case of just 4 possible states for the spectral tail process, one can choose for each estimator the parameters such that it has the smallest asymptotic variance. This example shows, that in general it cannot be stated whether the new projection based estimator $\hat{\hat{p}}_{n,A}$ is better or worse than $\hat{p}^f_{n,A}$ or $\hat{\hat{p}}^b_{n,A}$ in terms of a uniform smallest asymptotic variance, or even whether one of these three estimators is always better than at least one of the others.
For more complicated models than those considered, it is in general difficult to calculate and compare the asymptotic variances of $\hat{\hat{p}}_{n,A}$ with that of $\hat{p}^f_{n,A}$ or $\hat{p}^b_{n,A}$. However, note that a comparison of the asymptotic variances $\hat{p}^f_{n,A}$ or $\hat{p}^b_{n,A}$ is possible e.g. for Markovian processes. For such processes $\hat{p}^b_{n,A}$ has always a smaller asymptotic variance than $\hat{p}^f_{n,A}$, see Drees et al. (2015), Remark 4.2.
In Section 5.6 we will study the finite sample performance of the three estimators (with estimated α), to find out whether one estimator is advantageous even if this can not be seen in the asymptotic variances.

5.5.2 Stochastic recurrence equations

In this section, we want to consider an example for which we verify the conditions of the previous theorems about the asymptotic normality of the estimation errors from $\hat{p}_{n,A}$ and $\hat{\hat{p}}_{n,A}$. In particular, this example will show that the conditions can be satisfied by a reasonable class of time series models. We focus on stationary solutions to stochastic recurrence equations

$$X_t = C_t X_{t-1} + D_t, \qquad \forall\, t \in \mathbb{Z}, \tag{5.5.1}$$

where C_t are random $d \times d$-matrices with non-negative entries and D_t are $[0,\infty)^d$-valued random vectors such that $(C_t, D_t) \in [0,\infty)^{d\times d} \times [0,\infty)^d$, $t \in \mathbb{Z}$, are iid.
Such stochastic recurrence equations are often considered in the literature in various settings, they can be used e.g. to analyze GARCH time series (cf. Basrak et al. (2002)), which are popular models for financial time series. Some other possible applications for processes which solve a stochastic recurrence equation can be found e.g. in Vervaat (1979).
We consider the euclidean norm, i.e. the L_2-norm $\|\cdot\|$ as vector norm on $\mathbb{R}^d$. In addition, we consider the operator norm $\|\cdot\|_M$ of $\|\cdot\|$ as matrix norm on $\mathbb{R}^{d\times d}$. The operator norm is defined by $\|A\|_M = \sup_{\|x\|=1} \|Ax\|$ and is submultiplicative (i.e. $\|A \cdot B\|_M \le \|A\|_M \|B\|_M$ for $A, B \in \mathbb{R}^{d\times d}$) as well as compatible with $\|\cdot\|$ (i.e. $\|Ax\| \le \|A\|_M \|x\|$ for $A \in \mathbb{R}^{d\times d}$, $x \in \mathbb{R}^d$).

Existence of stationary, regularly varying solutions

Under the following conditions there exists a unique stationary solution to (5.5.1) and this solution is regularly varying with index α:

(SRE1) (i) The top Lyapunov exponent $\gamma := \inf_{n\in\mathbb{N}}\{1/nE[\log(\|C_n\cdots C_1\|_M)]\} < 0$ is negative;

(ii) there exists an $s > 0$ such that $h(s) := \lim_{n\to\infty}\left(E[\|C_n\cdots C_1\|_M^s]\right)^{1/n} \in (1,\infty)$;

(iii) $E[\|C_1\|_M^\alpha \log^+(\|C_1\|_M)] < \infty$, $E[\|D_1\|^\alpha] < \infty$ with $\alpha > 0$ denoting the unique solution to $h(\alpha) = 1$;

(iv) $P(D_1 = 0) < 1$, $P(((C_1)_{jk})_{1\le k\le d} \neq 0\,\forall\, 1 \le j \le d) = 1$;

(v) the additive subgroup generated by the logarithms of the spectral radii of arbitrary finite products of matrices in the support of P^{C_1} is dense in $\mathbb{R}$.

Condition (SRE1) (iii) implies $E[\log^+(\|C_1\|_M)] < \infty$ and $E[\log^+(\|D_1\|)] < \infty$, the latter holds since $\log^+(x) \le c_q x^q$ for $q > 0$ and some suitable constant $c_q > 0$. Thus, if (SRE1) (i) and (iii) hold, then (5.5.1) has a unique stationary solution $(X_t)_{t\in\mathbb{Z}}$ (cf. Basrak et al. (2002), Theorem 2.1). The top Lyapunov exponent γ for a sequence of random $d\times d$-matrices $(C_n)_{n\in\mathbb{N}}$ also has the representation $\gamma = \lim_{n\to\infty} 1/n\log(\|C_n\cdots C_1\|_M)$ a.s.
Under condition (SRE1) (i) and (iii) the solution to (5.5.1) admits the representation

$$X_t = D_t + \sum_{k=1}^{\infty}\prod_{j=t-k+1}^{t} C_j D_{j-k},$$

which in particular implies that $X_t \in [0,\infty)^d$ a.s. for all $t\in\mathbb{Z}$, since the components of C_t and D_t are non-negative. Denote

$$\Pi_{j,k} := C_k\cdot C_{k-1}\cdots C_j \qquad \text{and} \qquad R_k := \sum_{j=1}^{k}\Pi_{j+1,k}D_j.$$

Then, $(R_k, \Pi_{1,k})$ and X_0 are independent and it is $X_k = R_k + \Pi_{1,k}X_0$ for all $k\in\mathbb{N}$.
Under (SRE1) (i) and (ii) there exists an unique $\alpha > 0$ with $h(\alpha) = 1$ as required for condition (iii) (cf. Buraczewski et al. (2016), Lemma 4.4.2). The subgroup defined in (v) is specified in condition (A) of Buraczewski et al. (2016), page 171. Buraczewski et al. (2016), Theorem 4.4.5, shows that by (SRE1) (5.5.1) has a unique stationary solution $(X_t)_{t\in\mathbb{Z}}$ and X_0 is multivariate regularly varying with index α. Moreover, iteration of (5.5.1) and the regular variation of X_0 gives regular variation of $(X_t)_{t\in\mathbb{Z}}$ (cf. Buraczewski et al. (2016), Corollary 4.4.6). In particular, our condition (PR) is satisfied.
Note that the conditions (iv) and (v) are readily implied if the distribution of (C_1, D_1) is absolutely continuous. However, for sake of generality we assume the more technical conditions rather than the absolute continuity.

According to Janssen and Segers (2014), Theorem 2.1 and Example 6.1, the distribution of the forward spectral tail process $(\Theta_t)_{t\in\mathbb{N}}$ of $(X_t)_{t\in\mathbb{Z}}$ admits the representation

$$\Theta_t = \Theta_0 \prod_{j=1}^{t} \tilde{C}_j \tag{5.5.2}$$

for all $t > 0$, where $\tilde{C}_j$, $j \in \mathbb{N}$, are iid random variables with the same distribution as C_1 and Θ_0 is independent of $(\tilde{C}_j)_{j\in\mathbb{N}}$. The distribution of the backward spectral tail chain $(\Theta_{-t})_{t\in\mathbb{N}}$ is determined by this distribution of the forward spectral tail process, see Theorem 2.1.8.
We want to verify the conditions for Theorems 5.2.6, 5.2.9 and 5.3.1 with $\mathcal{A} = \{[0,x] = [0,x_1] \times ... \times [0,x_d] \mid x = (x_1,...,x_d) \in [0,\infty)^d\} \cup \{\emptyset, [0,\infty)^d\}$, which is distribution determining for Θ_i, $i \geq 0$.

General Markov theory as preparation

For the verification of these conditions we will apply some arguments from general Markov theory. For this we need that the unique solution $(X_t)_{t\in\mathbb{Z}}$ to (5.5.1) is an aperiodic and irreducible Markov process, which holds under the following assumptions:

(SRE2) (i) The interior of the support of P^{X_0} is non-empty;

(ii) there exists a σ-finite non-null measure ν on $\mathbb{R}^d$ and an open set $E \subset \mathbb{R}^d$ with $P(X_0 \in E) > 0$ such that $P^{C_1 x + D_1}$ has an absolutely continuous component with respect to ν for all $x \in E$.

If (SRE1) and (SRE2) are satisfied, then $(X_t)_{t\in\mathbb{Z}}$ is an aperiodic, positive Harris recurrent P^{X_0}-irreducible Feller process. This holds by Theorems 2.1, 2.2 and Corollary 2.3 of Alsmeyer (2003), since the conditions of these theorems are satisfied by Example 2.6 and Remark C of that reference. See also Proposition 4.2.1 of Buraczewski et al. (2016) and note for this proposition that the condition $P(C_1 x + D_1 = x) < 1$ for all $x \in \mathbb{R}^d$ is not needed since we consider only non-negative random variables (cf. Buraczewski et al. (2016), p.170).
Condition (SRE2) (ii) is in particular satisfied if (C_1, D_1) is absolutely continuous, see Lemma 4.2.2 of Buraczewski et al. (2016). The same lemma states alternative conditions to verify (SRE2) (ii).
We are going to apply general results for Markovian time series established by Kulik et al. (2019). To this end, we verify the Assumption 2.1 of that paper with a fixed $q \in (0,\alpha)$, $\mathbb{Y} := X$, $g(x) = \|x\|$, $V(x) = \|x\|^q + 1$ and $q_0 = q$. Due to the discussion above, $(X_t)_{t\in\mathbb{Z}}$ is a stationary, regularly varying Markov process and (i) and (ii) of Assumption 2.1 are directly satisfied. Part (v) of that assumption is obviously fulfilled for our g and V with some $c \geq 1$.

For the further theoretical Markov arguments, the existence of some specific small set is important. Mikosch and Wintenberger (2014) mentioned on page 161 that the set $\{x \in \mathbb{R}^d : \|x\|^q \leq M\}$ is small for some $M > 0$. This can be reasoned as follows: Due to (SRE2) (i) and since $(X_t)_{t\in\mathbb{Z}}$ is an irreducible Feller process, Meyn and Tweedie (1992), Theorem 3.4, implies that all compact sets are petite. Therefore, $\{x \in \mathbb{R}^d : \|x\|^q \leq M\}$ is petite as compact set and by Theorem 9.4.10 of Douc et al. (2018) it is small for all $M \geq 0$. By Proposition 9.2.13 of Douc et al. (2018) (applied with $V_0(x) = V_1(x) = \|x\|^q$) there exists a $M_0 \in \mathbb{N}$ such that $\{x \in \mathbb{R}^d : \|x\|^q \leq M_0\}$ is accessible. When $P(x, \cdot)$ denotes the Markov kernel associated to the Markov process $(X_t)_{t\in\mathbb{Z}}$, we denote with $P^m(x, \cdot)$ the Markov kernel of the m-skeleton (m-step Markov process). By Theorem 9.3.11 of Douc et al. (2018) the set $\{x \in \mathbb{R}^d : \|x\|^q \leq M_0\}$ is small and accessible by $P^m(x, \cdot)$. Corollary 14.1.6. of Douc et al. (2018) states equivalent conditions for sets being petite, part (ii) of that corollary is exactly assumption (iv) we want to verify and part (i) of that corollary is the formulation of the petite set which exists due to discussion above. Hence, part (iv) of Assumption 2.1 of Kulik et al. (2019) is fulfilled (with $m = 1$ or arbitrary $m \in \mathbb{N}$).

Mikosch and Wintenberger (2014), Section 5.2, showed that due to $h(\alpha) = 1$, the drift condition (DC_p) of this paper is satisfied for the d-dimensional solution of (5.5.1), using the function $\tilde{V}(x) = \|x\|^q$, for some $q \in (0, \alpha)$. We will check this drift condition as follows: Recall $X_k = \Pi_{1,k}X_0 + R_k$ for all $k \geq 0$. By $h(\alpha) = 1$ and the Jensen inequality it directly holds that $h(q) < 1$ for all $q \in (0, \alpha)$ (alternatively, h is a strictly convex function with $h(\alpha) = h(0) = 1$, this implies $h(q) < 1$ for all $q \in (0, \alpha)$), so that

$$\kappa := E[\|\Pi_{1,m}\|^q] < 1 \tag{5.5.3}$$

for m sufficiently large. Check that $(a + b)^q \leq ((1 + \eta)a)^q \mathbb{1}_{\{b\leq\eta a\}} + ((1 + \eta)b/\eta)^q \mathbb{1}_{\{b>\eta a\}}$ for all $a, b, \eta \geq 0$. Using this, we may conclude

$$\begin{aligned}
E[\|X_m\|^q \mid X_0 = y] &= E[\|\Pi_{1,m}X_0 + R_m\|^q \mid X_0 = y] \\
&= E[\|\Pi_{1,m}y + R_m\|^q] \\
&\leq E[(\|\Pi_{1,m}\|_M\|y\| + \|R_m\|)^q] \\
&\leq (1 + \eta)^q E[\|\Pi_{1,m}\|_M^q]\|y\|^q + ((1 + \eta)/\eta)^q E[\|R_m\|^q] \\
&= (1 + \eta)^q \kappa^q \|y\|^q + ((1 + \eta)/\eta)^q E[\|R_m\|^q] \\
&=: \tilde{\beta}\|y\|^q + b
\end{aligned} \tag{5.5.4}$$

with $\tilde{\beta} = (1 + \eta)^q \kappa^q < 1$ for sufficiently small $\eta > 0$ and $b = ((1 + \eta)/\eta)^q E[\|R_m\|^q] < \infty$. The last inequality holds due to $E[\|D_1\|^q] < \infty$ and $E[\|C_1\|_M^q] < \infty$, which holds by (SRE1) (iii) and $q \in (0, \alpha)$. According to the discussion above, $\{x \in \mathbb{R}^d : \|x\|^q \leq M_0\}$ is small and accessible for $P^m(x, \cdot)$ and by Theorem 14.1.4 of Douc et al. (2018) an analog to (5.5.4) with b replaced by $b\mathbb{1}_{\{x\in\mathbb{R}^d:\|x\|^q\leq M_0\}}$ and $\tilde{\beta}$ possibly replaced by some constant $\beta \in (0, 1)$ holds. Thus, part (a) of ($\text{DC}_{q,m}$) of Mikosch and Wintenberger (2014) holds (see also the remark before Lemma 2.1 in that paper). Furthermore, similarly we have

$$E[\|X_1\|^q|X_0 = y] \le (1+\eta)^q E[\|C_1\|_M^q]\|y\|^q + ((1+\eta)/\eta)^q E[\|D_1\|^q].$$

Thus, the condition ($\mathrm{DC}_{p,m}$) of Mikosch and Wintenberger (2014) is satisfied and by Lemma 2.1 of this paper the drift condition (DC_q) holds, i.e.

$$E[\|X_1\|^q|X_0 = y] \le \beta\|y\|^q + b'$$

for all $y \in \mathbb{R}^d$ and some $\beta \in (0,1)$ and $b' > 0$. Thus,

$$E[V(X_1) \mid X_0 = y] = E[\|X_1\|^q + 1 \mid X_0 = y] \le \beta(\|y\|^q + 1) + \tilde{b} = \beta V(y) + \tilde{b} \tag{5.5.5}$$

for some $\tilde{b} > 0$, which is why the drift condition in part (iii) of Assumption 2.1 of Kulik et al. (2019) holds.

It remains to check part (vi) of Assumption 2.1 of Kulik et al. (2019). Regular variation implies the weak convergence $\mathcal{L}(X_0/u_n \mid \|X_0\| > u_n) \xrightarrow{w} \mathcal{L}(Y_0)$ and with the continuous mapping theorem $\mathcal{L}((\|X_0\|/u_n)^q \mid \||X_0\| > u_n) \xrightarrow{w} \mathcal{L}((\|Y_0\|)^q)$ follows. Due to the regular variation of the time series $(X_t)_{t\in\mathbb{Z}}$, using the Potter bounds (see Theorem 2.1.2) we obtain $P((\|X_0\|/u_n)^q > x)/P(\|X_0\| > u_n) \le (1+\varepsilon)x^{-(\alpha-\varepsilon)/q}$ for $x > 1$, $\varepsilon > 0$ and sufficiently large n. Thus, by stationarity, for $h \in \mathbb{Z}$ and sufficiently large n it holds that

$$\begin{aligned}
E\Big[\Big(\frac{\|X_h\|}{u_n}\Big)^q \,\Big|\, \|X_0\| > u_n\Big] &\le \frac{1}{v_n} E\Big[\Big(\frac{\|X_h\|}{u_n}\Big)^q \mathbb{1}_{\{\|X_h\|>u_n\}}\mathbb{1}_{\{\|X_0\|>u_n\}}\Big] + 1 \\
&\le \frac{1}{v_n} E\Big[\Big(\frac{\|X_0\|}{u_n}\Big)^q \mathbb{1}_{\{\|X_0\|>u_n\}}\Big] + 1 \\
&= E\Big[\Big(\frac{\|X_0\|}{u_n}\Big)^q \,\Big|\, \|X_0\| > u_n\Big] + 1 \\
&= \int_0^\infty P\Big(\Big(\frac{\|X_0\|}{u_n}\Big)^q > x \,\Big|\, \|X_0\| > u_n\Big) dx + 1 \\
&\le \int_1^\infty (1+\varepsilon)x^{-(\alpha-\varepsilon)/q} dx + 2 < \infty
\end{aligned} \tag{5.5.6}$$

for $q < \alpha - \varepsilon$. In particular, $\sup_{n\in\mathbb{N}} E\left[(\|X_h\|/u_n)^q \,|\|X_0\| > u_n\right] < \infty$ for all $q \in (0,\alpha)$, i.e. $(\|X_h\|/u_n)^q \mathbb{1}_{\{\|X_0\|>u_n\}}/P(\|X_0\| > u_n)$, $n \in \mathbb{N}$, is uniform integrable. (This could also be shown with Karamata's theorem.) The weak convergence together with the uniform integrability imply the convergence of the expected values

$$E\Big[\Big(\frac{\|X_h\|}{u_n}\Big)^q \,\Big|\, \|X_0\| > u_n\Big] \to E[\|Y_h\|^q]. \tag{5.5.7}$$

This implies that for all $h \in \mathbb{Z}$ there exist an n_0 such that for all $n > n_0$ it holds

$$E\Big[\Big(\frac{\|X_h\|}{u_n}\Big)^q \mathbb{1}_{\{\|X_0\|>u_n\}}\Big] \le 2E[\|Y_h\|^q]P(\|X_0\| > u_n).$$

This bound for $h = 0$ and again the Potter bounds imply

$$E[(\|X_0\|^q + 1)\mathbb{1}_{\{\|X_0\|>su_n\}}] = E[\|X_0\|^q \mathbb{1}_{\{\|X_0\|>su_n\}}] + P(\|X_0\| > su_n)$$
$$\leq 2s^q u_n^q E[\|Y_0\|^q] P(\|X_0\| > su_n) + P(\|X_0\| > su_n) \leq C((su_n)^q + 1)s^{-\alpha+\eta} v_n$$

for $s > 0$, some constant $C > 0$ large enough, sufficiently large n and some $\eta \in \mathbb{R}$ ($\eta > 0$ for $s > 1$ and $\eta < 0$ for $s < 1$). Thus,

$$\limsup_{n\to\infty} \frac{1}{u_n^q v_n} E[(\|X_0\|^q + 1)\mathbb{1}_{\{\|X_0\|>su_n\}}] \leq \limsup_{n\to\infty} C(s^q + u_n^{-q})s^{-\alpha+\eta} = Cs^{-\alpha+\eta+p} < \infty,$$

which yields part (vi) and, thus, Assumption 2.1 of Kulik et al. (2019) holds. This will be used below.

β-mixing and verification of (P0)

According to Mikosch and Wintenberger (2014), page 161, the drift condition (5.5.5) together with irreducibility implies that $(X_t)_{t\in\mathbb{Z}}$ is geometrically β-mixing, they refer to Meyn and Tweedie (1993), p. 371, see also Kulik et al. (2019). We derive this below with some more details, always assuming the conditions (SRE1) and (SRE2). According to Corollary 14.1.6 in Douc et al. (2018), the drift condition (5.5.5) is equivalent to $E[V(X_1)|X_0 = y] \leq \beta V(y) + \tilde{b}\mathbb{1}_C$ for some petite set C. Note that the required petite sets $\{x : \|x\|^q \leq M\}$, $M < 0$, and the accessible petite set $\{x : \|x\|^q \leq M_0\}$ were established above. The drift condition corresponds to drift condition (V4) of Meyn and Tweedie (1993). Theorem 15.2.6 of Meyn and Tweedie (1993) yields that $\{x : \|x\|^q \leq M_0\}$ is V-geometrically regular and by $1 \leq \tilde{V}$ it is also geometrically regular. Since $\{x : \|x\|^q \leq M_0\}$ is also accessible, Theorem 14.2.6 Douc et al. (2018) implies that the Markov kernel associated to the considered Markov process is geometrically regular. By Corollary 15.1.4 Douc et al. (2018) this implies that the Markov kernel is geometrically ergodic, which, in turn, implies by Theorem 15.1.5 Douc et al. (2018) that $\|P^n(x, \cdot) - P^{X_0}\|_{TV} \leq V(x)\rho^n$ for some $\rho < 1$. Hence, Corollary F.3.4 of Douc et al. (2018) implies the β-mixing with geometrical rate.

The geometric β-mixing means that there exist constants $\rho \in (0, 1)$ and $\tau > 0$ such that $\beta_{n,k} \leq \tau\rho^k$. If one chooses the sequences v_n (or u_n) and s_n so that $v_n = \mathrm{o}(1/\log(n))$, $\log^2(n)/n = \mathrm{o}(v_n)$ and $s_n = \mathrm{o}(\min(v_n^{-1}, (nv_n)^{1/2}))$ (in particular $s_n = \mathrm{o}(\log(n))$), then condition (P0) is satisfied. One could e.g. choose $l_n \geq \max(s_n, \log(n)/|\log(\rho)|)$ and $l_n = \mathrm{o}(r_n)$ with $r_n = \mathrm{o}(\min(v_n^{-1}, (nv_n)^{1/2}))$. Then, the time series satisfies $(n/r_n)\beta^X_{n,l_n} \leq n/r_n\rho^{l_n} \leq n/r_n \exp(\log(\rho)\log(n)/|\log(\rho)|) = n/r_n n^{-1} = 1/r_n \to 0$ and all the other conditions for the rates of the sequences can be easily checked. Thus, $(X_t)_{t\in\mathbb{Z}}$ satisfies condition (P0) and (PR). Note that it suffices to assume conditions of this form for s_n and v_n because the existence of suitable r_n, l_n follows immediately.

Verification of (PP)

Condition (PP) can be shown similarly as for the univariate case in Drees et al. (2015), Example A.3: Recall $\Pi_{j,k} = C_k \cdot C_{k-1} \cdots C_j$, $R_k := \sum_{j=1}^{k} \Pi_{j+1,k} D_j$ and $X_k = R_k + \Pi_{1,k} X_0$. Denote $v_{n,a} = P(\|X_0\| > u_n a)$. This implies for all $a \in (0, 1]$

$$
\begin{aligned}
&P(\|X_k\| > u_n a \mid \|X_0\| > u_n a) \leq \frac{1}{v_{n,a}} P(\|X_0\| > u_n a, \|R_k\| + \|\Pi_{1,k} X_0\| > u_n a) \\
&\leq \frac{1}{v_{n,a}} \left(P\Big(\|X_0\| > u_n a, \|R_k\| > \frac{u_n a}{2} \Big) + P\Big(\|X_0\| > u_n a, \|\Pi_{1,k} X_0\| > \frac{u_n a}{2} \Big) \right) \\
&\leq \frac{1}{v_{n,a}} \left(P\Big(\|X_0\| > u_n a, \|R_k\| > \frac{u_n a}{2} \Big) + P\Big(\|X_0\| > u_n a, \|\Pi_{1,k}\|_M \|X_0\| > \frac{u_n a}{2} \Big) \right) \\
&= P\Big(\|R_k\| > \frac{u_n a}{2} \Big) + \frac{1}{v_{n,a}} \int_{u_n a}^{\infty} P\Big(\|\Pi_{1,k}\|_M > \frac{u_n a}{2t} \Big) P^{\|X_0\|}(dt).
\end{aligned}
$$

Observe that the operator norm is sub-multiplicative and, thus, with κ defined in (5.5.3)

$$
\begin{aligned}
E[\|\Pi_{1,k}\|_M^q] &\leq \Big(E[\|\Pi_{1,m}\|_M^q] \Big)^{\lfloor k/m \rfloor} E\Big[\|\Pi_{1,k-m\lfloor k/m \rfloor}\|_M^q \Big] \\
&\leq \Big(\kappa^{1/m} \Big)^k \max_{0 \leq j < m} \frac{E[\|\Pi_{1,j}\|_M^q]}{\kappa^j} =: \tilde{\kappa}^k c_m,
\end{aligned}
\tag{5.5.8}
$$

for all $k \in \mathbb{N}$, in particular $\tilde{\kappa} < 1$. Applying this and the generalized Markov inequality, we obtain

$$
P\Big(\|\Pi_{1,k}\|_M > \frac{u_n a}{2t} \Big) \leq E\left[\|\Pi_{1,k}\|_M^q \right] \left(\frac{2t}{u_n a} \right)^q \leq \tilde{\kappa}^k c_m \left(\frac{2t}{u_n a} \right)^q .
$$

Thus, using (5.5.7) in the last step we obtain

$$
\begin{aligned}
&\frac{1}{v_{n,a}} \int_{u_n a}^{\infty} P\Big(\|\Pi_{1,k}\|_M > \frac{u_n a}{2t} \Big) P^{\|X_0\|}(dt) \leq \frac{1}{v_{n,a}} \int_{u_n a}^{\infty} \tilde{\kappa}^k c_m \left(\frac{2t}{u_n a} \right)^q P^{\|X_0\|}(dt) \\
&= \tilde{\kappa}^k c_m E\left[\left(\frac{2\|X_0\|}{u_n a} \right)^q \Big| \|X_0\| > u_n a \right] \\
&\leq 2^{q+1} \tilde{\kappa}^k c_m E[\|Y_0\|^q]
\end{aligned}
$$

for all $k \in \mathbb{N}$ and sufficiently large $n \in \mathbb{N}$. Moreover, $\|R_k\| \leq \|X_k\|$ since we consider only non-negative random elements and the L_2-Norm on $\mathbb{R}^d$:

$$
\|R_k\| = \left(\sum_{i=1}^{d} |R_{k,i}|^2 \right)^{1/2} \leq \left(\sum_{i=1}^{d} |R_{k,i} + (\Pi_{1,k} X_0)_i|^2 \right)^{1/2} = \left(\sum_{i=1}^{d} |X_{k,i}|^2 \right)^{1/2} = \|X_k\|.
$$

Hence, $P(\|R_k\| > u_n a/2) \leq P(\|X_k\| > u_n a/2) \leq 2^{1-\alpha} v_{n,a}$ for all k and sufficiently large n, due to the regular variation. All in all, we obtain

$$
P(\|X_k\| > u_n a \mid \|X_0\| > u_n a) \leq c(v_{n,a} + \tilde{\kappa}^k) =: e_{n,a}(k) \tag{5.5.9}
$$

for a suitable constant $c > 0$. The regular variation of X_0 implies

$$r_n v_{n,a} = r_n v_n \frac{P(\|X_0\| > u_n a)}{P(\|X_0\| > u_n)} \to 0 \cdot a^{-\alpha} = 0$$

for all $a \in (0,1]$ and, therefore, $\lim_{n\to\infty} e_{n,a}(k) = c\tilde{\kappa}^k$ and

$$\sum_{k=1}^{r_n} e_{n,a}(k) = cr_n P(\|X_0\| > u_n a) + c\sum_{k=1}^{r_n} \tilde{\kappa}^k$$
$$\to c\sum_{k=1}^{\infty} \tilde{\kappa}^k = c\Big(\frac{1}{1-\tilde{\kappa}} - 1\Big) = \sum_{k=1}^{\infty} \lim_{n\to\infty} e_{n,a}(k) < \infty$$

for all $a \in (0,1]$. Thus, condition (PP) holds.

Verification of (PC)

For $A = [0,x]$ one has $P(\Theta_i \in \partial A) = P(\Theta_i \in \partial A^-) = P(\exists j \in \{1,...,d\} : \Theta_{i,j} = x_j) = 0$ for all $x \in [0,\infty)^d$ and, therefore, conditions (PC) and (PA) (iv) hold if Θ_i has continuous marginal distributions. The representation (5.5.2) combined with the absolute continuity of the marginal distributions of C_1 would imply the absolute continuity of the distribution of Θ_i, for $i > 0$, given that Θ_0 and C_1 are independent. Henceforth, we will assume that all marginal distributions of Θ_i are absolutely continuous. More generally, our results also apply if the marginal distributions of Θ_i are not absolutely continuous, provided that we consider only subsets of the family $\mathcal{A}$. More precisely, in this case we restrict ourself to subsets of the family $\mathcal{A}$ where all sets $[0,y]$, $y = (y_j)_{1\le j\le d} \in [0,\infty)^d$, for which y_j belongs to a neighborhood of some jump point of the j-th marginal distribution for some $1 \le j \le d$ are omitted. Then, by the above reasoning (PC) and (PA) (iv) are satisfied. If we allow only finitely many jump points in each coordinate, then one can check the other conditions in (PA) for all subsets of $\mathcal{A}$ including sets $(-\infty, y]$, where all y lie between the same jump points, separately. In this case (PA) and (PC) are verifiable even if the marginal distributions of Θ_i are not continuous.

Verification of (PT)

For the verification of (PT), we will need some stronger conditions on v_n and s_n:

(SRE3) $v_n = \mathrm{o}((\log n)^{-(3+\zeta)})$ for some $\zeta > 0$, $(\log^2 n)/n = o(v_n)$ and
$s_n = \mathrm{o}\Big(\min(v_n^{-1/(3+\zeta)}, (nv_n)^{1/2})\Big)$.

The first and last condition in (SRE3) ensures $s_n = \mathrm{o}(\log(n))$ and that a suitable r_n with $r_n^{1+\zeta} v_n \to 0$ as needed below for (PM) (i) can be chosen.

For $s_n \ge |j|$ the denominator in the conditional expectation in (PT) is larger or equal to

u_n^α. Moreover, $x^\alpha < x^q$ for $x \in (0,1)$, $q < \alpha$ and $\xi < 1$ for the ξ in condition (PT). Hence,

$$E\left[\frac{\sum_{\tilde{m}<|h|\le s_n} \|X_{h+j}\|^\alpha \mathbb{1}_{\{\|X_{h+j}\|\le \xi u_n\}}}{\sum_{|h|\le s_n} \|X_{h+j}\|^\alpha} \,\middle|\, \|X_0\| > u_n\right]$$
$$\le \sum_{\tilde{m}<|h|\le s_n} E\left[\left(\frac{\|X_{h+j}\|}{u_n}\mathbb{1}_{\{\|X_{h+j}\|<u_n\}}\right)^q \,\middle|\, \|X_0\| > u_n\right].$$

Thus, to verify (PT), it suffices to show

$$\limsup_{\tilde{m}\to\infty} \lim_{n\to\infty} \sum_{\tilde{m}<|h|\le s_n} E\left[\left(\frac{\|X_{h+j}\|}{u_n}\mathbb{1}_{\{\|X_{h+j}\|<u_n\}}\right)^q \,\middle|\, \|X_0\| > u_n\right] = 0$$

for some $q \in (0,\alpha)$ and all $j \in \mathbb{Z}$.
Choose $q \in (\alpha/(1+\zeta/2), \alpha)$ for some $\zeta > 0$ and some $\tau \in (1/q, (1+\zeta/2)/\alpha)$. Define $\varepsilon_h = |h|^{-\tau} < 1$ for all $h \in \mathbb{Z}$. In particular, this implies that $(\varepsilon_h^q)_{h\in\mathbb{Z}}$ is summable, since $q\tau > 1$. Then, by stationarity

$$\sum_{\tilde{m}<|h|\le s_n} E\left[\left(\frac{\|X_{h+j}\|}{u_n}\right)^q \mathbb{1}_{\{\|X_{h+j}\|<u_n\}} \,\middle|\, \|X_0\| > u_n\right]$$
$$\le \sum_{\tilde{m}-j<|h|\le s_n+j} E\left[\left(\frac{\|X_h\|}{u_n}\right)^q \mathbb{1}_{\{\|X_h\|<u_n\}} \,\middle|\, \|X_0\| > u_n\right]$$
$$= \sum_{\tilde{m}-j<|h|\le s_n+j} \left(E\left[\left(\frac{\|X_h\|}{u_n}\right)^q \mathbb{1}_{\{\varepsilon_h u_n<\|X_h\|<u_n\}} \,\middle|\, \|X_0\| > u_n\right]\right.$$
$$\left. + E\left[\left(\frac{\|X_h\|}{u_n}\right)^q \mathbb{1}_{\{\|X_h\|\le\varepsilon_h u_n\}} \,\middle|\, \|X_0\| > u_n\right]\right)$$
$$\le \sum_{\tilde{m}-j<|h|\le s_n+j} \left(P\left(\|X_h\| > \varepsilon_h u_n \mid \|X_0\| > u_n\right) + \varepsilon_h^q\right)$$
$$\le 2\sum_{h=\tilde{m}-j+1}^{s_n+j} \left(P\left(\|X_h\| > \varepsilon_h u_n \mid \|X_0\| > \varepsilon_h u_n\right) \frac{P(\|X_0\| > \varepsilon_h u_n)}{P(\|X_0\| > u_n)} + \varepsilon_h^q\right). \tag{5.5.10}$$

Choose some $q^* \in (\alpha, (1+\zeta/2)/\tau)$. Due to the regular variation of X_0 the sequence u_n is of larger order than $v_n^{-1/q^*} = P(X_0 > u_n)^{-1/q^*}$. Thus, for all $\tilde{m} - j + 1 \le h \le s_n + j$, $\varepsilon_h u_n \ge (s_n + j)^{-\tau} u_n$ is of larger order than $s_n^{-\tau} v_n^{-1/q^*} = (s_n v_n^{1/(\tau q^*)})^{-\tau}$. This term tends to ∞ by condition (SRE3) because $q^*\tau < 1 + \zeta/2 < 3 + \zeta$ by the choice of τ and, therefore, $s_n v_n^{1/(q^*\tau)} < s_n v_n^{1/(1+\zeta/2)} < s_n v_n^{1/(3+\zeta)} \to 0$. Thus, there exists for all $t > 0$ some $n_0 \in \mathbb{N}$ such that for all $n > n_0$ one has $\varepsilon_h u_n > t$ for all $\tilde{m} - j + 1 \le h \le s_n + j$. Then, for some $\varepsilon > 0$ the Potter bounds imply

$$\frac{P(\|X_0\| > \varepsilon_h u_n)}{P(\|X_0\| > u_n)} \le (1+\varepsilon)\varepsilon_h^{-q^*} \tag{5.5.11}$$

for all $\tilde{m} - j + 1 \le h \le s_n + j$. Moreover, $E\left[(2\|X_0\|/x)^q \mid \|X_0\| > x\right] \to 2^q E[\|Y_0\|^q]$ as $x \to \infty$ by (5.5.7). Hence, $E\left[(2\|X_0\|/x)^q \mid \|X_0\| > x\right] \le 2^{q+1} E[\|Y_0\|^q]$ for all $x > x_0$

and some sufficiently large x_0. Therefore, since $\varepsilon_{s_n} u_n \to \infty$, there exists an n_0 so that for all $n \geq n_0$ and $\tilde{m} - j + 1 \leq h \leq s_n + j$ we have $\varepsilon_h u_n \geq \varepsilon_{s_n} u_n \geq x_0$ and $E\left[(2\|X_0\|/(\varepsilon_h u_n))^q \mid \|X_0\| > \varepsilon_h u_n\right] \leq 2^{q+1} E[\|Y_0\|^q]$. Thus, with the same arguments leading to (5.5.9) and with (5.5.11) we obtain

$$\begin{aligned}
&P(\|X_h\| > u_n\varepsilon_h \mid \|X_0\| > u_n\varepsilon_h) \\
&\leq P\Big(\|R_h\| > \frac{u_n\varepsilon_h}{2}\Big) + c_m\tilde{\kappa}^h E\Big[\Big(\frac{2\|X_0\|}{\varepsilon_h u_n}\Big)^q \,\Big|\, \|X_0\| > \varepsilon_h u_n\Big] \\
&\leq 2^{1-\alpha} v_{n,\varepsilon_h} + c_m\tilde{\kappa}^h 2^{q+1} E[\|Y_0\|^q] \leq 2^{1-\alpha} v_n \varepsilon_h^{-q^*} + c_m\tilde{\kappa}^h 2^{q+1} E[\|Y_0\|^q] \\
&= c(v_n\varepsilon_h^{-q^*} + \tilde{\kappa}^h)
\end{aligned}$$

for some $c > 0$ and sufficiently large n uniformly for all $\tilde{m} - j + 1 \leq h \leq s_n + j$ (i.e. there exists an n_0 such that for all $n \geq n_0$ and $\tilde{m} - j + 1 \leq h \leq s_n + j$ this bound holds). Combining this last bound with (5.5.11) yields

$$\begin{aligned}
&\sum_{h=\tilde{m}-j+1}^{s_n+j} \Big(P\left(\|X_h\| > \varepsilon_h u_n \mid \|X_0\| > \varepsilon_h u_n\right) \frac{P(\|X_0\| > \varepsilon_h u_n)}{P(\|X_0\| > u_n)} + \varepsilon_h^q\Big) \\
&\leq \sum_{h=\tilde{m}-j+1}^{s_n+j} \Big(c(\varepsilon_h^{-q^*} v_n + \tilde{\kappa}^h)(1+\varepsilon)\varepsilon_h^{-q^*} + \varepsilon_h^q\Big) \\
&= (1+\varepsilon)c \sum_{h=\tilde{m}-j+1}^{s_n+j} h^{2q^*\tau} v_n + (1+\varepsilon)c \sum_{h=\tilde{m}-j+1}^{s_n+j} \tilde{\kappa}^h h^{q^*\tau} + \sum_{h=\tilde{m}-j+1}^{s_n+j} h^{-q\tau} \\
&= \mathrm{O}(s_n^{2q^*\tau+1} v_n) + \sum_{h=\tilde{m}-j+1}^{s_n+j} ((1+\varepsilon)c\tilde{\kappa}^h h^{q^*\tau} + h^{-q\tau}) = \mathrm{O}(1).
\end{aligned} \tag{5.5.12}$$

In view of (SRE3) the first term converges to 0 since q^*0 is chosen sufficiently such that $q^*\tau \leq 1 + \zeta/2$, since then $s_n^{2q^*\tau+1} v_n = \mathrm{o}(1)$. The sum can be bounded by $((1+\varepsilon)c \vee 1)\sum_{h=\tilde{m}-j+1}^{\infty}(\tilde{\kappa}^{|h|} h^{q^*\tau} + h^{-q\tau}) < \infty$, which is finite since $q\tau > 1$ and $\tilde{\kappa} < 1$. This last bound tends to 0 as $\tilde{m} \to \infty$. Thus, (PT) holds.
So far we have verified all assumptions of Theorem 5.2.6 except for the bias condition, which is always fulfilled if u_n is chosen sufficiently large.

Verification of (PA)

For the process convergence established in Theorem 5.2.9 we have to check Condition (PA). Parts (i) and (ii) are obvious. One can e.g. use the indexing map $[0,1]^d \to \mathcal{A}$, $t \mapsto A_t := [0, \tilde{t}_1] \times ... \times [0, \tilde{t}_d] \cap \mathbb{R}^d$, with

$$\tilde{t}_j := \begin{cases} 1/(1-t_j) - 1 - \varepsilon, & 0 \leq t_j < 1, \\ \infty, & t_j = 1 \end{cases}$$

for some $\varepsilon > 0$ and with the convention $[0, -\delta] = \emptyset$ for $\delta > 0$. One has $A_1 = [0, \infty)^d$ which is enough, since we consider only non-negative random variables. It obviously holds that $A_t = \emptyset$ if $t_j = 0$ for some $j = 1, ..., d$, i.e. (PA) (i) holds. This map is obviously non-decreasing and, therefore, condition (PA) (ii) holds. Moreover, (PA) (iii) holds because the processes are continuous from the right in each coordinate. For all $t \in \mathbb{R}$ and $1 \le k \le d$ one has

$$\bigcap_{s\in(t,1]} A_{s^{(k)}} \setminus A_{t^{(k)}} = [0,\infty)^{k-1} \times \bigcap_{s\in(t,1]} [0,s] \setminus [0,t] \times [0,\infty)^{d-k} = \emptyset,$$

which is why (PA) (v) and (vi) are trivially satisfied. (PA) (vii) is not needed since $0 \in A$ for all $\emptyset \neq A \in \mathcal{A}$. Still, it trivially holds with $0 \in A_w \setminus \bigcup_{s<w} A_s$ and $w_j = 1 - 1/(1+\varepsilon) = \varepsilon/(1+\varepsilon)$, $1 \le j \le d$. Finally, (PA) (iv) was shown above and, therefore, (PA) is verified. Thus, we have verified all conditions for Theorem 5.2.9, apart from the bias condition (PB_T) which always holds for sufficiently large u_n. For $\hat{\hat{p}}_{n,A}$ with unknown α and for Theorem 5.3.1 it remains to check conditions (PP1) and (PM), as the bias condition (PB_α) always holds with sufficiently large u_n.

Verification of (PP1)

The equation (2.9) in condition (C) of Drees and Knezevic (2020) is the univariate version of condition (PP1) and (PM) (i) corresponds to (2.11) in condition (C) of the cited paper, just replace ε by 0 there. Both conditions are verified for one-dimensional solutions to stochastic recurrence equations in Appendix B of Drees and Knezevic (2020). Their proof uses general techniques for Markov processes which will also be used for the d-dimensional solutions of stochastic recurrence equations considered here.

Since Assumption 2.1 of Kulik et al. (2019) is satisfied (see above), the following straight-forward generalization of Lemma 4.3 of the same paper holds: for all functions $\psi : \mathbb{R}^d \to \mathbb{R}$ that vanish on a neighborhood of 0 such that $|\psi(x)| \le c(\|x\|^{q/2} + 1)$ for some $c > 0$ and all $x \in \mathbb{R}^d$, one has

$$\lim_{L\to\infty} \limsup_{n\to\infty} \frac{1}{v_n} \sum_{k=L+1}^{r_n} E[|\psi(X_0/u_n)\psi(X_k/u_n)|] = 0, \tag{5.5.13}$$

provided $r_n v_n \to 0$. Note that Lemma 4.3 Kulik et al. (2019) proves this only for a specific function ψ_ε. However, following exactly the proof of this lemma, including the proof of Lemma 4.1 and 4.2 of that paper, yields the same assertion for the more general function ψ which satisfies $|\psi(x)| \le c(\|x\|^{q/2}+1)$ for some $c > 0$ and $\psi(x) = 0$ for $\|x\| < \varepsilon$ for some $\varepsilon > 0$. To this end, just replace $\|x\|^{q_0}$ by $V(x) = \|x\|^q + 1$ in the proofs there.

Define $\psi(x) := \max(\log^+ \|x\|, \mathbb{1}_{[1,\infty)}(\|x\|))$. It obviously holds $\psi(x) = 0$ for $\|x\| < 1$ and $|\max(\log \|x\|, \mathbb{1}_{[1,\infty)}(\|x\|))| \le c\|x\|^{q/2}\mathbb{1}_{[1,\infty)}(\|x\|) \le c(\|x\|^{q/2} + 1)$ for some $c > 0$ and all $x \in \mathbb{R}^d$, since for all $q > 0$ one has $\log(y) \le c_q y^q$ for some suitable $c_q > 0$. Thus, (5.5.13) holds for this function ψ, i.e.

$$\lim_{L\to\infty}\limsup_{n\to\infty}\sum_{k=L+1}^{r_n} E\Bigg[\max\Big(\log^+\Big(\frac{\|X_0\|}{u_n}\Big), \mathbb{1}_{\{\|X_0\|>u_n\}}\Big) \\ \times \max\Big(\log^+\Big(\frac{\|X_k\|}{u_n}\Big), \mathbb{1}_{\{\|X_k\|>u_n\}}\Big)\,\Big|\, \|X_0\|>u_n\Bigg] = 0. \quad (5.5.14)$$

Define $e'_n(k) := E[\psi(X_0/u_n)\psi(X_k/u_n) \mid \|X_0\| > u_n]$. The uniform integrability of $(\|X_0\|/u_n)^q(\|X_k\|/u_n)^q\mathbb{1}_{\{\|X_0\|>u_n\}}/v_n$, $n\in\mathbb{N}$, for all $q\in(0,\alpha/2)$ and $k\in\mathbb{N}$, which follows by the Cauchy-Schwartz inequality and the same arguments as (5.5.6), implies the uniform integrability of the random variables $\psi(X_0/u_n)\psi(X_k/u_n)/v_n$, $n\in\mathbb{N}$. Together with the weak convergence defining the tail process this implies

$$e'_\infty(k) := \lim_{n\to\infty} e'_n(k) = \lim_{n\to\infty} E[\psi(X_0/u_n)\psi(X_k/u_n) \mid \|X_0\|>u_n] = E[\psi(Y_0)\psi(Y_k)] < \infty$$

for all $k\in\mathbb{N}$. Moreover, the representation of the forward tail process $Y_k = \|Y_0\|\Theta_0\Pi_{1,k}$ and (5.5.8) shows that the $e'_\infty(k)$ are summable:

$$\sum_{k=1}^{\infty} e'_\infty(k) \le c^2E[\|Y_0\|^q]\sum_{k=1}^{\infty}E[\|\Pi_{1,k}\|^q] \le c^2c_mE[\|Y_0\|^q]\sum_{k=1}^{\infty}\tilde{\kappa}^k < \infty,$$

where we used $\psi(x)\le c\|x\|^{q/2}\mathbb{1}_{[1,\infty)}(\|x\|)$ in the first inequality.
Moreover, (5.5.14) implies

$$\begin{aligned}\lim_{n\to\infty}\sum_{k=1}^{r_n} e'_n(k) &= \lim_{n\to\infty}\sum_{k=1}^{L} E[\psi(X_0/u_n)\psi(X_k/u_n) \mid \|X_0\|>u_n] \\ &\quad + \lim_{n\to\infty}\sum_{k=L+1}^{r_n} E[\psi(X_0/u_n)\psi(X_k/u_n) \mid \|X_0\|>u_n] \\ &= \sum_{k=1}^{L} e'_\infty(k) + \mathrm{o}(1) < \infty\end{aligned}$$

as $L\to\infty$. Therefore, $\lim_{n\to\infty}\sum_{k=1}^{r_n} e'_n(k) = \sum_{k=1}^{\infty} e'_\infty(k) < \infty$ and (PP1) holds.

Verification of (PM) (i)

For part (i) of condition (PM) note that due to Proposition 14.1.8 in Douc et al. (2018) there exist a $\beta\in(0,1)$ and $b>1$ such that $E[\|X_k\|^q{+}1 \mid X_0=y] \le \beta^k(\|y\|^q{+}1){+}b/(1{-}\beta)$ for all $y\in\mathbb{R}^d$ and $q\in(0,\alpha)$ and, therefore, $E[\|X_k\|^q \mid X_0=y]\le\beta^k\|y\|^q+\tilde{b}$, with $\tilde{b}=b/(1-\beta)+\beta$. Then, part (i) of (PM) can be shown by following the verification of (2.11) in Appendix B of Drees and Knezevic (2020) while replacing ε with 0 and X_k with $\|X_k\|$. For convenience, we carry this out here: Using the aforementioned inequality yields for all $p, \tilde p > 0$ such that $p + \tilde p < \alpha$

$$
\begin{aligned}
&E\Big[\Big(\frac{\|X_k\|}{u_n}\Big)^p \Big(\frac{\|X_0\|}{u_n}\Big)^{\tilde p} \mathbb{1}_{\{\|X_k\|>u_n\}} \,\Big|\, \|X_0\| > u_n\Big] \\
&\le v_n^{-1} u_n^{-(p+\tilde p)} E[\|X_k\|^p \|X_0\|^{\tilde p} \mathbb{1}_{\{\|X_0\|>u_n\}}] \\
&= v_n^{-1} \int_{u_n}^{\infty} u_n^{-(p+\tilde p)} E[\|X_k\|^p \mid X_0 = y] \|y\|^{\tilde p} \; P^{X_0}(\mathrm{d}y) \\
&\le v_n^{-1} \int_{u_n}^{\infty} u_n^{-(p+\tilde p)} (\beta^k \|y\|^q + \tilde b) \|y\|^{\tilde p} \; P^{X_0}(\mathrm{d}y) \\
&= \beta^k E\Big[\Big(\frac{\|X_0\|}{u_n}\Big)^{p+\tilde p} \,\Big|\, \|X_0\| > u_n\Big] + \tilde b u_n^{-p} E\Big[\Big(\frac{\|X_0\|}{u_n}\Big)^{\tilde p} \,\Big|\, \|X_0\| > u_n\Big] \\
&\le 2\beta^k E[\|Y_0\|^{p+\tilde p}] + 2\tilde b u_n^{-p} E[\|Y_0\|^{\tilde p}]
\end{aligned} \tag{5.5.15}
$$

for sufficiently large n by (5.5.7). Also note, that $\log^+(x) \le c_q x^q$ for some $c_q > 0$. Therefore, (5.5.15) with $p \in (\alpha(1+\delta)/(1+\zeta), \alpha)$, $\delta < \zeta$, and $\tilde p \in (0, \alpha - p)$ shows that the sum on the left-hand side of Condition (PM) (i) can be bounded by a multiple of

$$
\begin{aligned}
&\sum_{k=1}^{r_n} \Big(\beta^k E[\|Y_0\|^{p+\tilde p}] + u_n^{-p} E[\|Y_0\|^{\tilde p}]\Big)^{1/(1+\delta)} \\
&\le \sum_{k=1}^{r_n} (\beta^{1/(1+\delta)})^k (E[\|Y_0\|^{p+\tilde p}])^{1/(1+\delta)} + r_n u_n^{-p/(1+\delta)} (E[\|Y_0\|^{\tilde p}])^{1/(1+\delta)} = \mathrm{O}(1).
\end{aligned}
$$

The last bound holds provided $r_n^{1+\zeta} v_n = \mathrm{O}(1)$, because u_n is of larger order than $v_n^{\eta - 1/\alpha}$ for all $\eta > 0$ due to regular variation. Note that due to (SRE3) one can choose an r_n such that $r_n^{1+\zeta} v_n \to 0$ for some $\zeta > 0$.

Verification of (PM) (ii)

By Lemma 5.3.2 (PM) part (ii) is implied by (PM1) part (i) and (ii), which we will verify here for the solutions to (5.5.1). For (PM1) (ii) note that for all $\varepsilon > 0$ there exists a sufficiently large constant $c_\varepsilon > 0$ so that $|\log(x)|^{1+\eta} x^\alpha \mathbb{1}_{\{x<1\}} \le c_\varepsilon x^{\alpha-\varepsilon}$, since $c_\varepsilon x^{\alpha-\varepsilon} + \log(x)^{1+\eta} x^\alpha = x^{\alpha-\varepsilon}(c_\varepsilon + \log(x)^{1+\eta} x^\varepsilon) \ge 0$ for all $x \in (0,1)$. Thus, to verify (PM1) (ii), it suffices to show

$$
\sum_{|h| \le s_n} E\Big[\Big(\frac{\|X_h\|}{u_n} \mathbb{1}_{\{\|X_h\|<u_n\}}\Big)^q \,\Big|\, \|X_0\| > u_n\Big] = \mathrm{O}(1)
$$

for some $q \in (0, \alpha)$. Note that the summand for $h = 0$ is always 0. This boundedness already follows from (5.5.10) and (5.5.12) with $j = \tilde m = 0$ in the verification of (PT). Next, we consider (PM1) (i). We have $\tilde\psi(x) := \max\Big((\log^+(\|x\|))^{1+\eta}, \mathbb{1}_{[1,\infty)}(\|x\|)\Big) \le c_q \|x\|^{q/2}$ for all $\eta > 0$ and some $c_q > 0$. Thus, by stationarity

$$
E\Big[\sup_{L \le |h| \le s_n} \Big(\log^+\Big(\frac{\|X_h\|}{u_n}\Big)\Big)^{1+\eta} \,\Big|\, \|X_0\| > u_n\Big]
$$

$$\le \frac{1}{v_n} E\Big[\sup_{L<|h|\le s_n} \tilde\psi(X_0/u_n)\tilde\psi(X_h/u_n)\Big] \le \frac{2}{v_n}\sum_{h=L+1}^{s_n} E\Big[\tilde\psi(X_0/u_n)\tilde\psi(X_h/u_n)\Big] \le 1$$

for sufficiently large L. The last inequality holds due to (5.5.13), which in turn holds since $\tilde\psi$ vanishes for $\|x\| < 1$ and $\tilde\psi$ can be bounded by a multiple of $\|x\|^{q/2}$. The uniform integrability of $(\|X_h\|/u_n)^q \mathbb{1}_{\{\|X_0\|>u_n\}}/v_n$, $n \in \mathbb{N}$, for all $q \in (0,\alpha)$ was shown in (5.5.6). Thus, it follows $E\Big[(\log^+(\|X_h\|/u_n))^\eta \mid \|X_0\| > u_n\Big] \to E[(\log^+(\|Y_h\|))^\eta]$. In particular, the expectation $E\Big[(\log^+(\|X_h\|/u_n))^\eta \mid \|X_0\| > u_n\Big]$ is bounded for all fixed $h \in \mathbb{N}$. Hence,

$$\begin{aligned} &E\Bigg[\sup_{0\le|h|\le s_n}\Big(\log^+\Big(\frac{\|X_h\|}{u_n}\Big)\Big)^{1+\eta}\,\Bigg|\,\|X_0\| > u_n\Bigg] \\ &\le 1 + 2\sum_{h=0}^{L} E\Bigg[\Big(\log^+\Big(\frac{\|X_h\|}{u_n}\Big)\Big)^{1+\eta}\,\Bigg|\,\|X_0\| > u_n\Bigg] = \mathrm{O}(1), \end{aligned}$$

for sufficiently large and fixed L, which proves (PM1) (i).

Verification of (PM) (iii)

Finally, (PM) (iii) can be established by the same arguments as (PM1) (ii). Since the denominator is at least 1 due to the conditioning event, and $\log(x)^- x^\alpha \le \tilde c_q x^q$ for $q < \alpha$ and some $\tilde c_q > 0$, it suffices to show

$$\lim_{m\to\infty}\limsup_{n\to\infty}\sum_{m\le|h|\le s_n} E\left[\left(\frac{\|X_h\|}{u_n}\right)^q \mathbb{1}_{\{\|X_h\|<u_n\}}\middle|\,\|X_0\| > u_n\right] = 0.$$

Similar to bound for (5.5.10) in the case $j = 0$, the sum can be bounded by

$$B(n,m) := 2\Big((1+\varepsilon)c\sum_{h=m}^{s_n} h^{2q^*\tau} v_n + (1+\varepsilon)c\sum_{h=m}^{s_n}\tilde\kappa^h h^{q^*\tau} + \sum_{h=m}^{s_n} h^{-q\tau}\Big),$$

where the bound is given in (5.5.12) and we have $\lim_{m\to\infty}\limsup_{n\to\infty} B(n,m) = 0$. Hence, (PM) (iii) holds.

Thus, all conditions were checked. This shows that the developed theory can be applied for d-dimensional solutions to stochastic recurrence equations, under some reasonable conditions. We summarize the result of the previous discussion in the following theorem.

Theorem 5.5.1. *Let (C_t, D_t), $t \in \mathbb{Z}$, be iid $[0,\infty)^{d\times d} \times [0,\infty)^d$-valued random variables. Suppose (SRE1) is satisfied, then (5.5.1) has a unique stationary and regularly varying solution $(X_t)_{t\in\mathbb{Z}}$. If (SRE2) and (SRE3) are satisfied and Θ_i has continuous marginal distributions, then (PR), (P0), (PP), (PT), (PC), (PA), (PP1) and (PM) hold.*
In particular, the statements of Theorems 5.2.9 and 5.3.1 are true for solutions to (5.5.1), provided u_n is chosen sufficiently large such that the bias conditions (PB_T) and (PB_α) are satisfied and $\log^4(n) = o(nv_n)$.

5.6 Simulation study

In the previous sections the new estimator $\hat{\hat{p}}_{n,A}$ for the estimation of $P(\Theta_i \in A)$ was introduced and the asymptotic normality of the estimator was shown. In a short example, we saw that compared with the forward and backward estimators $\hat{p}^f_{n,A}$ and $\hat{\hat{p}}^b_{n,A}$ there is no uniformly better estimator in terms of a smaller asymptotic variance.
In this section, we will analyze the finite sample performance of the new projection based estimator $\hat{\hat{p}}_{n,A}$. For this purpose, we present a Monte Carlo simulation study in which we simulate pseudo random data for different models with heavy tails and consider the bias, the standard deviation and the root mean squared error (RMSE) of the estimator. As competitors for $\hat{\hat{p}}_{n,A}$ we consider the forward estimator $\hat{p}^f_{n,A}$ defined in (5.1.1) and the backward estimators $\hat{\hat{p}}^b_{n,A}$ defined in (5.1.2). Note that only a comparison of $\hat{\hat{p}}_{n,A}$ with $\hat{\hat{p}}^b_{n,A}$ and not $\hat{p}^b_{n,A}$ is fair, since only the former two use estimates for α. We only consider the projection based estimator with estimated α, since α is generally unknown in a real data set and, therefore, $\hat{\hat{p}}_{n,A}$ is the only practically available estimator.
For sake of simplicity, we restrict the simulation study to the case of real valued time series models. We consider the sets $\mathcal{A} = \{A_x = (-\infty, x] \mid x \in \mathbb{R}\}$ and some $i \in \mathbb{N}$, the corresponding probabilities $P(\Theta_i \in A_x)$ describe the cumulative distribution function (cdf) of Θ_i.
Recall that the forward estimator is defined in (5.1.1) by

$$\hat{p}^f_{n,A_x} := \frac{1}{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}} \sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}} \mathbb{1}_{(-\infty,x]}(X_{t+i}/\|X_t\|).$$

The backward estimator from Davis et al. (2018) for the sets A_x is given by

$$\hat{\hat{p}}^b_{n,A_x} := \begin{cases} 1 - \frac{1}{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}} \sum_{t=1}^n \left(\frac{\|X_{t-i}\|}{\|X_i\|}\right)^{\hat{\alpha}_n} \mathbb{1}_{\{\|X_t\|>u_n\}} \mathbb{1}_{(x,\infty)}\left(\frac{X_t}{\|X_{t-i}\|}\right), & \text{if } x \geq 0, \\ \frac{1}{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}} \sum_{t=1}^n \left(\frac{\|X_{t-i}\|}{\|X_i\|}\right)^{\hat{\alpha}_n} \mathbb{1}_{\{\|X_t\|>u_n\}} \mathbb{1}_{(-\infty,x]}\left(\frac{X_t}{\|X_{t-i}\|}\right), & \text{if } x < 0. \end{cases}$$

Here, $\hat{\alpha}_n$ is an estimator for the tail index and is chosen as in (5.1.5) for $\hat{\hat{p}}_{n,A}$. Note that this definition of the backward estimator $\hat{\hat{p}}^b_{n,A_x}$ differs from the definition in (5.1.2). This way, 0 is not included in the sets of the indicators, which should improve the performance of the estimator, since the estimator performs worst for x near 0. Similarly, one could define the projection based estimator differently for $x > 0$ and $x \leq 0$ such that 0 is not in the indicator. However, this has no effect on our simulations, which is why the projection based estimator is calculated as defined in (5.1.6). The forward and backward estimator have already been compared in Davis et al. (2018) and Drees et al. (2015). Here, the focus lies on $\hat{\hat{p}}_{n,A_x}$.

GARCHt and SR model

We will study the performance of the projection based estimator $\hat{\hat{p}}_{n,A_x}$ and compare the results with the performance of $\hat{p}^f_{n,A_x}$ and $\hat{\hat{p}}^b_{n,A_x}$ in four well-known models. We start with two models where the spectral process $(\Theta_t)_{t\in\mathbb{Z}}$, has continuous marginal distribution.

GARCHt The first model is a GARCH(1,1) time series, i.e. $X_t = \sigma_t\varepsilon_t$ with $\sigma_t^2 = a_0 + a_1 X_{t-1}^2 + a_2\sigma_{t-1}^2$ and we consider independent innovations ε_t with Student's t_ν-distribution standardized to unit variance. We choose the parameters $a_0 = 0.1$, $a_1 = 0.14$, $a_2 = 0.84$ and $\nu = 4$. These are possible choices for financial time series. These parameters ensures that the time series is regularly varying with index $\alpha = 2.6$; see Mikosch and Stărică (2000), Section 2.2, and Davis et al. (2018), Section 4.

SR The second model is given by the one-dimensional solution of a stochastic recurrence equation $X_t = C_t X_{t-1} + D_t$, $t \in \mathbb{Z}$, with iid $\mathbb{R}^2$-valued random variables (C_t, D_t). Here, we choose C_t and D_t to be independent with $C_t \sim \mathcal{N}(1/3, 8/9)$ and $D_t \sim \mathcal{N}(-10, 1)$. This ensures that a stationary regularly varying solution exists and $E[C_1^2] = 1$, i.e. $\alpha = 2$ (Kesten, 1973). See also Drees et al. (2015), Section 5.2 and 6, for a detailed description of the model. (Note that this SR model does not exactly fit in the setting of Section 5.5.2, but it enables a comparison with the results of Drees et al. (2015).)

The bias and RMSE of the estimators are calculated with respect to the true asymptotic probabilities $P(\Theta_i \le x)$. For this, we have to specify the distribution of the forward spectral tail process $(\Theta_t)_{t\ge1}$:

GARCHt: For $t \ge 1$ one has

$$\Theta_t \stackrel{d}{=} \frac{\tilde{\varepsilon}_t}{|\tilde{\varepsilon_0}|}\prod_{i=1}^{t}(a_1\tilde{\varepsilon}_{t-i}^2 + a_2)^{1/2},$$

where $\tilde{\varepsilon}_h$, $h \ge 1$, are iid random variables with the same distribution as ε_1 (i.e. Student's t_ν distribution, standardized to unit variance) and $\tilde{\varepsilon}_0$ is a thereof independent random variable with density $f_\varepsilon(x)|x|^\alpha / E[|\varepsilon_1|^\alpha]$, where f_ε is the density of ε_1 and α is the index of the regular variation, see Proposition 6.2 of Ehlert et al. (2015).

SR: For $t \ge 1$ one has

$$\Theta_t \stackrel{d}{=} \Theta_0 \prod_{h=1}^{t} \tilde{C}_h,$$

where $\tilde{C}_h$, $h \ge 1$, are iid random variables with the same distribution as C_1 and $(\tilde{C}_h)_{h\ge1}$ is independent of Θ_0, cf. Janssen and Segers (2014), Example 6.1. Moreover, $P(\Theta_0 = 1) = P(\Theta_0 = -1) = 1/2$, cf. Goldie (1991), Theorem 4.1.

The true probabilities $P(\Theta_i \leq x)$ for $x \in [-2, 2]$ in steps of 0.01 in the GARCHt and SR model are derived numerically via a Monte Carlo simulation of 10^7 random variables with the aforementioned distributions.
For each model, we generate time series $(X_t)_{1 \leq t \leq n}$ of length $n = 2000$ and we perform $M = 1000$ Monte Carlo repetitions. We calculate the estimators for $P(\Theta_i \leq x)$ for $x \in [-2, 2]$ in steps of 0.01 and for lags $i \in \{1, ..., 10\}$. We set u_n to the 0.95 quantile level of the absolute values of one sample and choose the block length $s_n = 30$.
The choice of u_n as $\lfloor 0.95 \cdot n \rfloor$ order statistics is not directly in accordance with the theory from the previous sections, where we assumed u_n as deterministic threshold. However, it is common practice to use such data depending thresholds and for the forward and backward estimator this has no impact on the asymptotic results (cf. Drees and Knezevic (2020)).

Performance of $\hat{\hat{p}}_{n,A_x}$ in the GARCHt model

We start our simulation study with the consideration of the performance of the three estimators in the models introduced above. Figures 5.3 up to 5.5 show the performance of the projection based estimator $\hat{\hat{p}}_{n,A_x}$ and the competitors $\hat{p}^f_{n,A_x}$ and $\hat{\hat{p}}^b_{n,A_x}$ in the two models introduced above. All figures show the mean estimated value and the true values $P(\Theta_i \leq x)$ (left), as means over the $M = 1000$ Monte Carlo repetitions and the standard deviation (middle) of the three estimators as a function of $x \in [-2, 2]$. Furthermore, for better comparability of the RMSE, the RMSE ratios are shown in the right plots, i.e. the RMSE of $\hat{\hat{p}}_{n,A_x}$ divided by the RMSE of $\hat{\hat{p}}^b_{n,A_x}$ or $\hat{p}^f_{n,A_x}$, respectively. If this ratio is smaller than 1, then $\hat{\hat{p}}_{n,A_x}$ is more efficient than the competing estimator, otherwise $\hat{\hat{p}}_{n,A_x}$ is worse in terms of the RMSE.
Figure 5.3 shows the estimators in the GARCHt model for lags $i \in \{1, 5, 10\}$. For estimating α in this model we observed a bias of 0.014 and standard deviation of 0.461. The bias is small for large $|x|$ and all three estimators, but for small values of $|x|$ the bias of $\hat{\hat{p}}_{n,A_x}$ and $\hat{\hat{p}}^b_{n,A_x}$ is larger, due to the typical artificial point mass in $\{0\}$, which is due to the construction. This explains why for $|x| < 0.2$, in comparison to $\hat{p}^f_{n,A_x}$, our estimator underperforms, while for $|x| > 0.2$ the RMSE is significantly lowered by the new estimator. For larger lag i there is a trend, that the bias for small $|x|$ increases which is in accordance with more artificial point mass in $\{0\}$ for larger i.
The standard deviation is also largest for x close to 0. This fits with the results of Davis et al. (2018) for the forward and backward estimator. The standard deviation of $\hat{\hat{p}}_{n,A_x}$ is smallest for almost all cases, only near 0 it is larger than the standard deviation of $\hat{p}^f_{n,A_x}$ for small lag i. The advantage in the standard deviation of $\hat{\hat{p}}_{n,A_x}$ is more pronounced for larger lags i, for lag $i = 10$ the variance of $\hat{\hat{p}}_{n,A_x}$ is even smallest for all values of x. In the RMSE one can see that this trend for the variance for larger lags is stronger than the trend for the bias with larger lag. The RMSE of $\hat{\hat{p}}_{n,A_x}$ is constantly smaller than the

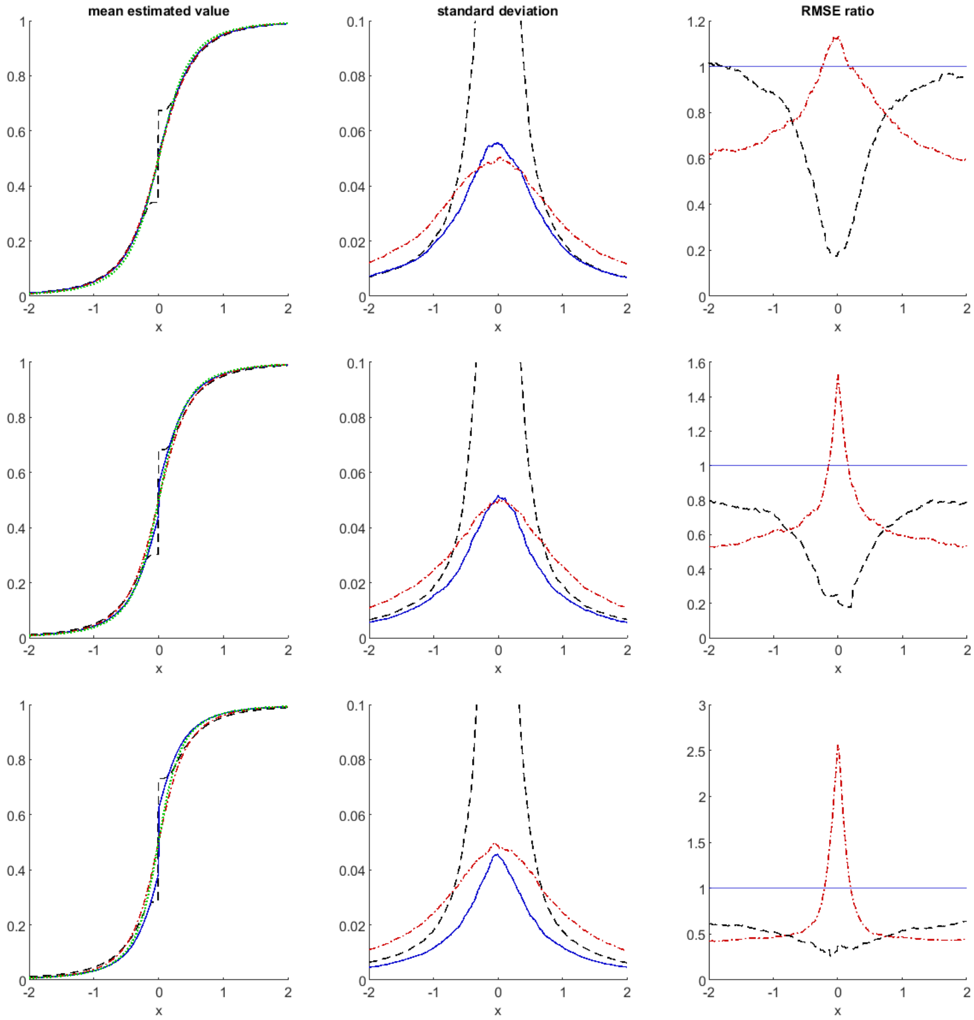

Figure 5.3: *GARCHt model*

Mean (left), standard deviation (middle) and relative RMSE w.r.t. $\hat{\hat{p}}_{n,A_x}$ *of* $\hat{\hat{p}}_{n,A_x}$ *(blue solid line),* $\hat{\hat{p}}^b_{n,A_x}$ *(black dashed line) and* $\hat{p}^f_{n,A_x}$ *(red dashed-dotted line) for lag* $i = 1$ *(top),* $i = 5$ *(middle) and* $i = 10$ *(bottom). The true cdf is indicated by the green dotted line.*

RMSE of $\hat{\hat{p}}^b_{n,A_x}$, for $i = 1$ in the neighborhood of 0 it is half as large. For larger i the general advantage of $\hat{\hat{p}}_{n,A_x}$ is even more pronounced. Except for an environment of 0, the RMSE of $\hat{\hat{p}}_{n,A_x}$ is also smaller than that of $\hat{p}^f_{n,A_x}$. For large i and $|x|$ the relative efficiency of the forward estimator w.r.t. $\hat{\hat{p}}_{n,A_x}$ is less than 1/2. For larger lag $i = 10$ the advantage of $\hat{\hat{p}}_{n,A_x}$ against $\hat{\hat{p}}^b_{n,A_x}$ is even stronger, the RMSE ratio is constantly smaller than 0.7 and the RMSE of $\hat{\hat{p}}_{n,A_x}$ is smaller than the RMSE of $\hat{p}^f_{n,A_x}$ for all $|x| > 0.2$.
These are the results in the GARCHt model with the parameters given above, a different choice of parameters, or, e.g., the use of normally distributed innovations leads to the same qualitatively results.

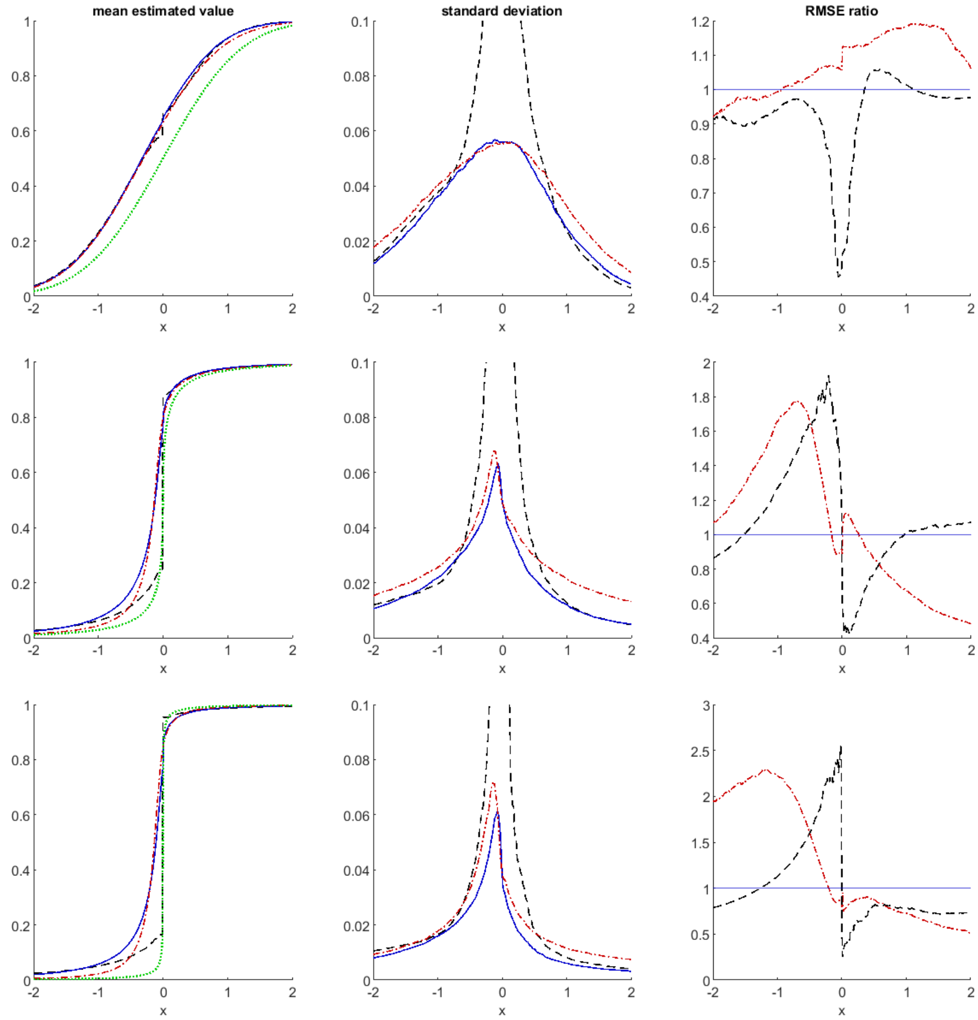

Figure 5.4: *SR model*

Mean (left), standard deviation (middle) and relative RMSE w.r.t. $\hat{\hat{p}}_{n,A_x}$ *of* $\hat{\hat{p}}_{n,A_x}$ *(blue solid line),* $\hat{p}^b_{n,A_x}$ *(black dashed line) and* $\hat{p}^f_{n,A_x}$ *(red dashed-dotted line) for lag* $i = 1$ *(top),* $i = 5$ *(middle) and* $i = 10$ *(bottom). The true cdf is indicated by the green dotted line.*

Performance of $\hat{\hat{p}}_{n,A_x}$ in the SR model

In Figure 5.4 the results of the simulation for the SR model and the lags $i \in \{1, 5, 10\}$ are shown. The plots are more asymmetrical than for the GARCHt model. For estimating α, we observed a bias of 0.2 and a standard deviation of 0.385. In the SR model, the bias is significantly larger for almost all values of x, but the bias of all three estimators is similar. For larger lag i the bias is even larger. For all three estimators the observed strong bias is due to the distribution of D_t, which is irrelevant for the distribution of $(\Theta_t)_{t\in\mathbb{Z}}$ but adds a negative drift in the pre-asymptotic behavior of $(X_t)_{t\in\mathbb{Z}}$. Thus, the bias becomes the dominant part of the RMSE. The standard deviation of $\hat{\hat{p}}_{n,A_x}$ is smaller than that of $\hat{p}^b_{n,A_x}$ and comparable to that of $\hat{p}^f_{n,A_x}$ for small $|x|$, while it becomes smaller compared to $\hat{p}^f_{n,A_x}$ and comparable for $\hat{p}^b_{n,A_x}$, for large $|x|$. Again, using a larger lag i further reduces the standard deviation of $\hat{\hat{p}}_{n,A_x}$ in comparison to the other estimators and the standard

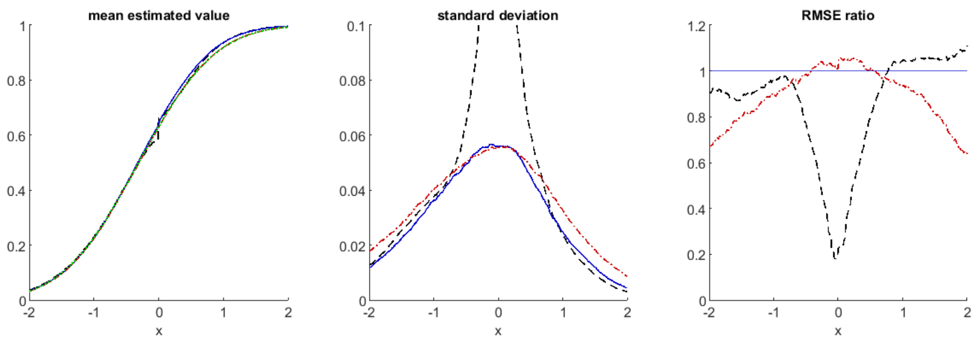

Figure 5.5: *SR model w.r.t. pre-asymptotic probabilities*

Mean (left), standard deviation (middle) and relative RMSE w.r.t. $\hat{\hat{p}}_{n,A_x}$ *of* $\hat{\hat{p}}_{n,A_x}$ *(blue solid line),* $\hat{p}^b_{n,A_x}$ *(black dashed line) and* $\hat{p}^f_{n,A_x}$ *(red dashed-dotted line) as estimator of the pre-asymptotic cdf of* $X_i/|X_0|$ *given* $|X_0| > F^{\leftarrow}_{|X|}(0.95)$ *in the SR model for lag* $i = 1$*. The true cdf is indicated by the green dotted line.*

deviation of $\hat{\hat{p}}_{n,A_x}$ is the smallest throughout for $i = 10$. The asymmetric bias is dominant in the RMSE, the variance advantage of $\hat{\hat{p}}_{n,A_x}$ is not always reflected, in particular for $i \in \{5, 10\}$ and $x < 0$ the projection based estimator $\hat{\hat{p}}_{n,A_x}$ has the largest RMSE due to the larger bias compared to $\hat{p}^f_{n,A_x}$.

In the SR model the bias is large for all three estimators because the distribution of D_t affects the pre-asymptotic probabilities $P(X_i/\|X_0\| < x \mid \|X_0\| > F^{\leftarrow}_{\|X\|}(l))$ for fixed l but not the limiting quantity $P(\Theta_i \le x)$. Here, $F^{\leftarrow}_{\|X\|}(l)$ denotes the quantile function of $\|X_0\|$ at level l. Note that the backward estimator and the projection based estimator are constructed based on principles which only hold in the limit. However, the estimators themselves are motivated as empirical counterparts to the pre-asymptotic probabilities and the mean only converges asymptotically to $P(\Theta_i \le x)$. Therefore, we want to compare the performance of all three estimators w.r.t. the pre-asymptotic probabilities, analogous to the analysis in Davis et al. (2018) for $\hat{p}^b_{n,A_x}$. The quantiles $F^{\leftarrow}_{\|X\|}(l)$ and the pre-asymptotic probabilities needed for this analysis are calculated numerically via 10^7 Monte Carlo repetitions for time series of length 10^4. Figure 5.5 shows the results when the bias and the RMSE are calculated w.r.t. the pre-asymptotic probabilities $P(X_i/|X_0| < x \mid |X_0| > F^{\leftarrow}_{|X|}(0.95))$, instead of the true values $P(\Theta_i \le x)$, for lag $i = 1$. Of course, this only affects the true cdf and RMSE, the mean estimated values and standard deviations remain unchanged. Now the bias of all estimators is corrected and becomes much smaller, compared to Figure 5.4. The change to the pre-asymptotic probabilities has little effect on the RMSE ratio of $\hat{\hat{p}}_{n,A_x}$ w.r.t. $\hat{p}^b_{n,A_x}$ but it improves the RMSE ratio w.r.t. $\hat{p}^f_{n,A_x}$, which is then only (slightly) larger than 1 for $|x| < 0.2$. We can observe that $\hat{\hat{p}}_{n,A_x}$ performs very well here when estimating pre-asymptotic probabilities, even though it was not constructed for this task in particular. However, this consideration of pre-asymptotic probabilities somewhat distorts the picture, so in the following we will again restrict ourselves to the bias regarding the asymptotic probabilities.

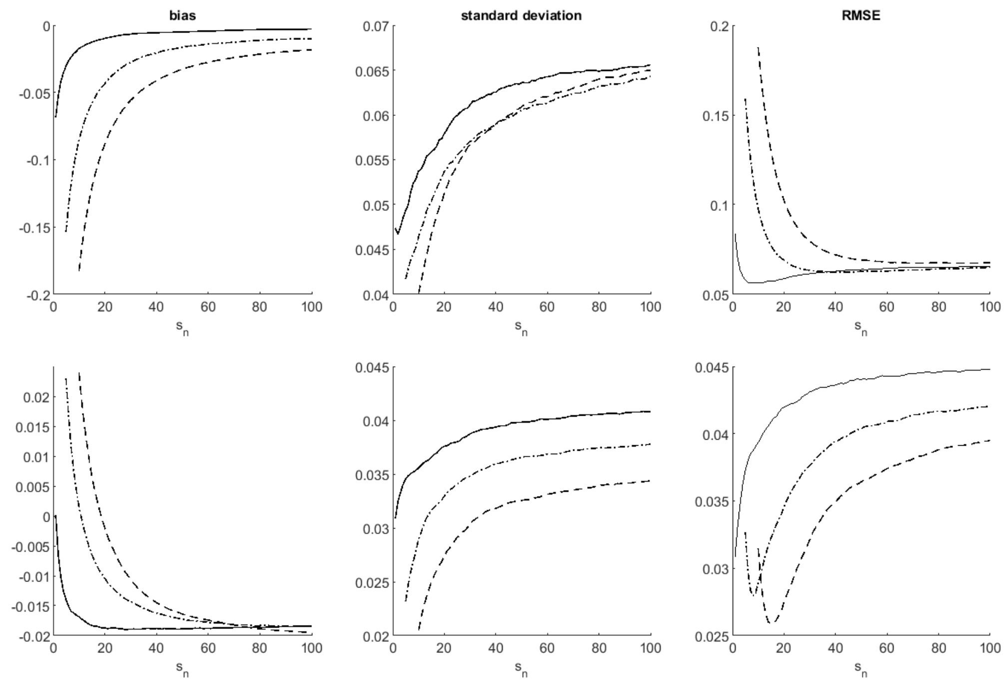

Figure 5.6: *$\hat{\hat{p}}_{n,A_x}$ as function in s_n, GARCHt model*

Bias (left), standard deviation (middle) and RMSE (right) of $\hat{\hat{p}}_{n,A_x}$ as function in s_n for lag $i = 1$ (solid line), $i = 5$ (dashed-dotted line) and $i = 10$ (dashed line) with $x = 0$ (top) and $x = 1/2$ (bottom) in the GARCHt model.

Tuning parameter: block size s_n

Next, we consider the sensitivity of $\hat{\hat{p}}_{n,A_x}$ w.r.t. the two tuning parameters s_n and u_n. The calculation of $\hat{\hat{p}}_{n,A_x}$ requires the choice of the tuning parameter $s_n > i$. So far we used $s_n = 30$ and we will justify that in a moment. A priori it is unclear how large s_n should be chosen. The theory suggests that s_n is small enough compared with n, but with growing n the sequence s_n should also tend to infinity, i.e. s_n should be sufficiently large. If s_n is chosen too small, the observed clusters entering the estimator $\hat{\hat{p}}_{n,A_x}$ are cut off prematurely and the projection based estimator places a larger artificial point mass in $\{0\}$, an artifact due to the construction and the cut-off observations in each summand. This would typically add a bias for $A_x = (-\infty, x]$ with x close to 0. On the other hand, if we choose s_n too large, then almost independent clusters of large observations are compounded which can add bias and increase the variance.
Figure 5.6 shows the bias, standard deviation and RMSE of $\hat{\hat{p}}_{n,A_x}$ in the GARCHt model as function of $s_n \in [i, 100]$ for the lags $i \in \{1, 5, 10\}$ and $x \in \{0, 1/2\}$. Especially for larger lags one can observe a larger bias for $x = 0$ and small values of s_n, which results in a much larger RMSE. Thus, s_n should not be chosen to small. On the other hand, choosing $s_n \geq 20$ has no big influence on the bias and variance, and, therefore, the RMSE is relatively stable against changes of s_n. For $x = 1/2$ and very small s_n ($s_n = 2$ for $i = 1$ and $s_n = 18$ for $i = 10$) the RMSE has a minimum and is afterwards almost constant but

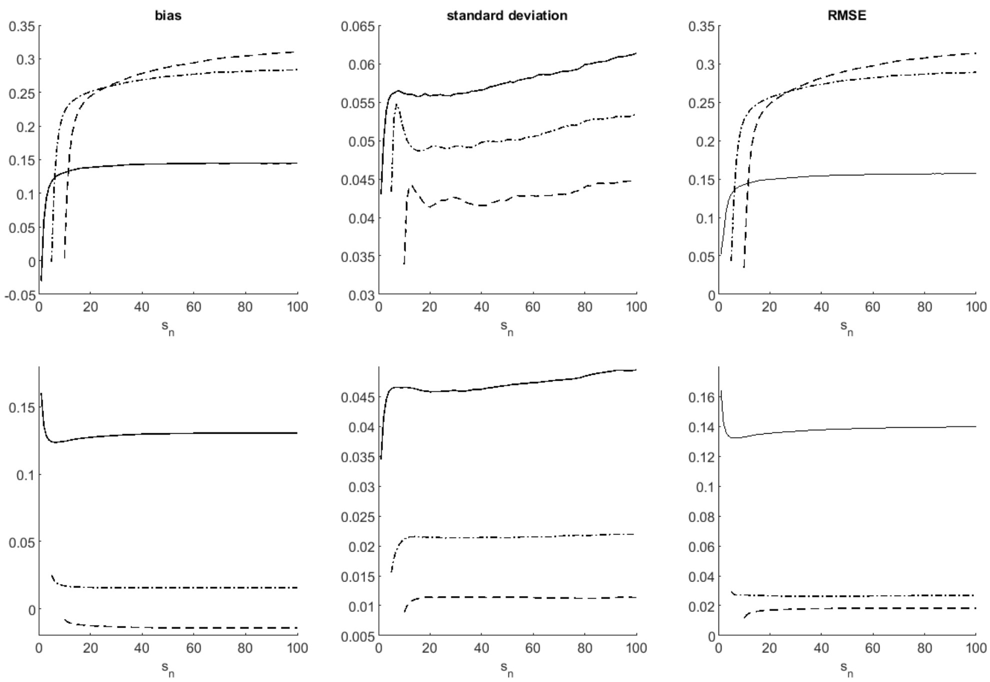

Figure 5.7: *$\hat{\hat{p}}_{n,A_x}$ as function in s_n, SR model*

Bias (left), standard deviation (middle) and RMSE (right) of $\hat{\hat{p}}_{n,A_x}$ as function in s_n for lag $i = 1$ (solid line), $i = 5$ (dashed-dotted line) and $i = 10$ (dashed line) with $x = 0$ (top) and $x = 1/2$ (bottom) in the SR model.

increasing. Thus, a change of s_n does not have high impact on the quality of the estimate, but s_n should not be chosen too large for larger $|x|$.

Basically the same behavior can be observed in the SR model, as shown in Figure 5.7. Note that the small bias for $x = 0$ and small s_n results from a negative bias in the pre-asymptotic model and the artificial point mass in $\{0\}$ of $\hat{\hat{p}}_{n,A_x}$ for really small s_n. In general the RMSE is again relatively stable w.r.t. the choice of $s_n \geq 20$.

Overall, we conclude from both Figures 5.6 and 5.7 that the performance of the estimator is relatively stable once s_n is chosen not too small, i.e. $s_n \geq 25$. Unless an excessively large value is used for s_n, the performance of the estimator is not very sensitive to this tuning parameter.

In practice, one cannot try which s_n fits best, since only one observed time series is available. Therefore, for practical applications, in order to find a suitable value for s_n for an unknown model, we suggest to produce a graph similar to the Hill plot, which shows the estimated value as function of different s_n. Figure 5.8 shows such a plot for the GARCHt model for several choices of x and lags i. A suitable choice of s_n should be chosen not too large but in a region where the estimator becomes stable, which is between 30 and 40 for most values i in Figure 5.8.

We plotted s_n only up to 100. Beyond that the graphs continue constantly until about $n/3$. After that the variance and, thus, the RMSE increases dramatically. Since s_n has no great influence on the performance of the projection estimator $\hat{\hat{p}}_{n,A_x}$, we used $s_n = 30$

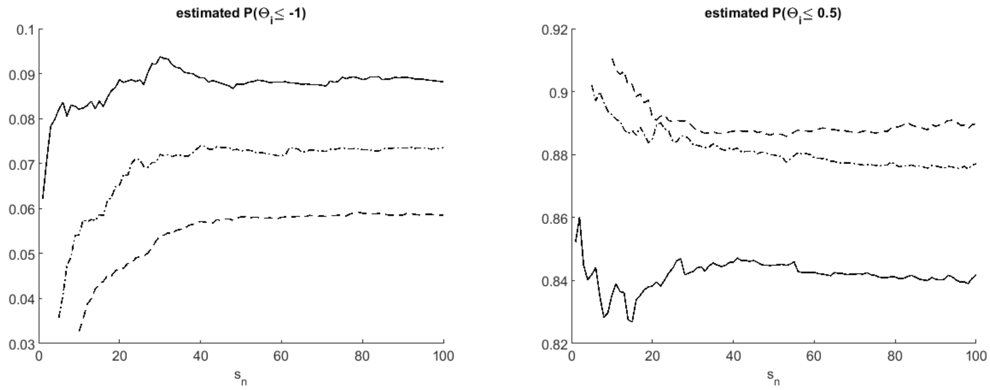

Figure 5.8: *Hill-Plot for $\hat{\hat{p}}_{n,A_x}$, GARCHt model*

$\hat{\hat{p}}_{n,A_x}$ as function of s_n for a single realization of the GARCHt model for lag $i = 1$ (solid line), $i = 5$ (dashed-dotted line) and $i = 10$ (dashed line) with $x = -1$ (left) and $x = 1/2$ (right).

for all simulations, which yields a good overall performance across different models, lags and values for x. However one should note, that in in some cases it could be advisable to choose different s_n for different x and i. The choice $s_n = 30$ is a compromise over all different lags and values for x and works quite well here.

Tuning parameter: threshold u_n

As always in the peak-over-threshold setting we have to choose the second tuning parameter, the threshold u_n. So far we have fixed it as $\lfloor 0.95 \cdot n \rfloor$-th empirical order statistic of the absolute values of the observations $(X_t)_{1 \le t \le n}$. We say that this threshold is of level 0.95. In the next plots we want to address the problem of choosing this tuning parameter and threshold u_n, which is used for the estimators $\hat{\hat{p}}_{n,A_x}$, $\hat{p}^f_{n,A_x}$ and $\hat{\hat{p}}^b_{n,A_x}$ as well as for the estimation of α. An obvious question is whether the level of the threshold affects the qualitative performance of the estimators, or how sensitive the estimators is w.r.t. u_n. Another question is whether the level should be chosen differently for $\hat{\hat{p}}_{n,A_x}$ and $\hat{p}^f_{n,A_x}$, since the latter uses far fewer observations and could therefore benefit from using a smaller threshold.

Varying the level between 0.9 and 0.99 does typically not affect the qualitative results, in particular the RMSE ranking of $\hat{\hat{p}}_{n,A_x}$ w.r.t. $\hat{\hat{p}}^b_{n,A_x}$ and $\hat{p}^f_{n,A_x}$, respectively. This is illustrated in Figure 5.9 for two cases. The figure shows the estimators as a function of different levels of u_n for the GARCHt model with $i = 10$, $x = 1$ and for $i = 1$, $x = 0.1$. We observe that the performance of $\hat{p}^f_{n,A_x}$ and $\hat{\hat{p}}_{n,A_x}$ is relatively stable for a broad range of values for u_n, but that the variance sharply increases for values of u_n larger than the 0.99 quantile. The backward estimator is more affected by the choice of the threshold and $\hat{\hat{p}}_{n,A_x}$ is superior to $\hat{\hat{p}}^b_{n,A_x}$ in terms of RMSE for all u_n and both parameter constellations considered here. In addition, one can see for $i = 10$, $x = 1$ that the RMSE of $\hat{\hat{p}}_{n,A_x}$ is smaller than the other two RMSE regardless of the level of u_n and also the global

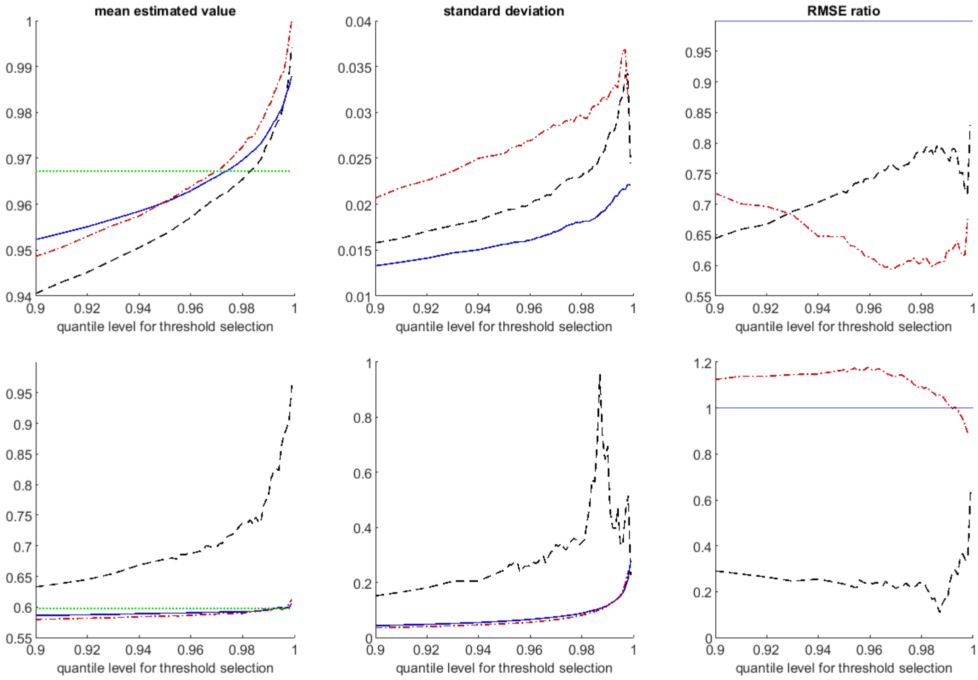

Figure 5.9: *Estimators as function of quantile level for u_n-selection, GARCHt model*

Mean (left), standard deviation (middle) and relative RMSE w.r.t. $\hat{\hat{p}}_{n,A_x}$,as function of quantile levels $F_{|X|}(u_n)$, of $\hat{\hat{p}}_{n,A_x}$ (blue solid line), $\hat{\hat{p}}^b_{n,A_x}$ (black dashed line) and $\hat{p}^f_{n,A_x}$ (red dashed-dotted line) in the GARCHt model for lag $i = 10$ and $x = 1$ (top) and $i = 1$, $x = 0.1$ (bottom). The true value is indicated by the green dotted line.

minimum of the RMSE of $\hat{\hat{p}}_{n,A_x}$ is smaller than the global minimum of the other two RMSEs. Similar results can be observed for $i = 1$ and $x = 0.1$, where $\hat{\hat{p}}_{n,A_x}$ has a larger RMSE than $\hat{p}^f_{n,A_x}$ and smaller RMSE than $\hat{\hat{p}}^b_{n,A_x}$, regardless of the level of u_n. The only exception is observed for level larger 0.99, see Figure 5.9. This shows, that the choice of the level (or different levels for different estimators) essentially has no influence on the ranking of $\hat{\hat{p}}_{n,A_x}$ w.r.t. $\hat{\hat{p}}^b_{n,A_x}$ or $\hat{p}^f_{n,A_x}$, respectively. This is not so clear, if one only compares $\hat{\hat{p}}^b_{n,A_x}$ and $\hat{p}^f_{n,A_x}$, as can be seen in the top row of Figure 5.9. Basically the same results can be observed in the SR model, see Figure 5.10. Unless the threshold is chosen very high, the performance of all three estimators is quite stable, the least stable estimator is the backward estimator.

All in all, the new proposed projection based estimator $\hat{\hat{p}}_{n,A_x}$ is fairly insensitive w.r.t. changes of the threshold u_n or the block length s_n. This is a useful property of the estimator, since it means, that it can be practically used without too much additional information.

Models with discontinuities: SV and AR model

So far, we considered models where $(\Theta_t)_{t\in\mathbb{Z}}$ has continuous marginal distributions. To complete the analysis, we now consider the performance of $\hat{\hat{p}}_{n,A_x}$ in models where P^{Θ_i} has point mass for some $i > 0$. To this end, we introduce two additional models.

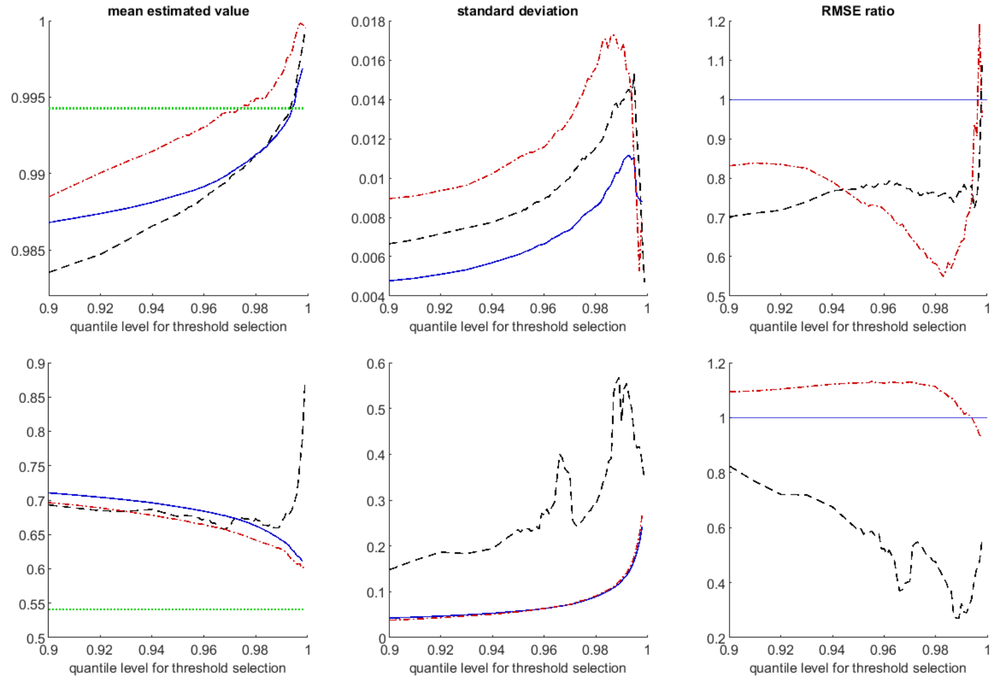

Figure 5.10: *Estimators as function of quantile level for u_n-selection, SR model*

Mean (left), standard deviation (middle) and relative RMSE w.r.t. $\hat{\hat{p}}_{n,A_x}$ as function of quantile levels $F_{|X|}(u_n)$ of $\hat{\hat{p}}_{n,A_x}$ (blue solid line), $\hat{\hat{p}}^b_{n,A_x}$ (black dashed line) and $\hat{p}^f_{n,A_x}$ (red dashed-dotted line) in the GARCHt model for lag $i = 10$ and $x = 1$ (top) and $i = 1$, $x = 0.1$ (bottom). The true value is indicated by the green dotted line.

SV We look at the stochastic volatility model defined by $X_t = \sigma_t \varepsilon_t$ with $\log(\sigma_t) = a_1 \log(\sigma_{t-1}) + Z_t$ where Z_t are iid standard normally distributed random variables and ε_t are iid random variables with Student's t_ν-distribution. In this model we choose the parameters $a_1 = 0.9$ and $\nu = 2.6$. Then $(X_t)_{t\in\mathbb{Z}}$ is a stationary regularly varying time series with index $\alpha = 2.6$, see Davis et al. (2018), Section 4.

AR The last model is an $AR(1)$ time series $X_t = aX_{t-1} + \varepsilon_t$ with $a \in (0,1)$ and the innovations ε_t are independent and symmetric around 0. Here we choose $(|\varepsilon_t| + 1)$ as $Par(\alpha)$-distributed and we choose the parameters $a = 0.95$ and $\alpha = 2$. Then, the index of regular variation is $\alpha = 2$.

In these models the distribution of the spectral tail process can be specified as follows.

SV: Since the volatility σ_t has light tails, the extremal behavior of $(X_t)_{t\in\mathbb{Z}}$ is dominated by the iid heavy-tailed innovations ε_t. Thus, in this model we have $P(\Theta_t = 0) = 1$ for all $t > 0$, see Davis and Mikosch (2009a).

AR: This is a special case of a general stochastic recurrence equation as in the SR model. Inserting $C_t = a$ in the formula for the spectral tail process in the SR model yields $\Theta_t = a^t \Theta_0$, $t \geq 0$, with $P(\Theta_0 = 1) = P(\Theta_0 = -1) = 1/2$. In particular, Θ_t can only take two different values for all $t \geq 0$.

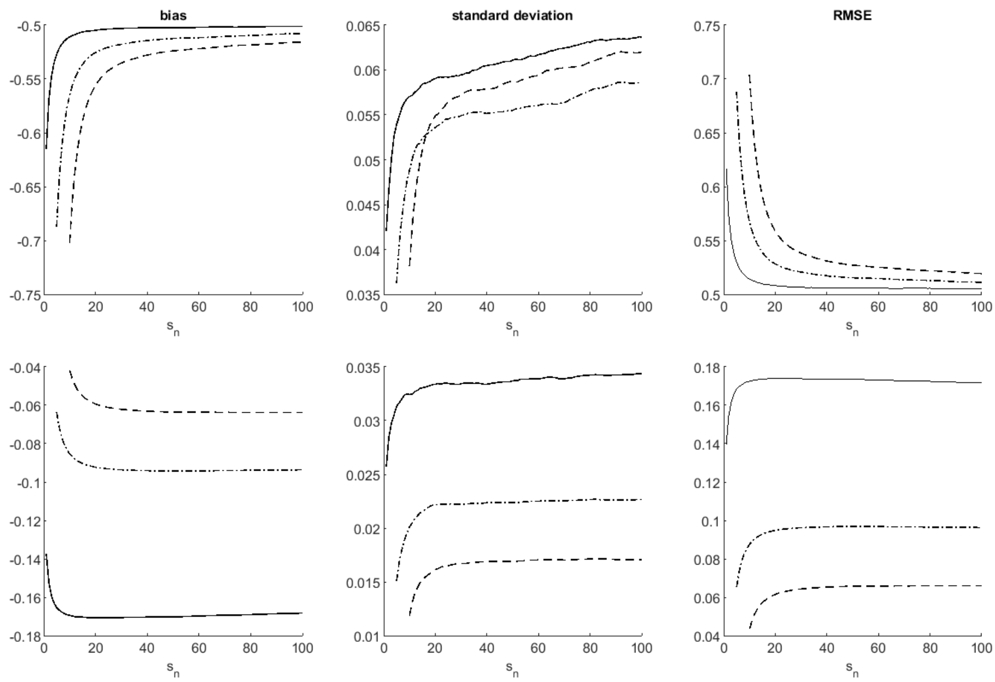

Figure 5.11: *$\hat{\hat{p}}_{n,A_x}$ as function in s_n, SV model*

Bias (left), standard deviation (middle) and RMSE (right) of $\hat{\hat{p}}_{n,A_x}$ as function in s_n for lag $i = 1$ (solid line), $i = 5$ (dashed-dotted line) and $i = 10$ (dashed line) with $x = 0$ (top) and $x = 1/2$ (bottom) in the SR model.

For the SV and AR model the true asymptotic probabilities $P(\Theta_i \leq x)$ can be calculated directly from the above representation of Θ_i, in particular the spectral tail process has discrete mass in some points. Note that therefore condition (PC) is not satisfiable for these discontinuity points. Thus, the conditions could be satisfied only for the family of sets $(-\infty, x]$ if a neighborhood of 0 is omitted from the range of x-values. Nevertheless, in this simulation study we present the results for the full range $x \in [-2, 2]$.
First we consider once more the choice of s_n. Figure 5.11 shows the estimator and RMSE as function in s_n, now for the SV model, similar to Figure 5.6 and Figure 5.7. Again, the estimator and RMSE are almost constant for $s_n \geq 25$, i.e. s_n does not have much effect on the performance, as long as it is chosen large enough. Hence, we choose $s_n = 30$ as before as a uniform parameter for all values of x and lags i.

Performance of $\hat{\hat{p}}_{n,A_x}$ in the SV model

Simulation results for the SV model are presented in Figure 5.12 for the lags $i \in \{1, 5, 10\}$. In the SV model, the only point of mass is 0, i.e. we have $P(\Theta_i \leq x) = \mathbb{1}_{[0,\infty)}(x)$. The jump point of the distribution function leads to a large bias for all three considered estimators for x close to 0. For the standard deviation, the picture is the same as for the GARCHt model. For large $|x|$ the standard deviation of $\hat{\hat{p}}_{n,A_x}$ is smaller than that of $\hat{p}^f_{n,A_x}$, for small $|x|$ it is considerably smaller than that of $\hat{\hat{p}}^b_{n,A_x}$. For lag $i = 1$ there is a

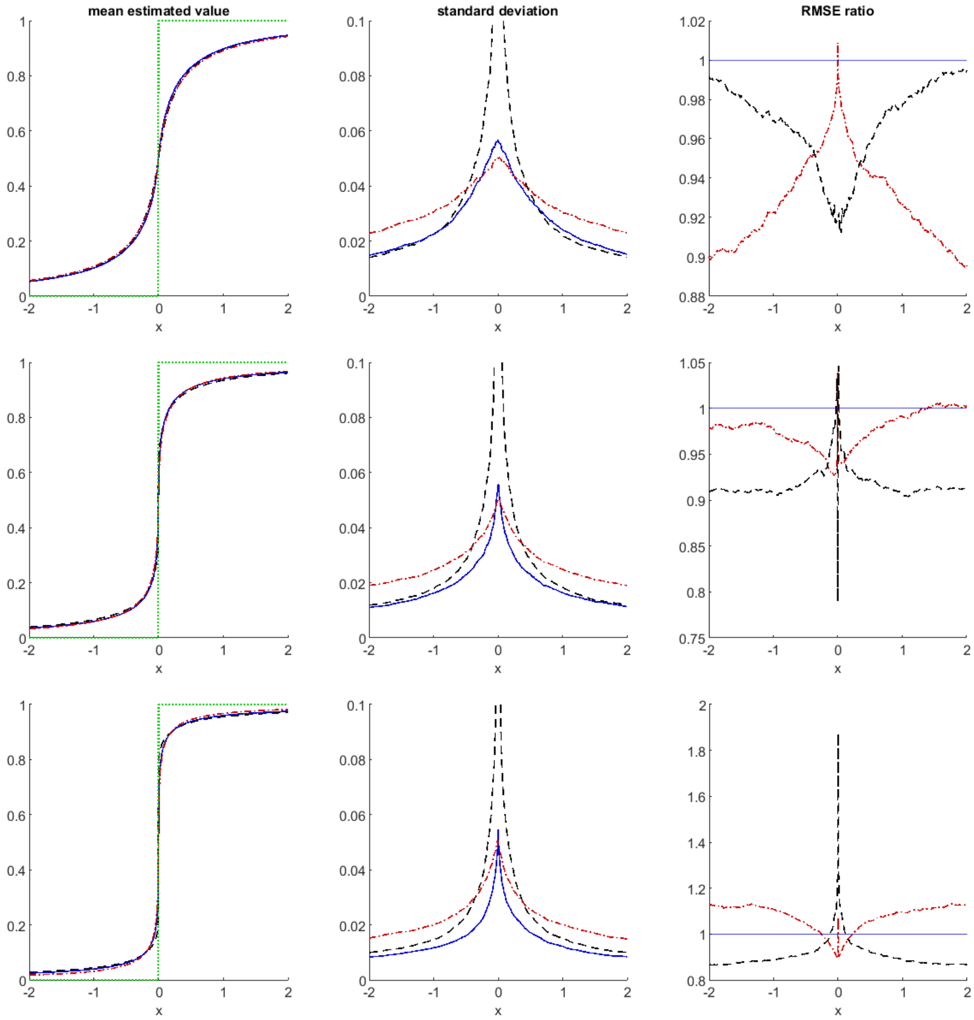

Figure 5.12: *SV model*

Mean (left), standard deviation (middle) and relative RMSE w.r.t. $\hat{\hat{p}}_{n,A_x}$ *of* $\hat{\hat{p}}_{n,A_x}$ *(blue solid line),* $\hat{\hat{p}}^b_{n,A_x}$ *(black dashed line) and* $\hat{p}^f_{n,A_x}$ *(red dashed-dotted line) for lag* $i = 1$ *(top),* $i = 5$ *(middle) and* $i = 10$ *(bottom). The true cdf is indicated by the green dotted line.*

weak advantage of $\hat{\hat{p}}_{n,A_x}$ in terms of the RMSE. For larger lag i, the results change as in the previous models, the standard deviation becomes even smaller relative to the other estimators, but the bias gets worse, which is why $\hat{p}^f_{n,A_x}$ becomes better than $\hat{\hat{p}}_{n,A_x}$. The backward estimator still has the largest RMSE. However, the efficiency advantage of the estimators against each other is at most 10% outside a neighborhood from $x = 0$, which is less than in the other models.

Performance of $\hat{\hat{p}}_{n,A_x}$ in the AR model

Finally, we consider the AR model in Figure 5.13. The distribution of Θ_i has two points of mass in $-a^i$ and a^i, and the bias is much larger in the surrounding of and between these points. The shape of the standard deviation is similar for all three estimators but it is slightly smaller for $\hat{\hat{p}}_{n,A_x}$ and larger for $\hat{\hat{p}}^b_{n,A_x}$ for $|x| < a^i$. The RMSE of $\hat{\hat{p}}_{n,A_x}$

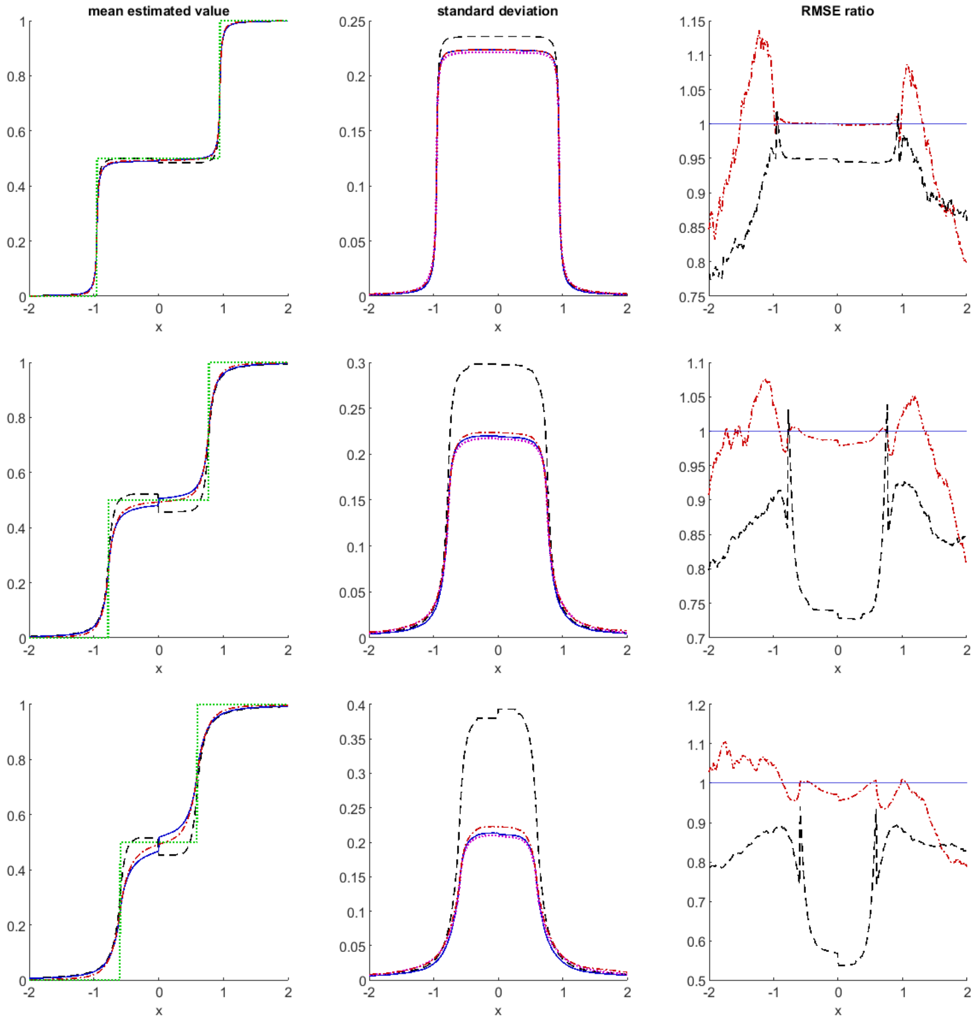

Figure 5.13: *AR model*

Mean (left), standard deviation (middle) and relative RMSE w.r.t. $\hat{\hat{p}}_{n,A_x}$ *of* $\hat{\hat{p}}_{n,A_x}$ *(blue solid line),* $\hat{p}^b_{n,A_x}$ *(black dashed line) and* $\hat{p}^f_{n,A_x}$ *(red dashed-dotted line) for lag* $i = 1$ *(top),* $i = 5$ *(middle) and* $i = 10$ *(bottom). The true cdf is indicated by the green dotted line. The standard deviation of* $\hat{p}_{n,A_x}$ *(magenta dotted line) is added.*

is smaller than the RMSE of $\hat{\hat{p}}^b_{n,A_x}$ almost everywhere. The estimators $\hat{\hat{p}}_{n,A_x}$ and $\hat{p}^f_{n,A_x}$ are comparably efficient with deviations of at most 15%, depending on x both can be advantageous.

Figure 5.13 visualizes the standard deviation of the estimator $\hat{p}_{n,A_x}$ with known α as considered in Section 5.2.2 as a magenta dotted line. According to the example on page 152 the asymptotic variance of this estimator is 0. This is not yet visible in this simulation. Rather, the standard deviation and the RMSE are almost equal to that of $\hat{\hat{p}}_{n,A_x}$ and only minimally smaller.

Conclusion

Overall, $\hat{\hat{p}}_{n,A_x}$ tends to have the smallest variance of the three estimators, especially for higher lags i and larger values of $|x|$. If pre-asymptotic probabilities differ significantly from the limit values, or alternatively for small values of $|x|$, a bias can lead to $\hat{\hat{p}}_{n,A_x}$ having larger RMSE's than alternative estimators. Still, $\hat{\hat{p}}_{n,A_x}$ provides a robust way of estimating limiting quantities and is superior or competitive to $\hat{p}^f_{n,A_x}$ and $\hat{\hat{p}}^b_{n,A_x}$ in terms of RMSE for different models and a large range of sets A. While $\hat{\hat{p}}^b_{n,A_x}$ was specifically designed to perform well for larger $|x|$, the exact threshold for which it outperforms $\hat{p}^f_{n,A_x}$ is a priori unknown and the projection based estimator $\hat{\hat{p}}_{n,A_x}$ provides a robust alternative that can be used for all x. The tuning parameters s_n and u_n have no huge impact on the performance of $\hat{\hat{p}}_{n,A_x}$.
Apart from the numerical advantage of $\hat{\hat{p}}_{n,A_x}$ demonstrated here, there is the theoretical advantage that the new estimator actually gives a distribution of a spectral tail process. All in all, $\hat{\hat{p}}_{n,A}$ performs reasonably well on finite samples and, therefore, it is a good alternative to the existing estimators for $P(\Theta_i \in A)$ for $d = 1$, which is motivated and defined for $d > 1$ and can even be extended to multiple time points easily.
This concludes this chapter on the new projection based estimator for the spectral tail process. In the final chapter a short outlook on open research questions concludes this thesis.

5.7 Proofs

Finally, we compile the proofs of the lemmas, propositions and theorems from this chapter.

5.7.1 Proofs for Section 5.1

In this section, we only have to prove Lemma 5.1.1.

Proof of Lemma 5.1.1. Direct calculations for an arbitrary set $A \in \mathcal{B}(l_\alpha)$ result in

$$\begin{aligned}
(Q^{RS})^{RS}(A) &= \sum_{k\in\mathbb{Z}} \int \frac{\|z_k\|^\alpha}{\|z\|_\alpha^\alpha} \mathbb{1}_A\left(\left(\frac{z_{s+k}}{\|z_k\|}\right)_{s\in\mathbb{Z}}\right) Q^{RS}(dz)\\
&= \sum_{k\in\mathbb{Z}}\sum_{l\in\mathbb{Z}} \int \frac{\|z_l\|^\alpha}{\|z\|_\alpha^\alpha} \frac{\|z_{l+k}\|^\alpha/\|z_l\|_\alpha^\alpha}{\|z/z_l\|_\alpha^\alpha} \mathbb{1}_A\left(\left(\frac{z_{s+k+l}/\|z_l\|}{\|z_{k+l}\|/\|z_l\|}\right)_{s\in\mathbb{Z}}\right) Q(dz)\\
&= \sum_{k\in\mathbb{Z}}\sum_{l\in\mathbb{Z}} \int \frac{\|z_l\|^\alpha}{\|z\|_\alpha^\alpha} \frac{\|z_{l+k}\|^\alpha}{\|z\|_\alpha^\alpha} \mathbb{1}_A\left(\left(\frac{z_{s+k+l}}{\|z_{k+l}\|}\right)_{s\in\mathbb{Z}}\right) Q(dz)\\
&= \sum_{i\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} \int \frac{\|z_{i-k}\|^\alpha}{\|z\|_\alpha^\alpha} \frac{\|z_i\|^\alpha}{\|z\|_\alpha^\alpha} \mathbb{1}_A\left(\left(\frac{z_{s+i}}{\|z_i\|}\right)_{s\in\mathbb{Z}}\right) Q(dz)\\
&= \sum_{i\in\mathbb{Z}} \int \frac{\|z_i\|^\alpha}{\|z\|_\alpha^\alpha} \mathbb{1}_A\left(\left(\frac{z_{s+i}}{\|z_i\|}\right)_{s\in\mathbb{Z}}\right) Q(dz)
\end{aligned}$$

$$= Q^{RS}(A).$$

In the fourth step we have substituted $i = l+k$ and in the penultimate step $\sum_{k\in\mathbb{Z}} \|z_{i-k}\|^\alpha = \|z\|_\alpha^\alpha$ was applied. All the interchanges of sums and integrals made here can be justified by monotone convergence. This proves the assertion. □

5.7.2 Proofs for Section 5.2.1

First, we prove that $T_{n,A}$ is asymptotically unbiased.

Proof of Proposition 5.2.1. By stationarity

$$\frac{1}{nv_n}E[T_{n,A}] = \frac{1}{nv_n}E\Big[\sum_{t=1}^{n} g_A(W_{n,t})\Big] = E\left[g_A(W_{n,0}) \mid \|X_0\| > u_n\right].$$

The function $g_A : l_\alpha \to [0,1]$ is absolutely bounded by 1. Define the approximating functions $g_A^{(m)} : l_\alpha \to [0,1]$ by

$$g_A^{(m)}\Big((w_h)_{h\in\mathbb{Z}}\Big) := \frac{\mathbb{1}_{\{\|w_0\|>1\}}}{\sum_{|h|\le m}\|w_h\|^\alpha}\sum_{|h|\le m}\|w_h\|^\alpha \mathbb{1}_A\Big(\frac{w_{h+i}}{\|w_h\|}\Big) \tag{5.7.1}$$

for all $m \in \mathbb{N}$. This $g_A^{(m)}$ is bounded by 1. As a finite sum and composition of continuous functions, $g_A^{(m)}$ is continuous P^Y-a.s. if $P(\exists j \in \mathbb{Z} : Y_{j+i}/\|Y_j\| \in \partial C, \|Y_j\| > 0) = 0$ for $C \in \{A, B\}$ and $P(\|Y_j\| = 1) = 0$ for all $j \in \mathbb{Z}$. While the former equality is ensured by (PC) and Lemma 5.2.3, the latter follows from $Y_j = \Theta_j\|Y_0\|$, where Θ_j and $\|Y_0\|$ are independent and $\|Y_0\|$ has a Pareto(α)-distribution.
Thus, the weak convergence defining the tail process implies

$$\lim_{n\to\infty} E\left[g_A^{(m)}(W_{n,0}) \mid \|X_0\| > u_n\right] = E\left[g_A^{(m)}((Y_h)_{h\in\mathbb{Z}})\right].$$

Since $g_A^{(m)}(w) \to g_A(w)$ as $m \to \infty$ for all $w \in l_\alpha$, and $|g_A^{(m)}| \le 1$, $m \in \mathbb{N}$, dominated convergence implies $\lim_{m\to\infty} E\left[g_A^{(m)}((Y_h)_{h\in\mathbb{Z}})\right] = E\left[g_A((Y_h)_{h\in\mathbb{Z}})\right]$.
In the next calculations we use an idea similar to the following argument: for $a_h, b_k \ge 0$ one has

$$\begin{aligned}
&\sum_{|h|\le s_n} a_h \sum_{|k|\le s_n} b_k - \sum_{|h|\le m} a_h \sum_{|k|\le m} b_k \\
&= \sum_{|h|\le s_n} a_h \sum_{|k|\le s_n} b_k - \sum_{|h|\le s_n} a_h \sum_{|k|\le m} b_k + \sum_{|h|\le s_n} a_h \sum_{|k|\le m} b_k - \sum_{|h|\le m} a_h \sum_{|k|\le m} b_k \\
&= \sum_{|h|\le s_n} a_h \sum_{m<|k|\le s_n} b_k + \sum_{m<|h|\le s_n} a_h \sum_{|k|\le m} b_k. \qquad (5.7.2)
\end{aligned}$$

Then, w.l.o.g. for n sufficiently large such that $m+|i| \le s_n$ (this is only needed to simplify the notation with the indicators in the sums over $|h| \le m$ a little bit, the calculations remains true if $m+|i|>s_n$, but then one has to consider the indicator $h\in H_{n,i}$ for these sums),

$$\begin{aligned}
&\left|E\left[g_A^{(m)}(W_{n,0})-g_A(W_{n,0})\Big|\|X_0\|>u_n\right]\right|\\
&=\Bigg|E\Bigg[\sum_{|h|\le m}\frac{\|X_{n,h}\|^\alpha}{\sum_{|h|\le m}\|X_{n,h}\|^\alpha}\mathbb{1}_A\Big(\frac{X_{n,h+i}}{\|X_{n,h}\|}\Big)\\
&\quad-\sum_{|h|\le s_n}\frac{\|X_{n,h}\|^\alpha}{\sum_{|h|\le s_n}\|X_{n,h}\|^\alpha}\Big(\mathbb{1}_{\{h\in H_{n,i}\}}\mathbb{1}_A\Big(\frac{X_{n,h+i}}{\|X_{n,h}\|}\Big)+\mathbb{1}_{\{h\in H_{n,i}^C\}}\mathbb{1}_A(0)\Big)\Big|\|X_0\|>u_n\Bigg]\Bigg|\\
&=\Bigg|E\Bigg[\frac{1}{\sum_{|h|\le m}\|X_{n,h}\|^\alpha\sum_{|h|\le s_n}\|X_{n,h}\|^\alpha}\Bigg(\sum_{|h|\le m}\|X_{n,h}\|^\alpha\mathbb{1}_A\Big(\frac{X_{n,h+i}}{\|X_{n,h}\|}\Big)\sum_{|h|\le s_n}\|X_{n,h}\|^\alpha\\
&\quad-\sum_{|h|\le m}\|X_{n,h}\|^\alpha\mathbb{1}_A\Big(\frac{X_{n,h+i}}{\|X_{n,h}\|}\Big)\sum_{|h|\le m}\|X_{n,h}\|^\alpha\\
&\quad+\sum_{|h|\le m}\|X_{n,h}\|^\alpha\mathbb{1}_A\Big(\frac{X_{n,h+i}}{\|X_{n,h}\|}\Big)\sum_{|h|\le m}\|X_{n,h}\|^\alpha\\
&\quad-\sum_{|h|\le s_n}\|X_{n,h}\|^\alpha\sum_{|k|\le m}\|X_{n,k}\|^\alpha\\
&\qquad\qquad\times\Big(\mathbb{1}_{\{h\in H_{n,i}\}}\mathbb{1}_A\Big(\frac{X_{n,h+i}}{\|X_{n,h}\|}\Big)+\mathbb{1}_{\{h\in H_{n,i}^C\}}\mathbb{1}_A(0)\Big)\Bigg)\ \Big|\ \|X_0\|>u_n\Bigg]\Bigg|\\
&=\Bigg|E\Bigg[\frac{\sum_{|h|\le m}\|X_{n,h}\|^\alpha\mathbb{1}_A\Big(\frac{X_{n,h+i}}{\|X_{n,h}\|}\Big)\sum_{m<|h|\le s_n}\|X_{n,h}\|^\alpha}{\sum_{|h|\le m}\|X_{n,h}\|^\alpha\sum_{|h|\le s_n}\|X_{n,h}\|^\alpha}\\
&\quad-\frac{\sum_{m<|h|\le s_n}\|X_{n,h}\|^\alpha\Big(\mathbb{1}_{\{h\in H_{n,i}\}}\mathbb{1}_A\Big(\frac{X_{n,h+i}}{\|X_{n,h}\|}\Big)+\mathbb{1}_{\{h\in H_{n,i}^C\}}\mathbb{1}_A(0)\Big)}{\sum_{|h|\le s_n}\|X_{n,h}\|^\alpha}\ \Big|\ \|X_0\|>u_n\Bigg]\Bigg|\\
&\le E\left[\frac{\sum_{m<|h|\le s_n}\|X_{n,h}\|^\alpha}{\sum_{|h|\le s_n}\|X_{n,h}\|^\alpha}\ \Big|\ \|X_0\|>u_n\right]+E\left[\frac{\sum_{m<|h|\le s_n}\|X_{n,h}\|^\alpha}{\sum_{|h|\le s_n}\|X_{n,h}\|^\alpha}\ \Big|\ \|X_0\|>u_n\right]\\
&=2E\left[\frac{\sum_{m<|h|\le s_n}\|X_h\|^\alpha}{\sum_{|h|\le s_n}\|X_h\|^\alpha}\ \Big|\ \|X_0\|>u_n\right].
\end{aligned}\tag{5.7.3}$$

We state the next argument with some additional parameter $j\in\mathbb{Z}$, since we want to apply this argument later on in the more general form: Applying condition (PP) for some $0<c<1$ for which (PT) holds, we conclude

$$\begin{aligned}
&E\left[\frac{\sum_{m<|h|\le s_n}\|X_{h+j}\|^\alpha\mathbb{1}_{\{\|X_{h+j}\|>cu_n\}}}{\sum_{|h|\le s_n}\|X_{h+j}\|^\alpha}\Big|\|X_0\|>u_n\right]\\
&\le\sum_{m<|h|\le s_n}P(\|X_{h+j}\|>cu_n\mid\|X_0\|>cu_n)\frac{P(\|X_0\|>cu_n)}{P(\|X_0\|>u_n)}\le 4c^{-\alpha}\sum_{m<|h|\le s_n}e_{n,c}(h+j)
\end{aligned}$$

for sufficiently large n due to regular variation of $\|X_0\|$. Therefore,

$$E\left[\frac{\sum_{m<|h|\le s_n}\|X_{h+j}\|^\alpha}{\sum_{|h|\le s_n}\|X_{h+j}\|^\alpha}\Big|\|X_0\|>u_n\right]$$

$$
\begin{aligned}
&= E\left[\frac{\sum_{m<|h|\le s_n}\|X_{h+j}\|^\alpha \mathbb{1}_{\{\|X_{h+j}\|>cu_n\}}}{\sum_{|h|\le s_n}\|X_{h+j}\|^\alpha}\middle|\|X_0\|>u_n\right]\\
&\qquad + E\left[\frac{\sum_{m<|h|\le s_n}\|X_{h+j}\|^\alpha \mathbb{1}_{\{\|X_{h+j}\|\le cu_n\}}}{\sum_{|h|\le s_n}\|X_{h+j}\|^\alpha}\middle|\|X_0\|>u_n\right]\\
&\le 4c^{-\alpha}\sum_{m<|h|\le s_n} e_{n,c}(h+j) + E\left[\frac{\sum_{m<|h|\le s_n}\|X_{h+j}\|^\alpha \mathbb{1}_{\{\|X_{h+j}\|\le cu_n\}}}{\sum_{|h|\le s_n}\|X_{h+j}\|^\alpha}\middle|\|X_0\|>u_n\right] \qquad (5.7.4)
\end{aligned}
$$

for all $j\in\mathbb{Z}$ and sufficiently large $n\in\mathbb{N}$. Thus, conditions (PP) and (PT) imply

$$
\lim_{m\to\infty}\limsup_{n\to\infty}|E[g_A^m(W_{n,0})-g_A(W_{n,0})\mid\|X_0\|>u_n]|=0. \qquad (5.7.5)
$$

Therefore,

$$
\begin{aligned}
\lim_{n\to\infty}E[g_A(W_{n,0})\mid\|X_0\|>u_n] &= \lim_{m\to\infty}\lim_{n\to\infty}E[g_A^{(m)}(W_{n,0})\mid\|X_0\|>u_n]\\
&= \lim_{m\to\infty}E[g_A^{(m)}((Y_h)_{h\in\mathbb{Z}})]=E[g_A((Y_h)_{h\in\mathbb{Z}})]
\end{aligned}
$$

Finally,

$$
\begin{aligned}
&E\left[g_A((Y_t)_{t\in\mathbb{Z}})\right] = E\Big[\mathbb{1}_{\{\|Y_0\|>1\}}\sum_{h\in\mathbb{Z}}\frac{\|Y_h\|^\alpha}{\|Y\|_\alpha^\alpha}\mathbb{1}_A\Big(\frac{Y_{h+i}}{\|Y_h\|}\Big)\Big]\\
&= E\Big[\sum_{h\in\mathbb{Z}}\frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha}\mathbb{1}_A\Big(\frac{\Theta_{h+i}}{\|\Theta_h\|}\Big)\Big] = E\left[\mathbb{1}_A\left(\Theta_i^{RS}\right)\right]=P(\Theta_i^{RS}\in A)=P(\Theta_i\in A),
\end{aligned}
$$

where the last step holds due to the invariance of the spectral tail process under the RS-transformation. This concludes the proof. □

The two following proofs establish the lemmas regarding the conditions (PP) and (PC).

Proof of Lemma 5.2.2. We start with the case $c\ge 1$. Then,

$$
\begin{aligned}
P(\|X_k\|>u_nc|\|X_0\|>u_nc) &\le P(\|X_{|k|}\|>u_nc\mid\|X_0\|>u_n)\frac{P(\|X_0\|>u_n)}{P(\|X_0\|>u_nc)}\\
&\le P(\|X_{|k|}\|>u_n\mid\|X_0\|>u_n)\frac{P(\|X_0\|>u_n)}{P(\|X_0\|>u_nc)}\\
&\le 2c^\alpha e_n(|k|)
\end{aligned}
$$

for all $k\in\mathbb{Z}$ and a sufficiently large $n\in\mathbb{N}$. The last inequality holds by the regular variation of X and (PP). For $k<0$ we also applied the stationarity of X.
Therefore, by condition (PP) we obtain

$$
\begin{aligned}
&\lim_{m\to\infty}\limsup_{n\to\infty}P\left(\max_{m\le|t|\le r_n}\|X_t\|>u_nc\middle|\|X_0\|>u_nc\right)\\
&\le\lim_{m\to\infty}\limsup_{n\to\infty}\sum_{m\le|t|\le r_n}P(\|X_t\|>u_nc\mid\|X_0\|>u_nc)
\end{aligned}
$$

$$\leq \lim_{m\to\infty} 4c^\alpha \sum_{t=m}^{\infty} \lim_{n\to\infty} e_n(k) = 0.$$

Now, consider $c < 1$. Here the assertion is a direct consequence of (PP):

$$\begin{aligned}
&\lim_{m\to\infty} \limsup_{n\to\infty} P\left(\max_{m\leq |t|\leq r_n} \|X_t\| > u_n c \middle| \|X_0\| > u_n c \right) \\
&\leq \lim_{m\to\infty} \limsup_{n\to\infty} \sum_{m\leq |t|\leq r_n} P(\|X_t\| > u_n c \mid \|X_0\| > u_n c) \\
&\leq \lim_{m\to\infty} 2 \sum_{t=m}^{\infty} \lim_{n\to\infty} e_{n,c}(k) = 0.
\end{aligned}$$

This shows that (AC) holds. □

Proof of Lemma 5.2.3. Using $\Theta_i \overset{d}{=} \Theta_i^{RS}$, which holds since Θ is a spectral tail process (cf. Theorem 2.2.5), we obtain

$$P(\Theta_i \in \partial A) = P(\Theta_i^{RS} \in \partial A) = \int \sum_{h\in\mathbb{Z}} \frac{\|\theta_h\|^\alpha}{\|\theta\|_\alpha^\alpha} \mathbb{1}_{\partial A}\Big(\frac{\theta_{h+i}}{\|\theta_h\|}\Big) P^\Theta(d\theta).$$

Hence, if $P(\Theta_i \in \partial A) = 0$, it is

$$\begin{aligned}
0 &= P\left(\sum_{h\in\mathbb{Z}} \frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbb{1}_{\partial A}\left(\frac{\Theta_{h+i}}{\|\Theta_h\|}\right) > 0 \right) = P\left(\sum_{h\in\mathbb{Z}} \mathbb{1}_{\{\|\Theta_h\|>0\}} \mathbb{1}_{\partial A}\left(\frac{\Theta_{h+i}}{\|\Theta_h\|}\right) > 0 \right) \\
&= P\left(\exists h \in \mathbb{Z} : \frac{\Theta_{h+i}}{\|\Theta_h\|} \in \partial A, \|\Theta_t\| > 0 \right) = P\left(\exists t \in \mathbb{Z} : \frac{Y_{t+i}}{\|Y_t\|} \in \partial A, \|Y_t\| > 0 \right).
\end{aligned}$$

In the last step we used $\|\Theta_t\| \overset{d}{=} \|Y_t\|/\|Y_0\|$ for all $t \in \mathbb{Z}$ and $\|Y_0\| \geq 1$ a.s. since $\|Y_0\|$ is Par(α) distributed.
Conversely, starting with $P(\exists t \in \mathbb{Z} : Y_{t+i}/\|Y_t\| \in \partial A, \|Y_t\| > 0) = 0$, it follows

$$\begin{aligned}
P(\Theta_i \in \partial A) &\leq \int \sum_{h\in\mathbb{Z}} \mathbb{1}_{\{\|\theta_h\|>0\}} \mathbb{1}_{\partial A}\left(\frac{\theta_{h+i}}{\|\theta_h\|}\right) P^\Theta(d\theta) \\
&= \int \sum_{h\in\mathbb{Z}} \mathbb{1}_{\{\|y_h\|>0\}} \mathbb{1}_{\partial A}\left(\frac{y_{h+i}}{\|y_h\|}\right) P^Y(dy) = 0.
\end{aligned}$$

This proves the assertion. □

Next we turn to the main proofs of Section 5.2.1. For these proofs we will apply the theory of Section 3.2, but for simpler presentation of the results developed here, we diverge from the notation in that chapter. In order to directly fit in the setting of Section 3.2 one has to consider the transformed random variables $X'_{n,t} = X_{n,t-s_n}$, $1 \leq t \leq n' := n + 2s_n$ and rename the block lengths $s'_n := 2s_n + 1$ and $l'_n := 2s'_n - 1$. Furthermore, we need some family of functions, which is given here by $\mathcal{G} = \{g_A | A \in \mathcal{A}\}$.

In line with the notation in Section 3.2, we denote $m_n = \lfloor n/r_n \rfloor$ for the sequence r_n, $n \in \mathbb{N}$. We have $v_n = P(\|X_0\| > u_n)$ and

$$p_n = P\Big(\sum_{t=1}^{r_n} g_{\mathbb{R}^d}(W_{n,t}) \neq 0\Big) = P(\exists 1 \leq t \leq r_n : \|X_t\| > u_n).$$

In the setting of Theorem 3.2.1 we choose $b_n(g) = \sqrt{nv_n/p_n}$ uniformly for all $g \in \mathcal{G}$. As already discussed on page 78, under conditions (PR), (P0) (which contains (θ1)) and (PP) (which is stronger than (θP)), (4.2.1) holds, i.e.

$$\frac{p_n}{r_n v_n} = \frac{P(\exists 1 \leq t \leq r_n : \|X_t\| > u_n)}{r_n P(\|X_0\| > u_n)} \to \theta > 0,$$

in particular, p_n is of the same order as $r_n v_n$. (By (2.1.2), and thus due to Lemma 5.2.2 under (PP), this was also shown by Basrak and Segers (2009), Proposition 4.2.) Thus, the choice $b_n(g_A) = \sqrt{nv_n/p_n}$ for all $g_A \in \mathcal{G}$ is of the order $\sqrt{m_n}$. Hence, the normalization in Condition (C) in Section 3.2 is asymptotically equivalent to $(r_n v_n)^{-1}$ as considered in Lemma 5.2.4 (i.e. the normalization Condition (C) divided by $(r_n v_n)^{-1}$ converges to 1).

Proof of Lemma 5.2.4. We calculate the covariance straightforwardly. It holds

$$\begin{aligned}
&Cov\Big(\sum_{t=1}^{r_n} g_A(W_{n,t}), \sum_{t=1}^{r_n} g_B(W_{n,t})\Big) \\
&\quad = E\Big[\sum_{t=1}^{r_n}\sum_{s=1}^{r_n} g_A(W_{n,t}) g_B(W_{n,s})\Big] - E\Big[\sum_{t=1}^{r_n} g_A(W_{n,t})\Big] E\Big[\sum_{t=1}^{r_n} g_B(W_{n,t})\Big] =: I + II.
\end{aligned}$$

In the following, both summands I and II will be considered separately. We start with I. By the stationarity and since $g_A(W_{n,0}) = 0$ if $\|X_0\| < u_n$, we have

$$\begin{aligned}
I &= E\Big[\sum_{t=1}^{r_n}\sum_{s=1}^{r_n} g_A(W_{n,t}) g_B(W_{n,s})\Big] \qquad (5.7.6)\\
&= \sum_{j=-r_n}^{r_n} (r_n - |j|) E\left[g_A(W_{n,0}) g_B(W_{n,j})\right] \\
&= r_n v_n \sum_{j=-r_n}^{r_n} \Big(1 - \frac{|j|}{r_n}\Big) E\left[g_A(W_{n,0}) g_B(W_{n,j}) \mid \|X_0\| > u_n\right] \\
&= r_n v_n \sum_{j=-r_n}^{r_n} \Big(1 - \frac{|j|}{r_n}\Big) E\Big[\mathbb{1}_{\{\|X_j\| > u_n\}} \Big(\sum_{h \in H_{n,i}} \frac{\|X_{j+h}\|^\alpha}{\sum_{k=-s_n}^{s_n} \|X_{j+k}\|^\alpha} \mathbb{1}_A\Big(\frac{X_{j+h+i}}{\|X_{j+h}\|}\Big) \\
&\qquad\qquad + \sum_{h \in H_{n,i}^C} \frac{\|X_{j+h}\|^\alpha}{\sum_{k=-s_n}^{s_n} \|X_{j+k}\|^\alpha} \mathbb{1}_A(0)\Big) \\
&\quad \times \Big(\sum_{l \in H_{n,i}} \frac{\|X_l\|^\alpha}{\sum_{k=-s_n}^{s_n} \|X_k\|^\alpha} \mathbb{1}_B\Big(\frac{X_{l+i}}{\|X_l\|}\Big) + \sum_{l \in H_{n,i}^C} \frac{\|X_l\|^\alpha}{\sum_{k=-s_n}^{s_n} \|X_k\|^\alpha} \mathbb{1}_B(0)\Big) \,\Big|\, \|X_0\| > u_n\Big].
\end{aligned}$$

Define the function $f = f_{A,B,j} : l_\alpha \times l_\alpha \to [0,1]$ by

$$f((y_t)_{t\in\mathbb{Z}}, (z_t)_{t\in\mathbb{Z}}) := \mathbb{1}_{\{\|y_0\|>1\}}\mathbb{1}_{\{\|z_0\|>1\}} \times \left(\sum_{h\in\mathbb{Z}} \frac{\|z_h\|^\alpha}{\sum_{k\in\mathbb{Z}}\|z_k\|^\alpha}\mathbb{1}_A\Big(\frac{z_{h+i}}{\|z_h\|}\Big)\right)\left(\sum_{l\in\mathbb{Z}} \frac{\|y_l\|^\alpha}{\sum_{k\in\mathbb{Z}}\|y_k\|^\alpha}\mathbb{1}_B\Big(\frac{y_{l+i}}{\|y_l\|}\Big)\right)$$

with the convention $0/0 := 0$. This function is obviously bounded with $0 \le f \le 1$. By the usual embedding of $(X_{n,t})_{|t|\le s_n}$ in l_α through $X_{n,t} = 0$ for $|t| > s_n$ we have

$$I = r_n v_n \sum_{j=-r_n}^{r_n}\Big(1 - \frac{|j|}{r_n}\Big)E\Big[f(W_{n,0}, W_{n,j}) \mid \|X_0\| > u_n\Big].$$

Define the approximating function $f^{(m)} : l_\alpha \times l_\alpha \to [0,1]$ by

$$f^{(m)}((y_t)_{t\in\mathbb{Z}}, (z_t)_{t\in\mathbb{Z}}) := \mathbb{1}_{\{\|y_0\|>1\}}\mathbb{1}_{\{\|z_0\|>1\}} \times \left(\sum_{|h|\le m} \frac{\|z_h\|^\alpha}{\sum_{|k|\le m}\|z_k\|^\alpha}\mathbb{1}_A\left(\frac{z_{h+i}}{\|z_h\|}\right)\right)\left(\sum_{|l|\le m} \frac{\|y_l\|^\alpha}{\sum_{|k|\le m}\|y_k\|^\alpha}\mathbb{1}_B\left(\frac{y_{l+i}}{\|y_l\|}\right)\right)$$

for all $m \in \mathbb{Z}$. Check that $f^{(m)}(y,z) = g_B^{(m)}(y)g_A^{(m)}(z)$ and $f(y,z) = g_B(y)g_A(z)$ with $g_A^{(m)}$ defined in (5.7.1). Note also, that $f^{(m)}$ is as a finite sum and combination of continuous functions $P^{((Y_h)_{h\in\mathbb{Z}},(Y_{h+j})_{h\in\mathbb{Z}})}$-a.s. continuous, if the occurring indicator functions are P^Y-a.s. continuous, i.e. if

$$\begin{aligned}0 &= P\left(\left\{\exists C \in \{A,B\}, j,h \in \mathbb{Z} : \frac{Y_{j+h+i}}{\|Y_{j+h}\|} \in \partial C, \|Y_{j+h}\| > 0\right\} \cup \{\exists j \in \mathbb{Z} : \|Y_j\| = 1\}\right)\\ &\le \sum_{C\in\{A,B\}} P\left(\exists j \in \mathbb{Z} : \frac{Y_{j+i}}{\|Y_j\|} \in \partial C, \|Y_j\| > 0\right) + P(\exists j \in \mathbb{Z} : \|Y_j\| = 1).\end{aligned}$$

Outside of these sets $f^{(m)}$ is a continuous function. The indicators are P^Y-a.s. continuous, if $P(\exists j \in \mathbb{Z} : Y_{j+i}/\|Y_j\| \in \partial C, \|Y_j\| > 0) = 0$ for $C \in \{A,B\}$ and $P(\|Y_j\| = 1) = 0$ for all $j \in \mathbb{Z}$. The former applies on the basis of condition (PC) and Lemma 5.2.3. The latter follows from the fact that $\|Y_j\| \overset{d}{=} \|\Theta_j\|\|Y_0\|$, where $\|\Theta_j\|$ and $\|Y_0\|$ are independent and $\|Y_0\|$ has a Par(α) distribution. Thus, $f^{(m)}$ is $P^{((Y_h)_{h\in\mathbb{Z}},(Y_{h+j})_{h\in\mathbb{Z}})}$-a.s. continuous. Therefore, by the weak convergence defining the tail process

$$E\left[f^{(m)}((X_t/u_n)_{|t|\le s_n}, (X_{t+j}/u_n)_{|t|\le s_n}) \mid \|X_0\| > u_n\right] \to E\left[f^{(m)}((Y_t)_{t\in\mathbb{Z}}, (Y_{t+j})_{t\in\mathbb{Z}})\right].$$

Since $\lim_{m\to\infty} f^{(m)}(y,z) = f(y,z)$ for all $y,z \in l_\alpha$ and $|f^{(m)}| \le 1$, dominated convergence implies $\lim_{m\to\infty} E\left[f^{(m)}((Y_t)_{t\in\mathbb{Z}}, (Y_{t+j})_{t\in\mathbb{Z}})\right] = E\left[f((Y_t)_{t\in\mathbb{Z}}, (Y_{t+j})_{t\in\mathbb{Z}})\right]$. Moreover, using $|g_A| \le 1$ yields

$$|E\left[f^{(m)}(W_{n,0}, w_{n,j}) - f(W_{n,0}, W_{n,j})\Big|\|X_0\| > u_n\right]|$$

$$
\begin{aligned}
&= |E\Big[g_A^{(m)}(W_{n,0})g_B^{(m)}(W_{n,j}) - g_A^{(m)}(W_{n,0})g_B(W_{n,j}) \\
&\qquad + g_A^{(m)}(W_{n,0})g_B(W_{n,j}) - g_A(W_{n,0})g_B(W_{n,j})\Big|\|X_0\| > u_n\Big]| \\
&\le E\Big[|g_A^{(m)}(W_{n,0})(g_B^{(m)}(W_{n,j}) - g_B(W_{n,j}))| \\
&\qquad + |(g_A^{(m)}(W_{n,0}) - g_A(W_{n,0}))g_B(W_{n,j})|\Big|\|X_0\| > u_n\Big] \\
&\le E\left[|(g_B^{(m)}(W_{n,j}) - g_B(W_{n,j}))| + |(g_A^{(m)}(W_{n,0}) - g_A(W_{n,0}))|\Big|\|X_0\| > u_n\right]| \\
&\le 2E\left[\frac{\sum_{m<|h|\le s_n}\|X_{h+j}\|^\alpha}{\sum_{|h|\le s_n}\|X_{h+j}\|^\alpha}\Big|\|X_0\| > u_n\right] + 2E\left[\frac{\sum_{m<|h|\le s_n}\|X_h\|^\alpha}{\sum_{|h|\le s_n}\|X_h\|^\alpha}\Big|\|X_0\| > u_n\right],
\end{aligned}
$$

where the last inequality holds by (5.7.3). Thus, by (5.7.4) conditions (PP) and (PT) imply

$$
\lim_{m\to\infty}\limsup_{n\to\infty}|E\left[f^{(m)}(W_{n,0}, W_{n,j}) - f(W_{n,0}, W_{n,j}) \mid \|X_0\| > u_n\right]| = 0.
$$

Combine this with the previous results for $f^{(m)}$ to conclude

$$
\begin{aligned}
&\lim_{n\to\infty} E[f(W_{n,0}, W_{n,j}) \mid \|X_0\| > u_n] \\
&\quad = \lim_{m\to\infty}\lim_{n\to\infty} E[f^{(m)}(W_{n,0}, W_{n,j}) \mid \|X_0\| > u_n] \\
&\quad = \lim_{m\to\infty} E[f^{(m)}((Y_h)_{h\in\mathbb{Z}}, (Y_{h+j})_{h\in\mathbb{Z}})] = E[f((Y_h)_{h\in\mathbb{Z}}, (Y_{h+j})_{h\in\mathbb{Z}})] \\
&\quad = E\left[\mathbb{1}_{\{\|Y_j\|>1\}}\left(\sum_{h\in\mathbb{Z}}\frac{\|Y_{j+h}\|^\alpha}{\sum_{k\in\mathbb{Z}}\|Y_{j+k}\|^\alpha}\mathbb{1}_A\left(\frac{Y_{j+h+i}}{\|Y_{j+h}\|}\right)\right)\left(\sum_{l\in\mathbb{Z}}\frac{\|Y_l\|^\alpha}{\sum_{k\in\mathbb{Z}}\|Y_k\|^\alpha}\mathbb{1}_B\left(\frac{Y_{l+i}}{\|Y_l\|}\right)\right)\right]
\end{aligned}
$$

for all $j \in \mathbb{Z}$. Condition (PP), and for $j < 0$ stationarity, imply

$$
\begin{aligned}
&\left(1 - \frac{|j|}{r_n}\right)E\left[f((X_t/u_n)_{|t|\le s_n}, (X_{t+j}/u_n)_{|t|\le s_n}) \mid \|X_0\| > u_n\right] \\
&\quad \le E\left[\mathbb{1}_{\{\|X_j\|>u_n\}} \mid \|X_0\| > u_n\right] = P(\|X_{|j|}\| > u_n \mid \|X_0\| > u_n) \le e_n(|j|)
\end{aligned}
$$

for all $j \in \mathbb{Z}$. Therefore, with condition (PP), by Pratt's Lemma (Pratt, 1960)

$$
\begin{aligned}
&\sum_{j=-r_n}^{r_n}\left(1 - \frac{|j|}{r_n}\right)E\left[f((X_t/u_n)_{|t|\le s_n}, (X_{t+j}/u_n)_{|t|\le s_n}) \mid \|X_0\| > u_n\right] \\
&\to \sum_{j\in\mathbb{Z}} E\left[\mathbb{1}_{\{\|Y_j\|>1\}}\left(\sum_{h\in\mathbb{Z}}\frac{\|Y_{j+h}\|^\alpha}{\sum_{k\in\mathbb{Z}}\|Y_{j+k}\|^\alpha}\mathbb{1}_A\left(\frac{Y_{j+h+i}}{\|Y_{j+h}\|}\right)\right)\left(\sum_{l\in\mathbb{Z}}\frac{\|Y_l\|^\alpha}{\sum_{k\in\mathbb{Z}}\|Y_k\|^\alpha}\mathbb{1}_B\left(\frac{Y_{l+i}}{\|Y_l\|}\right)\right)\right].
\end{aligned}
$$

Thus, for I from (5.7.6) we obtain

$$
\begin{aligned}
&\frac{I}{r_n v_n} = \frac{1}{r_n v_n}E\left[\sum_{t=1}^{r_n}\sum_{s=1}^{r_n} g_A(W_{n,t})g_B(W_{n,s})\right] \\
&\to \sum_{j\in\mathbb{Z}} E\left[\mathbb{1}_{\{\|Y_j\|>1\}}\left(\sum_{h\in\mathbb{Z}}\frac{\|Y_{j+h}\|^\alpha}{\sum_{k\in\mathbb{Z}}\|Y_{j+k}\|^\alpha}\mathbb{1}_A\left(\frac{Y_{j+h+i}}{\|Y_{j+h}\|}\right)\right)\left(\sum_{l\in\mathbb{Z}}\frac{\|Y_l\|^\alpha}{\sum_{k\in\mathbb{Z}}\|Y_k\|^\alpha}\mathbb{1}_B\left(\frac{Y_{l+i}}{\|Y_l\|}\right)\right)\right]
\end{aligned}
$$

Next, we consider the second summand II. Here one has

$$E\left[\sum_{t=1}^{r_n} g_A(Y_t)\right] = E\left[\sum_{t=1}^{r_n} \mathbb{1}_{\{\|X_t\|>u_n\}}\left(\sum_{h\in H_{n,i}} \frac{\|X_{t+h}\|^\alpha}{\sum_{k=-s_n}^{s_n}\|X_{t+k}\|^\alpha}\mathbb{1}_A\left(\frac{X_{t+h+i}}{\|X_{t+h}\|}\right)\right.\right.$$
$$\left.\left.+\sum_{h\in H_{n,i}^C}\frac{\|X_{t+h}\|^\alpha}{\sum_{k=-s_n}^{s_n}\|X_{t+k}\|^\alpha}\mathbb{1}_A(0)\right)\right]$$
$$\leq E\left[\sum_{t=1}^{r_n}\mathbb{1}_{\{\|X_t\|>u_n\}}\right] = r_n E\left[\mathbb{1}_{\{\|X_0\|>u_n\}}\right] = r_n v_n.$$

Thus, for II it follows

$$\frac{|II|}{r_n v_n} = \frac{1}{r_n v_n} E\left[\sum_{t=1}^{r_n} g_A(W_{n,t})\right] E\left[\sum_{t=1}^{r_n} g_B(W_{n,t})\right] = \mathrm{O}(r_n v_n) \to 0. \tag{5.7.7}$$

Here $r_n v_n \to 0$ applies per assumption. To sum up, the above calculations for I and II together now result in

$$\frac{1}{r_n v_n} Cov\left(\sum_{t=1}^{r_n} g_A(W_{n,t}), \sum_{t=1}^{r_n} g_B(W_{n,t})\right) \tag{5.7.8}$$
$$= \sum_{j=-r_n}^{r_n}\left(1-\frac{|j|}{r_n}\right)E\left[f_{A,B,j}((X_t/u_n)_{|t|\leq s_n},(X_{t+j}(u_n))_{|t|\leq s_n}) \mid \|X_0\|>u_n\right]+\mathrm{o}(1)$$
$$\to \sum_{j\in\mathbb{Z}} E\left[\mathbb{1}_{\{\|Y_j\|>1\}}\left(\sum_{h\in\mathbb{Z}}\frac{\|Y_{j+h}\|^\alpha}{\sum_{k\in\mathbb{Z}}\|Y_{j+k}\|^\alpha}\mathbb{1}_A\left(\frac{Y_{j+h+i}}{\|Y_{j+h}\|}\right)\right)\left(\sum_{l\in\mathbb{Z}}\frac{\|Y_l\|^\alpha}{\sum_{k\in\mathbb{Z}}\|Y_k\|^\alpha}\mathbb{1}_B\left(\frac{Y_{l+i}}{\|Y_l\|}\right)\right)\right]$$
$$= \sum_{j\in\mathbb{Z}} E\left[\mathbb{1}_{\{\|Y_j\|>1\}}\left(\sum_{h\in\mathbb{Z}}\frac{\|Y_h\|^\alpha}{\sum_{k\in\mathbb{Z}}\|Y_k\|^\alpha}\mathbb{1}_A\left(\frac{Y_{h+i}}{\|Y_h\|}\right)\right)\left(\sum_{l\in\mathbb{Z}}\frac{\|Y_l\|^\alpha}{\sum_{k\in\mathbb{Z}}\|Y_k\|^\alpha}\mathbb{1}_B\left(\frac{Y_{l+i}}{\|Y_l\|}\right)\right)\right]$$
$$= c(A,B).$$

Next, we will rewrite the limit $c(A,B)$. To this end, the relation $Y_t \overset{d}{=} \Theta_t\|Y_0\|$ is used, where $\|Y_0\|$ has a Par(α)-distribution, i.e. $P(\|Y_0\|>x) = x^{-\alpha}\wedge 1$, and $\|Y_0\|$ is independent of $(\Theta_t)_{t\in\mathbb{Z}}$. With this it follows for $j\in\mathbb{Z}$

$$E\left[\mathbb{1}_{\{\|Y_j\|>1\}}\left(\sum_{h\in\mathbb{Z}}\frac{\|Y_h\|^\alpha}{\|Y\|_\alpha^\alpha}\mathbb{1}_B\left(\frac{Y_{h+i}}{\|Y_h\|}\right)\right)\left(\sum_{l\in\mathbb{Z}}\frac{\|Y_l\|^\alpha}{\|Y\|_\alpha^\alpha}\mathbb{1}_A\left(\frac{Y_{l+i}}{\|Y_l\|}\right)\right)\right]$$
$$= E\left[\mathbb{1}_{\{\|\Theta_j\|\|Y_0\|>1\}}\left(\sum_{h\in\mathbb{Z}}\frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha}\mathbb{1}_B\left(\frac{\Theta_{h+i}}{\|\Theta_h\|}\right)\right)\left(\sum_{l\in\mathbb{Z}}\frac{\|\Theta_l\|^\alpha}{\|\Theta\|_\alpha^\alpha}\mathbb{1}_A\left(\frac{\Theta_{l+i}}{\|\Theta_l\|}\right)\right)\right]$$
$$= \int_{\mathbb{R}^d}\int_1^\infty \mathbb{1}_{\{y>\|\theta_j\|^{-1}\}}\left(\sum_{h\in\mathbb{Z}}\frac{\|\theta_h\|^\alpha}{\|\theta\|_\alpha^\alpha}\mathbb{1}_B\left(\frac{\theta_{h+i}}{\|\theta_h\|}\right)\right)$$
$$\times\left(\sum_{l\in\mathbb{Z}}\frac{\|\theta_l\|^\alpha}{\|\theta\|_\alpha^\alpha}\mathbb{1}_A\left(\frac{\theta_{l+i}}{\|\theta_l\|}\right)\right)P^{\|Y_0\|}(dy)P^\Theta(d\theta)$$

$$= \int (\|\theta_j\|^\alpha \wedge 1) \left(\sum_{h\in\mathbb{Z}} \frac{\|\theta_h\|^\alpha}{\|\theta\|_\alpha^\alpha} \mathbb{1}_B \left(\frac{\theta_{h+i}}{\|\theta_h\|} \right) \right) \left(\sum_{l\in\mathbb{Z}} \frac{\|\theta_l\|^\alpha}{\|\theta\|_\alpha^\alpha} \mathbb{1}_A \left(\frac{\theta_{l+i}}{\|\theta_l\|} \right) \right) P^\Theta(d\theta)$$
$$= E\left[(\|\Theta_j\|^\alpha \wedge 1) \left(\sum_{h\in\mathbb{Z}} \frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbb{1}_B \left(\frac{\Theta_{h+i}}{\|\Theta_h\|} \right) \right) \left(\sum_{l\in\mathbb{Z}} \frac{\|\Theta_l\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbb{1}_A \left(\frac{\Theta_{l+i}}{\|\Theta_l\|} \right) \right) \right]. \tag{5.7.9}$$

This results in

$$c(A,B) = \sum_{j\in\mathbb{Z}} E\left[(\|\Theta_j\|^\alpha \wedge 1) \left(\sum_{h\in\mathbb{Z}} \frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbb{1}_B \left(\frac{\Theta_{h+i}}{\|\Theta_h\|} \right) \right) \left(\sum_{l\in\mathbb{Z}} \frac{\|\Theta_l\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbb{1}_A \left(\frac{\Theta_{l+i}}{\|\Theta_l\|} \right) \right) \right],$$

which proves the claim. □

It follows the proof for the alternative representation of $c(A,B)$.

Proof of equation (5.2.3). In the beginning of the proof of Lemma 5.2.4, by stationarity, one could split up the sum in equation (5.7.6) into two sums summing over $1 \le j \le r_n$ and one summand for $j = 0$ instead of considering a single sum $-r_n \le j \le r_n$:

$$\begin{aligned} E\Big[\sum_{t=1}^{r_n} \sum_{s=1}^{r_n} g_A(W_{n,t}) g_B(W_{n,s}) \Big] =& E\left[g_A(W_{n,0}) g_B(W_{n,0}) \right] \\ &+ 2 \sum_{j=1}^{r_n} (r_n - |j|) E\left[g_A(W_{n,0}) g_B(W_{n,j}) \right]. \end{aligned}$$

Following the proof of Lemma 5.2.4 using the same arguments it follows that

$$\begin{aligned} c(A,B) = E\Big[\Big(\sum_{h\in\mathbb{Z}} \frac{\|Y_h\|^\alpha}{\|Y\|_\alpha^\alpha} \mathbb{1}_A \Big(\frac{Y_{h+i}}{\|Y_h\|} \Big) \Big) \Big(\sum_{l\in\mathbb{Z}} \frac{\|Y_l\|^\alpha}{\|Y\|_\alpha^\alpha} \mathbb{1}_B \Big(\frac{Y_{l+i}}{\|Y_l\|} \Big) \Big) \Big] \\ + 2 \sum_{j\in\mathbb{N}} E\Big[\mathbb{1}_{\{\|Y_j\|>1\}} \Big(\sum_{h\in\mathbb{Z}} \frac{\|Y_h\|^\alpha}{\|Y\|_\alpha^\alpha} \mathbb{1}_A \Big(\frac{Y_{h+i}}{\|Y_h\|} \Big) \Big) \Big(\sum_{l\in\mathbb{Z}} \frac{\|Y_l\|^\alpha}{\|Y\|_\alpha^\alpha} \mathbb{1}_B \Big(\frac{Y_{l+i}}{\|Y_l\|} \Big) \Big) \Big]. \end{aligned}$$

Note that $(\Theta_t)_{t\in\mathbb{Z}}$ satisfies the time change formula (TCF), which is why $(\Theta_t^{RS})_{t\in\mathbb{Z}} \overset{d}{=} (\Theta_t)_{t\in\mathbb{Z}}$. Hence, for the first summand in this representation one has

$$\begin{aligned} &E\left[\left(\sum_{h\in\mathbb{Z}} \frac{\|Y_h\|^\alpha}{\|Y\|_\alpha^\alpha} \mathbb{1}_A \left(\frac{Y_{h+i}}{\|Y_h\|} \right) \right) \left(\sum_{l\in\mathbb{Z}} \frac{\|Y_l\|^\alpha}{\|Y\|_\alpha^\alpha} \mathbb{1}_B \left(\frac{Y_{l+i}}{\|Y_l\|} \right) \right) \right] \\ &= E\left[\left(\sum_{h\in\mathbb{Z}} \frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbb{1}_A \left(\frac{\Theta_{h+i}}{\|\Theta_h\|} \right) \right) \left(\sum_{l\in\mathbb{Z}} \frac{\|\Theta_l\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbb{1}_B \left(\frac{\Theta_{l+i}}{\|\Theta_l\|} \right) \right) \right] \\ &= \int_{\mathbb{R}^d} \sum_{h\in\mathbb{Z}} \frac{\|\theta_h\|^\alpha}{\|\theta\|_\alpha^\alpha} \mathbb{1}_A \left(\frac{\theta_{h+i}}{\|\theta_h\|} \right) \left(\sum_{l\in\mathbb{Z}} \frac{\|\theta_{l+h}/\|\theta_h\|\|^\alpha}{\|\theta/\|\theta_h\|\|_\alpha^\alpha} \mathbb{1}_B \left(\frac{\theta_{l+h+i}/\|\theta_h\|}{\|\theta_{l+h}\|/\|\theta_h\|} \right) \right) P^\Theta(\theta) \\ &= \int_{\mathbb{R}^d} \mathbb{1}_A(\theta_i) \left(\sum_{l\in\mathbb{Z}} \frac{\|\theta_l\|^\alpha}{\|\theta\|_\alpha^\alpha} \mathbb{1}_B \left(\frac{\theta_{l+i}}{\|\theta_l\|} \right) \right) P^{\Theta^{RS}}(d\theta) \\ &= \int_{\mathbb{R}^d} \mathbb{1}_A(\theta_i) \left(\sum_{l\in\mathbb{Z}} \frac{\|\theta_l\|^\alpha}{\|\theta\|_\alpha^\alpha} \mathbb{1}_B \left(\frac{\theta_{l+i}}{\|\theta_l\|} \right) \right) P^\Theta(d\theta) \end{aligned}$$

$$= E\left[\mathbb{1}_A(\Theta_i)\left(\sum_{l\in\mathbb{Z}}\frac{\|\Theta_l\|^\alpha}{\|\Theta\|_\alpha^\alpha}\mathbb{1}_B\left(\frac{\Theta_{l+i}}{\|\Theta_l\|}\right)\right)\right].$$

All remaining summands in the previous representation can be transformed as in (5.7.9), which leads to the last representation in (5.2.3):

$$\begin{aligned}c(A,B) = E\Big[\mathbb{1}_A(\Theta_i)\Big(\sum_{l\in\mathbb{Z}}\frac{\|\Theta_l\|^\alpha}{\|\Theta\|_\alpha^\alpha}\mathbb{1}_B\Big(\frac{\Theta_{l+i}}{\|\Theta_l\|}\Big)\Big)\Big]\\ + 2\sum_{j\in\mathbb{N}}E\Big[(\|\Theta_j\|^\alpha\wedge 1)\Big(\sum_{h\in\mathbb{Z}}\frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha}\mathbb{1}_A\Big(\frac{\Theta_{h+i}}{\|\Theta_h\|}\Big)\Big)\Big(\sum_{l\in\mathbb{Z}}\frac{\|\Theta_l\|^\alpha}{\|\Theta\|_\alpha^\alpha}\mathbb{1}_B\Big(\frac{\Theta_{l+i}}{\|\Theta_l\|}\Big)\Big)\Big].\square\end{aligned}$$

As mentioned before, for the proof of Proposition 5.2.5 we apply Theorem 3.2.1 and verify that the necessary conditions are implied by (PR), (P0), (PP) and (PC).

Proof of Proposition 5.2.5. To prove this proposition we apply part (a) of the sliding blocks limit Theorem 3.2.1. First, notice that we have to rename the observations $X'_t = X_{t-s_n}$, $1 \le t \le n' = n+2s_n$ such that we are in exactly the setting of Theorem 3.2.1 with $s'_n = 2s_n+1$, $l'_n = 2s'_n - 1$. Moreover, recall that condition (PP) implies (2.1.2) (Lemma 5.2.2) and that therefore p_n is of the same order as $r_n v_n$.
The conditions (A), (A2) and (MX) of Theorem 3.2.1 are direct consequences of assumption (P0). Moreover, condition (D0) is directly fulfilled for finite families of sets $A \in \mathcal{A}$, which suffices here.
Since $0 \le g_A \le 1$ for all $g \in \mathcal{G}$ we can apply Theorem 3.2.1, for which condition (3.2.4) has to be verified. By the definition of g_A it follows $\mathbb{1}_{\{g(W_{n,j})\neq 0\}} \le \mathbb{1}_{\{\|X_j\|>u_n\}}$ for all $j = 1, ..., r_n$ and for all $g \in \mathcal{G}$. Therefore,

$$E\Big[\Big(\sum_{j=1}^{r_n}\mathbb{1}_{\{g(W_{n,j})\neq 0\}}\Big)^2\Big] \le E\Big[\Big(\sum_{j=1}^{r_n}\mathbb{1}_{\{\|X_j\|>u_n\}}\Big)^2\Big].$$

Hence, (3.2.4) is satisfied if

$$E\Big[\Big(\sum_{j=1}^{r_n}\mathbb{1}_{\{\|X_j\|>u_n\}}\Big)^2\Big] = \mathrm{O}\Big(\frac{p_n b_n(g)^2}{m_n}\Big) = \mathrm{O}\Big(\frac{nv_n}{m_n}\Big) = \mathrm{O}(r_n v_n). \tag{5.7.10}$$

To verify this, we apply Lemma 3.2.4 with the modified function $\tilde{g}((y_h)_{h\in\mathbb{Z}}) = \mathbb{1}_{\{\|y_0\|>1\}}$. By Lemma 3.2.4 this last equation is fulfilled, if condition (S) is satisfied for the function $\tilde{g}$. We can choose $e_{\tilde{g},n}(k) := e_n(k)$ for all $k \ge 1$, for $e_n(k)$ given in condition (PP). Thus, condition (S) directly follows from (PP). For this, note that we consider $b_n(g) = (nv_n/p_n)^{1/2}$ and therefore $p_n b_n(g)^2 = nv_n$. Hence, condition (3.2.4) holds for all $g \in \mathcal{G}$.
The condition (C) has been proved in Lemma 5.2.4. Now the assertion follows from Theorem 3.2.1. $\square$

Finally, we turn to the last proof of this section, namely the verification of the asymptotic normality of $\hat{p}_{n,A}$.

Proof of Theorem 5.2.6. By construction $\hat{p}_{n,A} = T_{n,A}/T_{n,\mathbb{R}^d}$ and the asymptotic behavior of $T_{n,A}$ is known from Proposition 5.2.5. Furthermore, we have $E[T_{n,A}] = nE\left[g_A(W_{n,0})\right] = nv_nE\left[g_A(W_{n,0}) \mid \|X_0\| > u_n\right]$. The asymptotic behavior of $\hat{p}_{n,A}$ should be derived from this by a continuous mapping argument. Direct calculations yield

$$\begin{aligned}
&\sqrt{nv_n}\left(\hat{p}_{n,A} - E\left[g_A(W_{n,0}) \mid \|X_0\| > u_n\right]\right)\\
&= \sqrt{nv_n}\left(\frac{T_{n,A}}{T_{n,\mathbb{R}^d}} - E\left[g_A(W_{n,0}) \mid \|X_0\| > u_n\right]\right)\\
&= \sqrt{nv_n}\frac{\sqrt{nv_n}Z_n(A) + E[T_{n,A}] - E\left[g_A(W_{n,0}) \mid \|X_0\| > u_n\right]\sqrt{nv_n}Z_n(\mathbb{R}^d) - E[T_{n,A}]}{\sqrt{nv_n}Z_n(\mathbb{R}^d) + nv_n}\\
&= \frac{Z_n(A) - E\left[g_A(W_{n,0}) \mid \|X_0\| > u_n\right]Z_n(\mathbb{R}^d)}{(nv_n)^{-1/2}Z_n(\mathbb{R}^d) + 1}.
\end{aligned}$$

Here we have used $E[T_{n,\mathbb{R}^d}] = nv_n$, which follows from (5.2.2). The equality holds uniformly over any finite family of sets $\tilde{\mathcal{A}} \subset \mathcal{A}$. From Proposition 5.2.5 we know that $Z_n(A))_{A\in\tilde{\mathcal{A}}} \to (Z(A))_{A\in\tilde{\mathcal{A}}}$ weakly, and by assumption it holds $(nv_n)^{-1/2} \to 0$. Hence, all in all we achieve the weak convergence

$$\begin{aligned}
\left(\sqrt{nv_n}\left(\hat{p}_{n,A} - E\left[g_A(W_{n,0}) \mid \|X_0\| > u_n\right]\right)\right)_{A\in\tilde{\mathcal{A}}}
&\to \left(\frac{Z(A) - P(\Theta_i \in A)Z(\mathbb{R}^d)}{0+1}\right)_{A\in\tilde{\mathcal{A}}}\\
&= \left(Z(A) - P(\Theta_i \in A)Z(\mathbb{R}^d)\right)_{A\in\tilde{\mathcal{A}}} =: (Z^{pb}(A))_{A\in\tilde{\mathcal{A}}},
\end{aligned}$$

where we used $E\left[g_A(W_{n,0}) \mid \|X_0\| > u_n\right] \to P(\Theta_i \in A)$, which holds due to Proposition 5.2.1. This proves the asserted weak convergence of fidis. If in addition the bias condition $E\left[g_A(W_{n,0}) \mid \|X_0\| > u_n\right] - P(\Theta_i \in A) = \mathrm{o}((nv_n)^{-1/2})$ holds, then the asserted weak convergence holds obvious by the previous result and

$$\begin{aligned}
\sqrt{nv_n}\left(\hat{p}_{n,A} - P(\Theta_i \in A)\right) =& \sqrt{nv_n}\left(\hat{p}_{n,A} - E\left[g_A(W_{n,0}) \mid \|X_0\| > u_n\right]\right)\\
&+ \sqrt{nv_n}\left(E\left[g_A(W_{n,0}) \mid \|X_0\| > u_n\right] - P(\Theta_i \in A)\right). \qquad \square
\end{aligned}$$

5.7.3 Proofs for Section 5.2.2

In this section, we prove the process convergence of $(Z_n(A))_{A\in\mathcal{A}}$.

Proof of Proposition 5.2.7. We are going to apply Theorem 3.2.1 part (b) with condition set (ii). Condition (D0) follows immediately from the separability assumed in (PA) (iii). All the other conditions used for the fidis convergence (namely (A), (A2), (MX), equation (3.2.4) and (C)) have already been verified in the proof of Proposition 5.2.5.
Thus, by Theorem 3.2.1 part (b) with condition set (ii) it suffices to show conditions (D1) and (D2) for some semi-metric on $\mathcal{G} = \{g_A \mid A \in \mathcal{A}\}$ such that $\mathcal{G}$ it totally bounded.

We start by showing the totally boundedness of $\mathcal{G}$ w.r.t. the $\tilde{\rho}(g_A, g_B) := \rho(A, B)$, where the semi-metric ρ on $\mathcal{A} \cup \{\emptyset\}$ is given by

$$\rho(A,B) = \sum_{j\in\mathbb{Z}} E\Bigg[\mathbb{1}_{\{\|Y_j\|>1\}} \sum_{h\in\mathbb{Z}} \frac{\|Y_h\|^\alpha}{\|Y\|_\alpha^\alpha}\Bigg|\mathbb{1}_A\Big(\frac{Y_{h+i}}{\|Y_h\|}\Big) - \mathbb{1}_B\Big(\frac{Y_{h+i}}{\|Y_h\|}\Big)\Bigg|\Bigg].$$

This is a semi-metric, since $\rho(A, A) = 0$, $\rho(A, B) \geq 0$ and $\rho(A, B) = \rho(B, A)$ for all $A, B \in \mathcal{A}$. Moreover, for $A, B, C \in \mathcal{A}$ we have

$$\begin{aligned}
\rho(A,C) &= \sum_{j\in\mathbb{Z}} E\left[\mathbb{1}_{\{\|Y_j\|>1\}} \sum_{h\in\mathbb{Z}} \frac{\|Y_h\|^\alpha}{\|Y\|_\alpha^\alpha}\left|\mathbb{1}_A\left(\frac{Y_{h+i}}{\|Y_h\|}\right) - \mathbb{1}_C\left(\frac{Y_{h+i}}{\|Y_h\|}\right)\right|\right] \\
&\leq \sum_{j\in\mathbb{Z}}\left(E\left[\mathbb{1}_{\{\|Y_j\|>1\}} \sum_{h\in\mathbb{Z}} \frac{\|Y_h\|^\alpha}{\|Y\|_\alpha^\alpha}\left|\mathbb{1}_A\left(\frac{Y_{h+i}}{\|Y_h\|}\right) - \mathbb{1}_B\left(\frac{Y_{h+i}}{\|Y_h\|}\right)\right|\right]\right. \\
&\qquad \left. + E\left[\mathbb{1}_{\{\|Y_j\|>1\}} \sum_{h\in\mathbb{Z}} \frac{\|Y_h\|^\alpha}{\|Y\|_\alpha^\alpha}\left|\mathbb{1}_B\left(\frac{Y_{h+i}}{\|Y_h\|}\right) - \mathbb{1}_C\left(\frac{Y_{h+i}}{\|Y_h\|}\right)\right|\right]\right) \\
&= \rho(A,B) + \rho(A,C).
\end{aligned}$$

Thus, ρ fulfills all properties of a semi-metric. In the next step, we prove that $\mathcal{A}$ is totally bounded with respect to ρ, which is part of condition (D1). From the definition of ρ it follows that $\rho(A, B) = \rho(A\backslash B, \emptyset)$ for $B \subset A$ and $\rho(\bigcup_{n\in\mathbb{N}} B_n, \emptyset) = \sum_{n\in\mathbb{N}} \rho(B_n, \emptyset)$ for disjoint sets $B_n \subset \mathbb{R}^d$, $n \in \mathbb{N}$. In addition, one has $\rho(\mathbb{R}^d, \emptyset) = c(\mathbb{R}^d, \mathbb{R}^d) < \infty$, as shown in Lemma 5.2.4.

Fix $\delta > 0$ and for the beginning $k \in \{1, ..., q\}$. Recall that, for $t \in [0, 1 + \iota]$, $t^{(k)} \in [0, 1]^q$ denotes the vector with k-th coordinate equal to $t \wedge 1$ and all other coordinates equal to 1.

Due to condition (PA) (ii) the mapping $H_k : t \mapsto \rho(A_{t^{(k)}}, \emptyset)$ is non-decreasing. With exactly the same arguments as in the proof of Lemma 5.2.3 condition (PA) (v) implies

$$P\Big(\exists h \in \mathbb{Z} : Y_{h+i}/\|Y_h\| \in \bigcap_{s\in(t,1]} A_{s^{(k)}} \setminus A_{t^{(k)}}\Big) = 0$$

for all $t \in [0, 1)$. By the monotone convergence theorem, this implies

$$\begin{aligned}
\lim_{s\downarrow t} H_k(s) &= \lim_{s\downarrow t} \rho(A_{s^{(k)}}, \emptyset) = \rho(A_{t^{(k)}}, \emptyset) + \lim_{s\downarrow t} \rho(A_{s^{(k)}} \setminus A_{t^{(k)}}, \emptyset) \\
&= \rho(A_{t^{(k)}}, \emptyset) + \rho\Big(\bigcap_{s\in(t,1]} A_{s^{(k)}} \setminus A_{t^{(k)}}, \emptyset\Big) \\
&= H_k(t) + \sum_{j\in\mathbb{Z}} E\Bigg[\mathbb{1}_{\{\|Y_j\|>1\}} \sum_{h\in\mathbb{Z}} \frac{\|Y_h\|^\alpha}{\|Y\|_\alpha^\alpha} \mathbb{1}_{\bigcap_{s\in(t,1]} A_{s^{(k)}} \setminus A_{t^{(k)}}}\Big(\frac{Y_{h+i}}{\|Y_h\|}\Big)\Bigg] \\
&= H_k(t), \qquad (5.7.11)
\end{aligned}$$

i.e. H_k is right-continuous and thus it is the measure generating function of some measure on $[0,1]$, which is finite because of $\rho(A_1,\emptyset)=\rho(\mathbb{R}^d,\emptyset)=c(\mathbb{R}^d,\mathbb{R}^d)<\infty$. Thus, there exist $J_k<\infty$ such that we can define a partition $\mathcal{T}_k := \Big\{[s_{k,j-1},s_{k,j}) \mid 1\le j\le J_k\Big\}\cup\{\{1\}\}$ of $[0,1]$, with $s_{k,0}=0$ and $s_{k,J_k}=1$ such that $H_k(s_{k,j}-)-H_k(s_{k,j-1})\le\delta$ for all $1\le j\le J_k$. (with $H_k(s-):=\lim_{t\downarrow s}H_k(t)$). Roughly speaking, the idea of this construction and choice of the $s_{k,j}$ is to split up the mass of $\rho(\mathbb{R}^d,\emptyset)$ onto a finite number of disjoint sets given by $\bigcup_{t<s_{k,j+1}} A_{t^{(k)}}\backslash A_{s^{(k)}_{k,j}}$ with mass smaller than δ w.r.t. ρ. One way to define this partition is to define the interval boundaries iteratively by

$$s_{k,j}=\inf\Big\{t\in(s_{k,j-1},1] \mid \rho\Big(A_{t^{(k)}},A_{s^{(k)}_{k,j-1}}\Big)>\delta\Big\}.$$

Observe that although it is possible that the measure pertaining to H_k has mass greater than δ at some of the $s_{k,j}$, such jumps of H_k do not play any role in the following calculations, in particular no continuity is required. If H_k has some point mass greater than δ, then this has to be at some point $s_{k,j}$.
For the fixed $\delta>0$ we then define the finite cover of $[0,1]^q$ by the sets

$$\mathcal{T}^{(\delta)} := \{\times_{k=1}^q T_k \mid T_k\in\mathcal{T}_k,\ \forall\, 1\le k\le q\}.$$

Since $\mathcal{T}_k$ is a partition of $[0,1]$, this $\mathcal{T}^{(\delta)}$ directly defines a partition of $[0,1]^q$. This partition contains $\prod_{k=1}^q J_k<\infty$ sets, in particular there are only finite many sets in this cover of $[0,1]^q$.
For any set $T=\times_{k=1}^q T_k\in\mathcal{T}^{(\delta)}$ define $K_T:=\{1\le k\le q \mid T_k\ne\{1\}\}$, i.e. as the set of all indexes k with $T_k\ne\{1\}$. Moreover, we define the smallest set and the upper bound for all sets in one family $\{A_t|t\in T\}$ of the above defined partition by

$$\begin{aligned}\underline{A}_T &:= \bigcap_{s\in T} A_s = A_{(\min T_1,\dots,\min T_q)}\in\mathcal{A} \quad \text{and} \\ \bar{A}_T &:= \bigcup_{s\in T} A_s \in \tilde{\mathcal{A}} := \{A_t^- \mid t\in[0,1+\iota]^q\}.\end{aligned} \tag{5.7.12}$$

Note that $\bar{A}_T\backslash\underline{A}_T\subset\bigcup_{k\in K_T}\Big(\bigcup_{s\in T_k}A_{s^{(k)}}\setminus A_{(\min T_k)^{(k)}}\Big)$. Hence, by the construction of $\mathcal{T}^{(\delta)}$, for all $t\in T$,

$$\begin{aligned}\rho\Big(A_t,\underline{A}_T\Big) &= \rho\Big(A_t\setminus\underline{A}_T,\emptyset\Big)\le\rho\Big(\bar{A}_T\setminus\underline{A}_T,\emptyset\Big) \\ &\le \sum_{k\in K_T}\rho\Big(\bigcup_{s\in T_k}A_{s^{(k)}}\setminus A_{(\min T_k)^{(k)}},\emptyset\Big) \\ &= \sum_{k\in K_T}H_k(\sup T_k-)-H_k(\min T_k)\le q\delta,\end{aligned} \tag{5.7.13}$$

by applying $\rho(\bigcup_{n\in\mathbb{N}}B_n,\emptyset)=\sum_{n\in\mathbb{N}}\rho(B_n,\emptyset)$ for disjoint sets $B_n\subset\mathbb{R}^d$, $n\in\mathbb{N}$.
Therefore, all sets of the form $\{A_t|t\in T\}$, $T\in\mathcal{T}^{(\delta)}$, have a radius of at most $q\cdot\delta$ w.r.t. ρ. Thus, $\mathcal{A}$ is totally bounded with respect to ρ and this implies that $\mathcal{G}$ is totally bounded

with respect to $\tilde{\rho}$.
Now, we turn to the continuity condition (3.1.5) of (D1). Define the semi-metric ρ as before and define $V_n(g_A) = \sum_{j=1}^{r_n} g(W_{n,j})$. Observe that

$$\begin{aligned}
&\frac{m_n}{p_n b_n(g_A) b_n(g_B)} E\left[(V_n(g_A) - V_n(g_B))^2\right] \\
&= \frac{m_n}{n v_n} E\left[\left(\sum_{t=1}^{r_n} g_A(W_{n,t}) - g_B(W_{n,t})\right)^2\right] \\
&= \frac{m_n}{n v_n} E\left[\left(\sum_{t=1}^{r_n} g_A(W_{n,t})\right)^2\right] + \frac{m_n}{n v_n} E\left[\left(\sum_{t=1}^{r_n} g_B(W_{n,t})\right)^2\right] \\
&\qquad - 2\frac{m_n}{n v_n} E\left[\sum_{t=1}^{r_n} g_A(W_{n,t}) \sum_{j=1}^{r_n} g_B(W_{n,j})\right] \\
&\to \sum_{j\in\mathbb{Z}} E\left[\mathbb{1}_{\{\|Y_j\|>1\}} \left(\sum_{l\in\mathbb{Z}} \frac{\|Y_l\|^\alpha}{\|Y\|_\alpha^\alpha} \mathbb{1}_A\left(\frac{Y_{l+i}}{\|Y_l\|}\right)\right)^2\right] \\
&\quad + \sum_{j\in\mathbb{Z}} E\left[\mathbb{1}_{\{\|Y_j\|>1\}} \left(\sum_{l\in\mathbb{Z}} \frac{\|Y_l\|^\alpha}{\|Y\|_\alpha^\alpha} \mathbb{1}_B\left(\frac{Y_{l+i}}{\|Y_l\|}\right)\right)^2\right] \\
&\quad - 2\sum_{j\in\mathbb{Z}} E\left[\mathbb{1}_{\{\|Y_j\|>1\}} \left(\sum_{l\in\mathbb{Z}} \frac{\|Y_l\|^\alpha}{\|Y\|_\alpha^\alpha} \mathbb{1}_A\left(\frac{Y_{l+i}}{\|Y_l\|}\right)\right)\left(\sum_{k\in\mathbb{Z}} \frac{\|Y_k\|^\alpha}{\|Y\|_\alpha^\alpha} \mathbb{1}_B\left(\frac{Y_{k+i}}{\|Y_k\|}\right)\right)\right] \\
&= \sum_{j\in\mathbb{Z}} E\left[\mathbb{1}_{\{\|Y_j\|>1\}} \left(\sum_{l\in\mathbb{Z}} \frac{\|Y_l\|^\alpha}{\|Y\|_\alpha^\alpha} \left(\mathbb{1}_A\left(\frac{Y_{l+i}}{\|Y_l\|}\right) - \mathbb{1}_B\left(\frac{Y_{l+i}}{\|Y_l\|}\right)\right)\right)^2\right] \\
&\le \sum_{j\in\mathbb{Z}} E\left[\mathbb{1}_{\{\|Y_j\|>1\}} \sum_{l\in\mathbb{Z}} \frac{\|Y_l\|^\alpha}{\|Y\|_\alpha^\alpha} \left|\mathbb{1}_A\left(\frac{Y_{l+i}}{\|Y_l\|}\right) - \mathbb{1}_B\left(\frac{Y_{l+i}}{\|Y_l\|}\right)\right|\right] \\
&= \rho(A,B).
\end{aligned}$$

The convergence holds pointwise for all $A, B \in \mathcal{A}$ due to Lemma 5.2.4, in particular due to (5.7.7) and (5.7.8). Note also that $m_n r_n v_n/(n v_n) \to 1$. By condition (PA) (iv) and the same arguments as in Lemma 5.2.4, this convergence holds also for $A, B \in \tilde{\mathcal{A}} = \{A_t^- \mid t \in [0, 1+\iota]^q\}$.
For $T, S \in \mathcal{T}^{(\delta)}$ and $\bar{A}_T \in \tilde{\mathcal{A}}$ and $\underline{A}_S \in \mathcal{A}$ it follows for sufficiently large $n \in \mathbb{N}$ that

$$\frac{1}{r_n v_n} E\left[\left(V_n(g_{\bar{A}_T}) - V_n(g_{\underline{A}_S})\right)^2\right] \le \rho(\bar{A}_T, \underline{A}_S) + \frac{\varepsilon}{2}. \tag{5.7.14}$$

This inequality holds uniformly for all $T, S \in \mathcal{T}^{(\delta)}$, since $\mathcal{T}^{(\delta)}$ contains only finite many sets.
Now consider $A_t, A_s \in \mathcal{A}$ with $\rho(A_t, A_s) \le \delta$. Since by construction $\mathcal{T}^{(\delta)}$ is a partition of $[0,1]^q$, there exist unique $S, T \in \mathcal{T}^{(\delta)}$ such that $s \in S$ and $t \in T$. By definition it holds $\underline{A}_T \subset A_t \subset \bar{A}_T$ and $\underline{A}_S \subset A_s \subset \bar{A}_S$. Hence, by the choice of A_t, A_s, the construction of the partition, the convergence leading to (5.7.14) and by the inequality (5.7.13) we may

conclude for sufficiently large $n \in \mathbb{N}$

$$\begin{aligned}
&\frac{m_n}{nv_n} E\left[(V_n(g_{A_t}) - V_n(g_{A_s}))^2\right] \\
&\quad \le \frac{m_n}{nv_n} E\left[\max\left((V_n(g_{\bar{A}_T}) - V_n(g_{\underline{A}_S}))^2, (V_n(g_{\underline{A}_T}) - V_n(g_{\bar{A}_S}))^2\right)\right] \\
&\quad \le \frac{m_n}{nv_n} E\left[(V_n(g_{\bar{A}_T}) - V_n(g_{\underline{A}_S}))^2\right] + \frac{m_n}{nv_n} E\left[(V_n(g_{\underline{A}_T}) - V_n(g_{\bar{A}_S}))^2\right] \\
&\quad \le \rho(\bar{A}_T, \underline{A}_S) + \rho(\underline{A}_T, \bar{A}_S) + \varepsilon \\
&\quad \le \rho(\bar{A}_T, A_t) + \rho(A_t, A_s) + \rho(A_s, \underline{A}_S) + \rho(\underline{A}_T, A_t) + \rho(A_t, A_s) + \rho(A_s, \bar{A}_S) + \varepsilon \\
&\quad \le 2\Big(\rho(\bar{A}_T, \underline{A}_T) + \rho(\bar{A}_S, \underline{A}_S) + \rho(A_t, A_s)\Big) + \varepsilon \\
&\quad \le (4q+2)\delta + \varepsilon.
\end{aligned}$$

This bound holds uniformly for all $s, t \in [0,1]^q$ with $\rho(A_t, A_s) \le \delta$. From this uniform bound it follows $m_n/(nv_n)E\left[(V_n(g_A) - V_n(g_B))^2\right] < (4q+2)\delta + \varepsilon_n$ for all $A, B \in \mathcal{A}$ with $\rho(A, B) < \delta$ and for some $\varepsilon_n > 0$ independent of A, B and with $\varepsilon_n \to 0$ as $n \to \infty$. Hence, $\sup_{A,B\in\mathcal{A}:\rho(A,B)<\delta} m_n/(nv_n)E\left[(V_n(g_A) - V_n(g_B))^2\right] \le \delta + \varepsilon_n$ follows for a sufficiently large $n \in \mathbb{N}$. Thus,

$$\limsup_{n\to\infty} \sup_{A,B\in\mathcal{A}:\rho(A,B)<\delta} \frac{m_n}{nv_n} E\left[(V_n(g_A) - V_n(g_B))^2\right] \le \delta$$

and

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \sup_{A,B\in\mathcal{A}:\rho(A,B)<\delta} \frac{m_n}{nv_n} E\left[(V_n(g_A) - V_n(g_B))^2\right] = 0$$

and condition (D1) is satisfied.
Finally we turn to condition (D2). As shorthand define

$$N_n := \frac{1}{r_n v_n} E\left[\Big(\sum_{j=1}^{r_n} \mathbb{1}_{\{\|X_j\| > u_n\}}\Big)^2\right].$$

It obviously holds $N_n \le r_n/v_n < \infty$ for all $n \in \mathbb{N}$.
Denote by $w = (w_k)_{1\le k\le q} \in [0,1]^q$ an index with $0 \in A_w$ but $0 \notin A_s$ for all $s < w$. This index w exists due to condition (PA) (vii). Condition (PA) (ii) implies that $t \mapsto A_t$ is non-decreasing in each coordinate (to this end, choose $s_i = 0$ in the non-decreasing functions in condition (PA) (ii)) and therefore $0 \in A_s$ for all $s = (s_1, ..., s_q)$ with $s \ge w$. Moreover, this is an equivalence, i.e condition (PA) (ii) implies $0 \notin A_s$ for all $s \not\ge w$, i.e. $s_i < w_i$ for some $i \in \{1, ..., q\}$. To this end, note that $0 \in A_w$ and $0 \notin A_{s\wedge w}$ with $s \wedge w$ is understood componentwise, i.e. $(s \wedge w)_i := \min(s_i, w_i)$. Likewise $s \vee w = (\max(s_i, w_i))_{1\le i\le q}$ is defined as the componentwise maximum. If $0 \in A_s$, then it would be $0 \in A_s \backslash A_{s\wedge w}$ but $0 \notin A_{s\vee w} \backslash A_w$. This contradicts condition (PA) (ii) since this condition implies $A_s \backslash A_{s\wedge w} \subset A_{s\vee w} \backslash A_w$. This can be seen by increasing successively s_j to w_j for all

$j = 1, ..., q$ with $s_j < w_j$ under consideration of the non-decreasing functions in condition (PA) (ii). Thus, $0 \notin A_s$. All in all, condition (PA) (ii) implies $0 \in A_s$ if and only if $s \geq w$. This will be important for the construction of the brackets below.

In a first step for the construction of brackets now fix $n \in \mathbb{N}$ and let $\varepsilon \in (0,1)$ be arbitrary. With exactly the same arguments as above for the functions H_k, we may conclude from condition (PA) (ii) and (vi) that for all $1 \leq k \leq q$ the function $F_k : [0,1] \to \mathbb{R}$,

$$t \mapsto \frac{1}{r_n v_n} E\left[\left(\sum_{t=1}^{r_n} \mathbb{1}_{\{\|X_t\| > u_n\}} \sum_{h \in H_{n,i}} \frac{\|X_{t+h}\|^\alpha}{\sum_{k=-s_n}^{s_n} \|X_{t+k}\|^\alpha} \mathbb{1}_{A_{t(k)}}\left(\frac{X_{t+h+i}}{\|X_{t+h}\|}\right)\right)^2\right]$$

is right-continuous and, therefore, the measure-generating function of a measure on $[0,1]$ with finite mass $F_k(1) \leq N_n < \infty$.

Hence, one can choose $t_{k,j} \in [0,1]$, $0 \leq j \leq J_k^* \leq \lceil N_n q^2/\varepsilon \rceil + 1$, such that $t_{k,0} := 0$, $t_{k,J_k^*} = 1$, $w_k \in \{t_{k,j} | 1 \leq j \leq J_k^*\}$, $t_{k,j-1} < t_{k,j}$ and $F_k(t_{k,j}-) - F_k(t_{k,j-1}) \leq \varepsilon/q^2$ for all $1 \leq j \leq J_k^*$. Roughly speaking, the idea is to choose an increasing sequence of J_k^* points such that they split up the mass of the measure pertaining to F_k into finite many disjoint sets $[t_{k,j}, t_{k,j-1})$ with mass less than ε/q^2. Again, one can chose this points $t_{k,j}$ for instance iteratively by

$$t_{k,j} = \inf\left\{t \in (t_{k,j-1}, 1] \mid F_k(t) - F_k(t_{k,j-1}) > \varepsilon/q^2\right\}.$$

These points define a partition $\mathcal{T}_k^* := \{[t_{k,j-1}, t_{k,j}) | 1 \leq j \leq J_k^*\} \cup \{\{1\}\}$ of $[0,1]$ and hence $\mathcal{T}^* := \{\times_{k=1}^q T_k \mid T_k \in \mathcal{T}_k^*, \forall\, 1 \leq k \leq q\}$ is a partition of $[0,1]^q$. Note that the construction of this partition $\mathcal{T}^*$ works completely analogously to the construction of $\mathcal{T}$ above, the only difference now is that the size of the sets in the partition is now bounded with respect to the pre-asymptotic function F_k instead of the asymptotic versions H_k. This also explains why we need condition (PA) (vi) instead of (PA) (v) here. From this partition we define brackets by

$$\mathcal{A}_T^{\varepsilon,n} := \{A_t \mid t \in T\}$$

for all $T \in \mathcal{T}^*$, so that $\mathcal{A}^{\varepsilon,n} = \{\mathcal{A}_T^{\varepsilon,n} \mid T \in \mathcal{T}^*\}$ forms a partition of $\mathcal{A}$. By the previous discussion, $0 \in A_s$ if and only if $s \geq w$. Thus, according to the construction of the partition and the definition of A_w for each set $\mathcal{S} \in \mathcal{A}^{\varepsilon,n}$ it is either $0 \in A$ for all $A \in \mathcal{S}$ or $0 \notin A$ for all $A \in \mathcal{S}$. In particular, this implies $\mathbb{1}_A(0) = \mathbb{1}_B(0)$ for all $A, B \in \mathcal{S}$ and therefore $\sup_{A,B \in \mathcal{S}} \mathbb{1}_{A \setminus B}(0) = 0$. This is why the indicator $\mathbb{1}_A(0)$ which occurs in the definition of $g_A(W_{n,t})$ does not occur in the following calculations.

In the subsequent calculations we will apply $(a-b)^2 = a^2 - 2ab + b^2 \leq a^2 - 2b^2 + b^2 \leq a^2 - b^2$, which hold for all $a \geq b$. Due to the construction of the partition it holds for $\mathcal{A}_T^{\varepsilon,n}$ with $T = \times_{k=1}^q T_k \in \mathcal{T}^*$ and $A_{T_k}^+ := \bigcup_{s \in T_k} A_{s^{(k)}}$ that

$$\frac{1}{r_n v_n} E\left[\sup_{A,B \in \mathcal{A}_T^{\varepsilon,n}} \left(\sum_{t=1}^{r_n} (g_A(W_{n,t}) - g_B(W_{n,t}))\right)^2\right]$$

$$
\begin{aligned}
&\le \frac{1}{r_n v_n} E\left[\left(\sum_{t=1}^{r_n}(g_{\bar{A}_T}(W_{n,t}) - g_{\underline{A}_T}(W_{n,t}))\right)^2\right] = \frac{1}{r_n v_n} E\left[\left(\sum_{t=1}^{r_n} g_{\bar{A}_T\setminus\underline{A}_T}(W_{n,t})\right)^2\right] \\
&= \frac{1}{r_n v_n} E\left[\left(\sum_{t=1}^{r_n} \mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h=-s_n}^{s_n} \frac{\|X_{t+h}\|^\alpha}{\sum_{k=-s_n}^{s_n}\|X_{t+k}\|^\alpha}\left(\mathbb{1}_{\{h\in H_{n,i}\}}\mathbb{1}_{\bar{A}_T\setminus\underline{A}_T}\left(\frac{X_{t+h+i}}{\|X_{t+h}\|}\right)\right.\right.\right. \\
&\qquad\qquad\qquad \left.\left.\left. + \mathbb{1}_{\{h\in H_{n,i}^C\}}\mathbb{1}_{\bar{A}_T\setminus\underline{A}_T}(0)\right)\right)^2\right] \\
&= \frac{1}{r_n v_n} E\left[\left(\sum_{t=1}^{r_n} \mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h\in H_{n,i}} \frac{\|X_{t+h}\|^\alpha}{\sum_{k=-s_n}^{s_n}\|X_{t+k}\|^\alpha}\mathbb{1}_{\bar{A}_T\setminus\underline{A}_T}\left(\frac{X_{t+h+i}}{\|X_{t+h}\|}\right)\right)^2\right] \\
&\le \frac{1}{r_n v_n} E\left[\left(\sum_{k\in K_T}\sum_{t=1}^{r_n} \mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h\in H_{n,i}} \frac{\|X_{t+h}\|^\alpha}{\sum_{k=-s_n}^{s_n}\|X_{t+k}\|^\alpha}\mathbb{1}_{A^+_{T_k}\setminus A_{(\min T_k)^{(k)}}}\left(\frac{X_{t+h+i}}{\|X_{t+h}\|}\right)\right)^2\right] \\
&\le \frac{|K_T|}{r_n v_n}\sum_{k\in K_T} E\left[\left(\sum_{t=1}^{r_n} \mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h\in H_{n,i}} \frac{\|X_{t+h}\|^\alpha}{\sum_{k=-s_n}^{s_n}\|X_{t+k}\|^\alpha}\mathbb{1}_{A^+_{T_k}\setminus A_{(\min T_k)^{(k)}}}\left(\frac{X_{t+h+i}}{\|X_{t+h}\|}\right)\right)^2\right] \\
&= \frac{|K_T|}{r_n v_n}\sum_{k\in K_T} E\left[\left(\sum_{t=1}^{r_n} \mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h\in H_{n,i}} \frac{\|X_{t+h}\|^\alpha}{\sum_{k=-s_n}^{s_n}\|X_{t+k}\|^\alpha}\left(\mathbb{1}_{A^+_{T_k}}\left(\frac{X_{t+h+i}}{\|X_{t+h}\|}\right)\right.\right.\right. \\
&\qquad\qquad\qquad \left.\left.\left. - \mathbb{1}_{A_{(\min T_k)^{(k)}}}\left(\frac{X_{t+h+i}}{\|X_{t+h}\|}\right)\right)\right)^2\right] \\
&\le |K_T|\sum_{k\in K_T}\frac{1}{r_n v_n} E\left[\left(\sum_{t=1}^{r_n} \mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h\in H_{n,i}} \frac{\|X_{t+h}\|^\alpha}{\sum_{k=-s_n}^{s_n}\|X_{t+k}\|^\alpha}\mathbb{1}_{A^+_{T_k}}\left(\frac{X_{t+h+i}}{\|X_{t+h}\|}\right)\right)^2\right] \\
&\quad - K_T\sum_{k\in K_T}\frac{1}{r_n v_n} E\left[\left(\sum_{t=1}^{r_n} \mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h\in H_{n,i}} \frac{\|X_{t+h}\|^\alpha}{\sum_{k=-s_n}^{s_n}\|X_{t+k}\|^\alpha}\mathbb{1}_{A_{(\min T_k)^{(k)}}}\left(\frac{X_{t+h+i}}{\|X_{t+h}\|}\right)\right)^2\right] \\
&= |K_T|\sum_{k\in K_T}(F_k(\sup T_k-) - F_k(\min T_k)) \\
&\le |K_T|\sum_{k\in K_T}\frac{\varepsilon}{q^2} \le \varepsilon,
\end{aligned}
$$

where the third last step holds by construction and the penultimate step holds since $|K_T| \le q$ by definition. Note that for $k \in \{1, ...q\} \setminus K_T$ one has $A^+_{T_k} \setminus A_{1^{(k)}} = \emptyset$, which is why the corresponding summands do not occur in the above calculation.

Thus, each set $\mathcal{A}_T^{\varepsilon,n}$ in the partition is indeed a $\sqrt{\varepsilon}$-bracket for $\mathcal{A}$ with respect to the L_2^n metric as considered in condition (D2).

The partition $\mathcal{A}^{\varepsilon,n}$ contains $\prod_{k=1}^q (J_k^* + 1)$ sets. This is an upper bound for the bracketing number $N_{[\cdot]}(\sqrt{\varepsilon}, \mathcal{A}, L_2^n)$ which is defined in condition (D2) for all $n \in \mathbb{N}$ and $\varepsilon > 0$. Thus, we obtain $N_{[\cdot]}(\sqrt{\varepsilon}, \mathcal{A}, L_2^n) \le \prod_{k=1}^q (J_k^* + 1) \le (\lceil N_n q^2/\varepsilon\rceil + 2)^q \le (N_n q^2/\varepsilon + 3)^q$. This implies for $\varepsilon < 1$

$$
\begin{aligned}
\log(N_{[\cdot]}(\varepsilon, \mathcal{A}, L_2^n)) &\le q\log\left(N_n\frac{q^2}{\varepsilon^2} + 3\right) = q\log\left(N_n\frac{q^2}{\varepsilon^2}\left(1 + 3\frac{\varepsilon^2}{N_n q^2}\right)\right) \le q\log\left(N_n\frac{q^2}{\varepsilon^2}4\right) \\
&= q\log(N_n) - q\log(\varepsilon) + 2q\log(q^2) + q\log(4) =: q\log(N_n) - 2q\log(\varepsilon) + c_1,
\end{aligned}
$$

since we have $N_n \geq 1$ by definition. Due to the verification of (5.7.10) in the proof of Proposition 5.2.5 we have $N_n = \mathrm{O}(1)$ for $n \to \infty$, in particular we have $\log(N_n) \leq c_2$ for some suitable constant $c_2 > 0$ and for all sufficiently large n.
Therefore, it follows for sufficiently large n and $\tau < 1$

$$\int_0^\tau \sqrt{\log(N_{[\cdot]}(\varepsilon, \mathcal{A}, L_2^n))} d\varepsilon \leq \int_0^\tau \sqrt{q \log(N_n) - 2q \log(\varepsilon) + c_1} d\varepsilon$$
$$\leq \int_0^\tau q \log(N_n) - 2q \log(\varepsilon) + c_1 d\varepsilon = c_1\tau + \tau q \log(N_n) - 2q \left[\varepsilon \log(\varepsilon) - \varepsilon\right]_{\varepsilon=0}^\tau$$
$$= c_1\tau + \tau q \log(N_n) - 2q\tau \log(\tau) + 2q\tau \leq \tau(c_1 + qc_2 + 2q) - 2q\tau \log(\tau).$$

This tends to 0 as $\tau \to 0$. Thus, condition (D2) is satisfied with the n-dependent partitions $\mathcal{A}^{\varepsilon,n}$ defined above.
All in all, the conditions of Theorem 3.2.1 part (b) are fulfilled which yields the asserted process convergence of $(Z_n(A))_{A\in\mathcal{A}}$ to centered Gaussian process $(Z(A))_{A\in\mathcal{A}}$ with covariance function c as given in Lemma 5.2.4. □

The next proof considers the special case of linearly ordered sets in $\mathcal{A}$.

Proof of Corollary 5.2.8. First note, that (PA) (ii) is trivially fulfilled due to the linear order of $\mathcal{A}$. Part (i) of (PA) can be fulfilled if $\emptyset, \mathbb{R}^d \in \mathcal{A}$, otherwise one can add this two sets. The proof of the assertion is essentially the same as the proof of Proposition 5.2.7. We want to apply Theorem 3.2.1, but now we want to verify the conditions for asymptotic equicontinuity in condition set (i) in part (b) of the theorem, i.e. condition (D1) and (D3) (instead of part (ii) with condition (D2)). Condition (D1) was established in the proof of Proposition 5.2.7.
Next we consider condition (D3). Since $\mathcal{A}$ is linearly ordered, the functions in $\mathcal{G}$ are linearly ordered. Therefore, $\mathcal{G} = \{g_A | A \in \mathcal{A}\}$ is a $VC(2)$-class. This is enough to show the entropy condition (D3) (cf. remark directly after condition (D3) or Remark 2.11 in Drees and Rootzén (2010) or Van der Vaart and Wellner (1996), Section 2.6). Thus, the assertion follows from Theorem 3.2.1.
Note that the conditions (PA) (vi) and (vii) were only used in the proof of (D2) and are hence not needed in the case $q = 1$ considered here. □

As before, for the proof of Theorem 5.2.9 the asymptotics of $(T_{n,A})_{A\in\mathcal{A}}$ is the main ingredient and we can derive our asymptotics for $(\hat{p}_{n,A})_{A\in\mathcal{A}}$ from Proposition 5.2.7.

Proof of Theorem 5.2.9. We have $\hat{p}_{n,A} = T_{n,A}/T_{n,\mathbb{R}^d}$ and we have already established the asymptotic behavior of $T_{n,A}$. Direct calculations yields

$$\sqrt{nv_n}\,(\hat{p}_{n,A} - p_A) = \sqrt{nv_n}\Big(\frac{T_{n,A}}{T_{n,\mathbb{R}^d}} - p_A\Big)$$
$$= \sqrt{nv_n}\frac{\sqrt{nv_n}Z_n(A) + E[T_{n,A}] - p_A\sqrt{nv_n}Z_n(\mathbb{R}^d) - p_A nv_n}{\sqrt{nv_n}Z_n(\mathbb{R}^d) + nv_n}$$

$$= \frac{Z_n(A) - p_A Z_n(\mathbb{R}^d) + \sqrt{nv_n}\,((nv_n)^{-1}E[T_{n,A}] - p_A)}{(nv_n)^{-1/2} Z_n(\mathbb{R}^d) + 1}.$$

Here we applied $E[T_{n,\mathbb{R}^d}] = nE[\mathbb{1}_{\{\|X_0\|>u_n\}}] = nv_n$. From Proposition 5.2.7 we know that $(Z_n(A))_{A\in\mathcal{A}} \to (Z(A))_{A\in\mathcal{A}}$, and by assumption it holds $(nv_n)^{-1/2} \to 0$. Hence, together with (PB$_T$) we achieve the weak process convergence

$$\begin{aligned}(\sqrt{nv_n}\,(\hat{p}_{n,A} - p_A))_{A\in\mathcal{A}} &\xrightarrow{w} \left(\frac{Z(A) - p_A Z(\mathbb{R}^d) + 0}{0+1}\right)_{A\in\mathcal{A}} \\ &= \Big(Z(A) - p_A Z(\mathbb{R}^d)\Big)_{A\in\mathcal{A}} = (Z^{pb}(A))_{A\in\mathcal{A}}.\end{aligned}$$

The covariance structure of the limit process $(Z^{pb}(A))_{A\in\mathcal{A}}$ follows by direct calculations: With the covariance function c given in Lemma 5.2.4 and using the symmetry of c it follows for $A, B \in \mathcal{A}$

$$\begin{aligned}
&c^{pb}(A,B) = Cov(Z^{pb}(A), Z^{pb}(B)) \\
&= Cov(Z(A) - p_A Z(\mathbb{R}^d), Z(B) - p_B Z(\mathbb{R}^d)) \\
&= c(A,B) + p_A p_B c(\mathbb{R}^d,\mathbb{R}^d) - p_B c(A,\mathbb{R}^d) - p_A c(B,\mathbb{R}^d) \\
&= \sum_{j\in\mathbb{Z}} E\bigg[(\|\Theta_j\|^\alpha \wedge 1)\Big(\sum_{h\in\mathbb{Z}} \frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha}\mathbb{1}_A\Big(\frac{\Theta_{h+i}}{\|\Theta_h\|}\Big)\Big)\Big(\sum_{l\in\mathbb{Z}} \frac{\|\Theta_l\|^\alpha}{\|\Theta\|_\alpha^\alpha}\mathbb{1}_B\Big(\frac{\Theta_{l+i}}{\|\Theta_l\|}\Big)\Big)\bigg] \\
&\qquad + p_A p_B \sum_{j\in\mathbb{Z}} E\left[(\|\Theta_j\|^\alpha \wedge 1)\right] \\
&\qquad - p_B \sum_{j\in\mathbb{Z}} E\bigg[(\|\Theta_j\|^\alpha \wedge 1)\Big(\sum_{l\in\mathbb{Z}} \frac{\|\Theta_l\|^\alpha}{\|\Theta\|_\alpha^\alpha}\mathbb{1}_A\Big(\frac{\Theta_{l+i}}{\|\Theta_l\|}\Big)\Big)\bigg] \\
&\qquad - p_A \sum_{j\in\mathbb{Z}} E\bigg[(\|\Theta_j\|^\alpha \wedge 1)\Big(\sum_{l\in\mathbb{Z}} \frac{\|\Theta_l\|^\alpha}{\|\Theta\|_\alpha^\alpha}\mathbb{1}_B\Big(\frac{\Theta_{l+i}}{\|\Theta_l\|}\Big)\Big)\bigg] \\
&= \sum_{j\in\mathbb{Z}} E\bigg[(\|\Theta_j\|^\alpha \wedge 1)\Big(p_B - \sum_{h\in\mathbb{Z}} \frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha}\mathbb{1}_B\Big(\frac{\Theta_{h+i}}{\|\Theta_h\|}\Big)\Big)\Big(p_A - \sum_{l\in\mathbb{Z}} \frac{\|\Theta_l\|^\alpha}{\|\Theta\|_\alpha^\alpha}\mathbb{1}_A\Big(\frac{\Theta_{l+i}}{\|\Theta_l\|}\Big)\Big)\bigg]. \quad \square
\end{aligned}$$

5.7.4 Proof of Theorem 5.3.1

Similarly to the proof of Theorem 5.2.9, we can prove the asymptotic normality of $\hat{\hat{p}}_{n,A}$. Again, we first prove asymptotic normality of the appearing statistics $T_{n,A}$ and $T_{n,\phi}$ using Theorem 3.2.1 and than apply a continuous mapping argument to derive asymptotic normality of the fraction. However, since the estimator $\hat{\alpha}_n$ appears in an exponent, the analysis here is more sophisticated involving some Taylor expansion arguments. The most complicated part of the proof will be the treatment of these single terms of the Taylor expansion.

Some parts of the proof of Theorem 5.3.1 are arranged in a series of lemmas, which are discussed subsequent to the main proof.

Proof of Theorem 5.3.1. One directly obtains

$$\begin{aligned}
&\left(\sqrt{nv_n}\left(\hat{\hat{p}}_{n,A} - P(\Theta_i \in A)\right)\right)_{A\in\mathcal{A}} \\
&= \left(\sqrt{nv_n}\left(\hat{p}_{n,A} - P(\Theta_i \in A)\right)\right)_{A\in\mathcal{A}} + \left(\sqrt{nv_n}\left(\hat{\hat{p}}_{n,A} - \hat{p}_{n,A}\right)\right)_{A\in\mathcal{A}}. \qquad (5.7.15)
\end{aligned}$$

The first summand converges weakly to $(Z^{pb}(A))_{A\in\mathcal{A}}$ due to Theorem 5.2.9. Thus, we have to check the convergence of the second summand and that the convergence of both summands holds jointly.

First, note that $\hat{\alpha}_n = T_{n,\mathbb{R}^d}/T_{n,\phi}$ with $T_{n,A}$ defined in (5.2.1) and the function ϕ is given by $\phi((x_h)_{h\in\mathbb{Z}}) = \log^+(\|x_0\|)$. With similar arguments as in Proposition 5.2.7 we will show the weak convergence of $((Z_n(A))_{A\in\mathcal{A}}, Z_n(\phi))$ to a centered Gaussian process. For $(Z_n(A))_{A\in\mathcal{A}}$ this is the statement of Proposition 5.2.7, in particular the conditions for fidis convergence were already checked there. The conditions for asymptotic tightness are still fulfilled if only the single function ϕ is added, so asymptotic tightness holds for $\{g_A : A \in \mathcal{A}\} \cup \{\phi\}$. Thus, to prove the process convergence $((Z_n(A))_{A\in\mathcal{A}}, Z_n(\phi))$ it suffices to prove fidis convergence of $Z_n(\phi)$ and the convergence of the covariance of $Z_n(\phi)$ and $Z_n(A)$, since apart from (C) the conditions for fidis convergence can be checked for each function of the index set individually.

For ϕ as an unbounded function the fidis conditions can be verified with Theorem 3.2.3 for unbounded function. Here we use $b_n(\phi) = \sqrt{nv_n/p_n} = b_n(g_A)$. For this theorem we have to check (3.2.7), condition (L) and the convergence of standardized covariance, the remaining conditions were already established in Proposition 5.2.7. To ease the formulas recall the notation $X_{n,t} = X_t/u_n$. By condition (PP1) and stationarity we obtain

$$\begin{aligned}
&\frac{1}{r_n v_n} E\left[\left(\sum_{t=1}^{r_n} \phi(W_{n,t})\right)^2\right] = \frac{1}{r_n v_n} E\left[\left(\sum_{t=1}^{r_n} \log^+(\|X_{n,t}\|)\right)^2\right] \\
&= \sum_{t=1}^{r_n}\sum_{j=1}^{r_n} \frac{1}{r_n v_n} E\left[\log^+(\|X_{n,j}\|)\log^+(\|X_{n,t}\|)\right] \\
&\le 2\sum_{k=0}^{r_n} \frac{(r_n - k)}{r_n} E\left[\log^+(\|X_{n,0}\|)\log^+(\|X_{n,k}\|) \mid \|X_0\| > u_n\right] \\
&\le 2\sum_{k=0}^{r_n} E\left[\log^+(\|X_{n,0}\|)\log^+(\|X_{n,k}\|) \mid \|X_0\| > u_n\right] \\
&\le 2E\left[\log^+(\|X_{n,0}\|)^2 \mid \|X_0\| > u_n\right] + 2\sum_{k=1}^{r_n} e_n'(k) = \mathrm{O}(1). \qquad (5.7.16)
\end{aligned}$$

Thus, condition (3.2.7) is met. Here, $E\left[\log^+(\|X_{n,0}\|)^2 \mid \|X_0\| > u_n\right] = \mathrm{O}(1)$ holds since $E\left[\log^+(\|X_{n,0}\|)^2 \mid \|X_0\| > u_n\right] \to E[\log^+(\|Y_0\|)^2]$, which holds due to the definition of the tail process $(Y_t)_{t\in\mathbb{Z}}$ and the uniform integrability of the family $\log^+(\|X_{n,0}\|)^2/v_n$, $n \in \mathbb{N}$. The latter holds since $\log^+(\|x\|) \le c_p\|x\|^p \mathbb{1}_{\{\|x\|>1\}}$ for some $c_p > 0$ and $p \in (0, \alpha/2)$ and $\|X_{n,0}\|^{2p}\mathbb{1}_{\{\|X_0\|>u_n\}}/v_n, n \in \mathbb{N}$, is uniform integrable due to the Potter bounds (the

argument is given more detailed in (5.5.6)).
Using (5.7.16) and Markov's inequality yields

$$P\Big(\sum_{t=1}^{r_n}\log^+\|X_{n,t}\|>\varepsilon\sqrt{nv_n}\Big)\le\frac{1}{\varepsilon^2 nv_n}E\Big[\Big(\sum_{t=1}^{r_n}\log^+(\|X_{n,t}\|)\Big)^2\Big]=\mathrm{O}\Big(\frac{r_n}{n}\Big)$$

for all $\varepsilon>0$. Applying Hölder's inequality, stationarity and condition (PM) (i), we conclude for $\delta>0$

$$\begin{aligned}
&E\Big[\Big(\sum_{t=1}^{r_n}\log^+\|X_{n,t}\|\Big)^2\mathbb{1}_{\{\sum_{j=1}^{r_n}\log^+\|X_{n,j}\|>\varepsilon\sqrt{nv_n}\}}\Big]\\
&=\sum_{t=1}^{r_n}\sum_{s=1}^{r_n}E\Big[\log^+\|X_{n,t}\|\log^+\|X_{n,s}\|\mathbb{1}_{\{\sum_{j=1}^{r_n}\log^+\|X_{n,j}\|>\varepsilon\sqrt{nv_n}\}}\Big]\\
&\le\sum_{t=1}^{r_n}\sum_{s=1}^{r_n}E\Big[(\log^+\|X_{n,s}\|\cdot\log^+\|X_{n,t}\|)^{1+\delta}\Big]^{1/(1+\delta)}P\Big(\sum_{j=1}^{r_n}\log^+\|X_{n,j}\|>\varepsilon\sqrt{nv_n}\Big)^{\delta/(1+\delta)}\\
&\le 2r_n\sum_{k=0}^{r_n}E\Big[(\log^+\|X_{n,0}\|\cdot\log^+\|X_{n,k}\|)^{1+\delta}\Big]^{1/(1+\delta)}\\
&\qquad\qquad\times P\Big(\sum_{j=1}^{r_n}\log^+\|X_{n,j}\|>\varepsilon\sqrt{nv_n}\Big)^{\delta/(1+\delta)}\\
&=2r_nv_n^{1/(1+\delta)}\sum_{k=0}^{r_n}E\Big[(\log^+\|X_{n,0}\|\cdot\log^+\|X_{n,k}\|)^{1+\delta}\mid\|X_0\|>u_n\Big]^{1/(1+\delta)}\\
&\qquad\qquad\times\mathrm{O}\Big(\Big(\frac{r_n}{n}\Big)^{\delta/(1+\delta)}\Big)\\
&=\mathrm{O}\Big(r_nv_n^{1/(1+\delta)}\Big(\frac{r_n}{n}\Big)^{\delta/(1+\delta)}\Big)=\mathrm{O}\Big(r_nv_n\Big(\frac{r_n}{nv_n}\Big)^{\delta/(1+\delta)}\Big)=\mathrm{o}(r_nv_n).
\end{aligned}$$

Note that $E\Big[(\log^+\|X_{n,0}\|)^{2(1+\delta)}\mid\|X_0\|>u_n\Big]=\mathrm{O}(1)$ by regular variation of $\|X_0\|$ as before. Thus, condition (L2) is satisfied for ϕ, which implies condition (L) (cf. Lemma 3.1.6). Hence, all condition of Theorem 3.2.3, part (a) for fidis convergence, are fulfilled, except for the covariance convergence.
The convergence of the covariance between $Z_n(\phi)$ and $Z_n(A)$, for any $A\in\mathcal{A}$, and of the standardized variance of $Z_n(\phi)$ is proven in Lemma 5.7.7. Thus, all conditions for the joint convergence of $((Z_n(A))_{A\in\mathcal{A}},Z_n(\phi))$ are fulfilled, such that

$$((Z_n(A))_{A\in\mathcal{A}},Z_n(\phi))\xrightarrow{w}((Z(A))_{A\in\mathcal{A}},Z(\phi)).$$

By similar arguments as in Theorem 5.2.9 and Lemma 3.3.5 and due to the bias conditions (PB_T) and (PB_α) we conclude the joint convergence

$$\sqrt{nv_n}\left(((\hat{p}_{n,A}-P(\Theta_i\in A)))_{A\in\mathcal{A}},\hat{\alpha}_n-\alpha\right)\xrightarrow{w}\left(\left(Z^{pb}(A)\right)_{A\in\mathcal{A}},Z_\alpha\right)\tag{5.7.17}$$

with $Z_\alpha:=\alpha Z(\mathbb{R}^d)-\alpha^2Z(\phi)$ and $Z^{pb}(A)=Z(A)-p_AZ(\mathbb{R}^d)$. This is the main argument

needed to show joint weak convergence of both terms on the right hand side of (5.7.15). To deal with the second term on the right hand side in (5.7.15) we further decompose this term. Recall the shorthands $H_{n,i} = \{(-s_n - i) \vee -s_n, ..., (s_n - i) \wedge s_n\}$ and $H_{n,i}^C = \{-s_n, ..., s_n\} \backslash H_{n,i}$.
Define the function $p_{n,A} : (0, \infty) \to \mathbb{R}$ by

$$p_{n,A}(a) := \frac{1}{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}} \sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h=-s_n}^{s_n} \frac{\|X_{n,t+h}\|^a}{\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^a} \times \left(\mathbb{1}_{\{h \in H_{n,i}\}} \mathbb{1}_A \left(\frac{X_{t+h+i}}{\|X_{t+h}\|} \right) + \mathbb{1}_{\{h \in H_{n,i}^C\}} \mathbb{1}_A(0) \right).$$

Then, $p_{n,A}(\alpha) = \hat{p}_{n,A}$ and $p_{n,A}(\hat{\alpha}_n) = \hat{\hat{p}}_{n,A}$. A Taylor expansion of this function at the point α yields

$$\begin{aligned}
&\hat{\hat{p}}_{n,A} - \hat{p}_{n,A} \\
&= \frac{(\hat{\alpha}_n - \alpha)}{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}} \sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h=-s_n}^{s_n} \Bigg(\frac{\log(\|X_{n,t+h}\|)\|X_{n,t+h}\|^\alpha}{\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^\alpha} \\
&\qquad - \frac{\|X_{n,t+h}\|^\alpha \sum_{k=-s_n}^{s_n} \log(\|X_{n,t+k}\|)\|X_{n,t+k}\|^\alpha}{(\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^\alpha)^2} \Bigg) \\
&\qquad \times \left(\mathbb{1}_{\{h \in H_{n,i}\}} \mathbb{1}_A \left(\frac{X_{t+h+i}}{\|X_{t+h}\|} \right) + \mathbb{1}_{\{h \in H_{n,i}^C\}} \mathbb{1}_A(0) \right) \\
&\quad + \frac{(\hat{\alpha}_n - \alpha)^2}{2\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}} \sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h=-s_n}^{s_n} \Bigg(\frac{\log^2(\|X_{n,t+h}\|)\|X_{n,t+h}\|^{\bar{\alpha}}}{\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar{\alpha}}} \\
&\qquad -2 \frac{\log(\|X_{n,t+h}\|)\|X_{n,t+h}\|^{\bar{\alpha}} \sum_{k=-s_n}^{s_n} \log(\|X_{n,t+k}\|)\|X_{n,t+k}\|^{\bar{\alpha}}}{(\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar{\alpha}})^2} \\
&\qquad - \frac{\|X_{n,t+h}\|^{\bar{\alpha}} \sum_{k=-s_n}^{s_n} \log^2(\|X_{n,t+k}\|)\|X_{n,t+k}\|^{\bar{\alpha}}}{(\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar{\alpha}})^2} \\
&\qquad +2 \frac{\|X_{n,t+h}\|^{\bar{\alpha}} (\sum_{k=-s_n}^{s_n} \log(\|X_{n,t+k}\|)\|X_{n,t+k}\|^{\bar{\alpha}})^2}{(\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar{\alpha}})^3} \Bigg) \\
&\qquad \times \left(\mathbb{1}_{\{h \in H_{n,i}\}} \mathbb{1}_A \left(\frac{X_{t+h+i}}{\|X_{t+h}\|} \right) + \mathbb{1}_{\{h \in H_{n,i}^C\}} \mathbb{1}_A(0) \right) \\
&= I(A) + II(A), \qquad (5.7.18)
\end{aligned}$$

with $\bar{\alpha} = \lambda\alpha + (1 - \lambda)\hat{\alpha}_n$ for some $\lambda \in (0, 1)$.
In Lemma 5.7.6 we show that $II(A)$ is asymptotically uniform negligible, i.e.

$$\sup_{A \in \mathcal{A}} \sqrt{nv_n} |II(A)| = o_P(1)$$

as $n \to \infty$. To deal with term $I(A)$, define

$$d_{A,n}^{(m)} := \frac{1}{nv_n} \sum_{t=1}^n f_A^{(m)}(W_{n,t}), \qquad d_{A,n} := \frac{1}{nv_n} \sum_{t=1}^n f_A(W_{n,t}) \qquad (5.7.19)$$

with $W_{n,t} = (X_{n,t})_{|t|\le s_n}$ and

$$f_A^{(m)}((y_t)_{h\in\mathbb{Z}}) := \mathbb{1}_{\{\|y_0\|>1\}} \sum_{h=-m}^{m} \left(\frac{\log(\|y_h\|)\|y_h\|^\alpha}{\sum_{k=-m}^m \|y_k\|^\alpha} - \frac{\|y_h\|^\alpha \sum_{k=-m}^m \log(\|y_k\|)\|y_k\|^\alpha}{(\sum_{k=-m}^m \|y_k\|^\alpha)^2} \right) \mathbb{1}_A\left(\frac{y_{h+i}}{\|y_h\|}\right) \tag{5.7.20}$$

for all $m \in \mathbb{N}$ and $f_A := f_A^{(\infty)}$ with the convention $\sum_{|h|\le\infty} := \sum_{h\in\mathbb{Z}}$. In addition, recall the definition of d_A from Theorem 5.3.1 and define $d_A^{(m)} := E[f_A^{(m)}((Y_h)_{h\in\mathbb{Z}})]$.
With the usual embedding of $(X_{n,t})_{|t|\le s_n}$ in l_α by defining $X_{n,t} = 0$ for $|t| > s_n$ one has

$$\begin{aligned} f_A(W_{n,t}) = \mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h=-s_n}^{s_n} &\left(\frac{\log(\|X_{n,t+h}\|)\|X_{n,t+h}\|^\alpha}{\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^\alpha} - \frac{\|X_{n,t+h}\|^\alpha \sum_{k=-s_n}^{s_n} \log(\|X_{n,t+k}\|)\|X_{n,t+k}\|^\alpha}{(\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^\alpha)^2} \right) \\ &\times \left(\mathbb{1}_{\{h\in H_{n,i}\}} \mathbb{1}_A\left(\frac{X_{t+h+i}}{\|X_{t+h}\|}\right) + \mathbb{1}_{\{h\in H_{n,i}^C\}} \mathbb{1}_A(0) \right). \end{aligned}$$

This definitions leads to

$$\begin{aligned} &\sqrt{nv_n} I(A) = \sqrt{nv_n}(\hat\alpha_n - \alpha)\frac{nv_n}{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}} d_{A,n} \\ &= \sqrt{nv_n}(\hat\alpha_n - \alpha)\frac{nv_n}{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}} d_{A,n}^{(m)} + \sqrt{nv_n}(\hat\alpha_n - \alpha)\frac{nv_n}{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}} (d_{A,n} - d_{A,n}^{(m)}) \\ &=: I^{(m)}(A) + I^{(m,R)}(A). \end{aligned}$$

The representation of $I(A)$, the convergences of $\sqrt{nv_n} II(A)$ combined with (5.7.15) and (5.7.18) yield

$$\begin{aligned} &\left(\sqrt{nv_n}\left(\hat{\hat{p}}_{n,A} - P(\Theta_i \in A)\right)\right)_{A\in\mathcal{A}} \\ &= \left(\sqrt{nv_n}\left(\hat p_{n,A} - P(\Theta_i \in A)\right)\right)_{A\in\mathcal{A}} + \left(\sqrt{nv_n}(\hat\alpha_n - \alpha)\frac{nv_n}{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}} d_{A,n}^{(m)}\right)_{A\in\mathcal{A}} \\ &\quad + \left(I^{(m,R)}(A)\right)_{A\in\mathcal{A}} + o_P(1) \end{aligned} \tag{5.7.21}$$

for all $m \in \mathbb{N}$.
The convergence $\sup_{A\in\mathcal{A}} |d_{A,n}^{(m)} - d_A^{(m)}| = o_P(1)$, i.e. $d_{A,n}^{(m)} \to d_A^{(m)}$ in probability uniformly for all $A \in \mathcal{A}$, is established in Lemma 5.7.3. The weak convergence in Proposition 5.2.7 for the set $\mathbb{R}^d$ readily implies $nv_n/(\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}) \to 1$ in probability. Note that these two convergences also imply $d_{A,n}^{(m)} nv_n/(\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}) \to d_A^{(m)}$ in probability and uniformly for all $A \in \mathcal{A}$ and for all $m \in \mathbb{N}$, provided $\sup_{A\in\mathcal{A}} |d_A^{(m)}| < \infty$, which is shown in Lemma 5.7.3. This convergence, the joint convergence (5.7.17) and Slutsky's Lemma

(Kosorok (2008), Theorem 7.15 (i)) imply the weak convergence

$$\left((\sqrt{nv_n}\,(\hat{p}_{n,A} - P(\Theta_i \in A)))_{A\in\mathcal{A}}\,, \sqrt{nv_n}(\hat{\alpha}-\alpha), \frac{nv_n}{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}}(d_{A,n}^{(m)})_{A\in\mathcal{A}} \right) \xrightarrow{w} ((Z^{pb}(A))_{A\in\mathcal{A}}, Z_\alpha, (d_A^{(m)})_{A\in\mathcal{A}}).$$

The map $H : \ell^\infty(\mathcal{A}) \times \mathbb{R} \times \ell^\infty(\mathcal{A}) \to \mathbb{R}$, $((x(A)_{A\in\mathcal{A}}), y, (z(A)_{A\in\mathcal{A}})) \mapsto (x(A) + z(A)\cdot y)_{A\in\mathcal{A}}$ is continuous with respect to the supremum norm:

$$\begin{aligned}
&\|(x(A) + z(A)y)_{A\in\mathcal{A}} - (x'(A) + z'(A)y')_{A\in\mathcal{A}}\|_\infty \\
&\le \|(x(A) - x'(A))_{A\in\mathcal{A}}\|_\infty + \sup_{A\in\mathcal{A}} |z(A)y - z(A)y'| + \sup_{A\in\mathcal{A}} |z(A)y' - z'(A)y'| \\
&\le \sup_{A\in\mathcal{A}}(x(A) - x'(A)) + \sup_{A\in\mathcal{A}} |z(A)||y-y'| + 2\sup_{A\in\mathcal{A}} |z(A)||y'| \le \varepsilon
\end{aligned}$$

if $\sup_{A\in\mathcal{A}} \max(|x(A) - x'(A)|, |y-y'|, |z(A) - z'(A)|) < \delta$ and for a suitable ε, i.e. H is continuous in all points with $\sup_{A\in\mathcal{A}} |z(A)| < \infty$. Choose $z(A) = d_A^{(m)}$ and note that $\sup_{A\in\mathcal{A}} |d_A^{(m)}| < \infty$ by Lemma 5.7.3.
Thus, the map H is continuous at the point $((Z^{pb}(A))_{A\in\mathcal{A}}, Z_\alpha, (d_A^{(m)})_{A\in\mathcal{A}})$. Thus, the joint convergence resulting from Slutsky's Lemma and the continuous mapping theorem applied with H imply

$$\begin{aligned}
&(\sqrt{nv_n}\,(\hat{p}_{n,A} - P(\Theta_i \in A)))_{A\in\mathcal{A}} + \left(\sqrt{nv_n}(\hat{\alpha}-\alpha)\frac{nv_n}{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}} d_{A,n}^{(m)}\right)_{A\in\mathcal{A}} \\
&\xrightarrow{w} (Z^{pb}(A) + d_A^{(m)} Z_\alpha)_{A\in\mathcal{A}}
\end{aligned} \tag{5.7.22}$$

for all $m \in \mathbb{N}$.
In Lemma 5.7.4 part (i), it is shown that $\lim_{m\to\infty} \limsup_{n\to\infty} E[|d_{A,n}^{(m)} - d_{A,n}|] = 0$ uniformly for all $A \in \mathcal{A}$, in particular $d_{A,n}^{(m)} - d_{A,n}$ converges to 0 in probability uniformly for all $A \in \mathcal{A}$. Since by the same arguments as before $\sqrt{nv_n}(\hat{\alpha}-\alpha)nv_n/(\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}) \to Z_\alpha$ weakly as $n \to \infty$, Slutsky's Lemma implies that

$$I^{(m,R)}(A) = \sqrt{nv_n}(\hat{\alpha}_n - \alpha)\frac{nv_n}{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}}(d_{A,n} - d_{A,n}^{(m)}) \xrightarrow{P} 0 \tag{5.7.23}$$

uniformly for all $A \in \mathcal{A}$ a.s. as $n \to \infty$ and then $m \to \infty$. Moreover, Lemma 5.7.4 part (ii) shows that $d_A^{(m)} \to d_A$ as $m \to \infty$ uniformly for all $A \in \mathcal{A}$. Thus,

$$(Z^{pb}(A) + d_A^{(m)} Z(A))_{A\in\mathcal{A}} \to (Z^{pb}(A) + d_A Z(A))_{A\in\mathcal{A}} \tag{5.7.24}$$

a.s. as $m \to \infty$. Combining (5.7.21), (5.7.22), (5.7.23) and (5.7.24), standard arguments yield

$$\left(\sqrt{nv_n}\left(\hat{\hat{p}}_{n,A} - P(\Theta_i \in A)\right)\right)_{A\in\mathcal{A}} \xrightarrow{w} (Z^{pb}(A) + d_A Z_\alpha)_{A\in\mathcal{A}} = (Z^{pb,\alpha}(A))_{A\in\mathcal{A}}.$$

The applied standard arguments are similar to e.g. Proposition 6.3.9 in Brockwell and Davis (1991). Since $Z_\alpha := \alpha Z(\mathbb{R}^d) - \alpha^2 Z(\phi)$ and $Z^{pb}(A) = Z(A) - p_A Z(\mathbb{R}^d)$, this concludes the proof. □

Auxiliary Lemmas for the proof of Theorem 5.3.1

In the following a series of lemmas is presented, which state useful assertions which are applied in the previous main proof of Theorem 5.3.1. In Lemma 5.7.1 the convergence of the expected value $\lim_{n\to\infty}(v_n)^{-1}E\left[f_A^{(m)}(W_{n,0})\right]$ is shown. This lemma is a preparation for Lemma 5.7.2 which establishes the pointwise convergence $d_{A,n}^{(m)} \to d_A^{(m)}$ for all $A \in \mathcal{A}$, $m \in \mathbb{N}$. Lemma 5.7.2, in turn, is the preparation of Lemma 5.7.3 which basically shows the uniform convergence $\sup_{A\in\mathcal{A}} |d_{A,n}^{(m)} - d_A^{(m)}| \to 0$ in probability as $n \to \infty$. In Lemma 5.7.4 we establish that the approximation error of using $f_A^{(m)}$ instead of f_A in the main proof is asymptotically negligible.
Lemma 5.7.5 addresses two optimization problems which occur in the proof of Lemma 5.7.6. Lemma 5.7.6 then shows $\sup_{A\in\mathcal{A}} |II(A)| = o_P((nv_n)^{-1/2})$. Finally Lemma 5.7.7 shows the convergence of the covariances needed to establish condition (C) in the previous main proof.

Lemma 5.7.1. *Suppose the conditions (PR), (P0) and (PC) are satisfied. Then,*

$$\lim_{n\to\infty} E\left[\frac{1}{r_n v_n}\sum_{t=1}^{r_n} f_A^{(m)}(W_{n,t})\right] = \lim_{n\to\infty}\frac{1}{v_n}E\left[f_A^{(m)}(W_{n,0})\right] = d_A^{(m)} \in \mathbb{R}$$

for all $A \in \mathcal{A}$ and $m \in \mathbb{N}$, with $d_A^{(m)} = E[f_A^{(m)}((Y_h)_{h\in\mathbb{Z}})]$ and $f_A^{(m)}$ defined in (5.7.20).

Proof. Due to the definition of the tail process $(Y_t)_{t\in\mathbb{Z}}$ one has

$$\mathcal{L}\left(u_n^{-1}(X_t)_{|t|\le m} \mid \|X_0\| > u_n\right) \xrightarrow{w} \mathcal{L}\left((Y_t)_{|t|\le m}\right).$$

The continuous mapping theorem implies

$$\begin{aligned}
&\mathcal{L}\left(f_A^{(m)}\left((X_t/u_n)_{|t|\le s_n}\right) \mid \|X_0\| > u_n\right) \qquad (5.7.25)\\
&\xrightarrow{w} \mathcal{L}\left(f_A^{(m)}\left((Y_t)_{t\in\mathbb{Z}}\right)\right)\\
&= \mathcal{L}\left(\sum_{|h|\le m}\left(\frac{\log(\|Y_h\|)\|Y_h\|^\alpha}{\sum_{|k|\le m}\|Y_k\|^\alpha} - \frac{\|Y_h\|^\alpha \sum_{|k|\le m}\log(\|Y_k\|)\|Y_k\|^\alpha}{(\sum_{|k|\le m}\|Y_k\|^\alpha)^2}\right)\mathbb{1}_A\left(\frac{Y_{h+i}}{\|Y_h\|}\right)\right)
\end{aligned}$$

if $f_A^{(m)}$ is a.s. continuous with respect to the distribution of $(Y_t)_{t\in\mathbb{Z}}$. As a finite sum and composition of continuous functions, the function $f_A^{(m)}$ is P^Y-a.s. continuous if the map $(Y_t)_{t\in\mathbb{Z}} \to \mathbb{1}_A(Y_{h+i}/\|Y_h\|)$ is a.s. continuous for all $|h| \le m$, which is the case if $P(Y_{h+i}/\|Y_h\| \in \partial A, \|Y_h\| > 0) = 0$ for all $|h| \le m$, $A \in \mathcal{A}$. This holds by assumption (PC) and Lemma 5.2.3, i.e. $f_A^{(m)}$ is P^Y-a.s. continuous.

Due to the weak convergence (5.7.25) there exist random variables with the same distribution as $f_A^{(m)}(W_{n,0})$, given $\|X_0\| > u_n$, $n \in \mathbb{N}$, which converge almost surely and therefore also in probability. Convergence in probability together with uniform integrability implies L_1 convergence which in turn implies the convergence of the expected value. The convergence of the expectation is a property of the distribution and therefore the expected value of the originally considered random variables holds. Thus, the weak convergence together with uniform integrability of $v_n^{-1} f_A^{(m)}(W_{n,0})$, $n \in \mathbb{N}$, would imply

$$\frac{1}{r_n v_n} E\Big[\sum_{t=1}^{r_n} f_A^{(m)}(W_{n,t})\Big] = \frac{1}{v_n} E[f_A^{(m)}(W_{n,0})] = E[f_A^{(m)}(W_{n,0}) \mid \|X_0\| > u_n]$$
$$\to E[f_A^{(m)}((Y_h)_{h\in\mathbb{Z}})] = d_A^{(m)}.$$

Hence, it remains to prove the uniform integrability of $v_n^{-1} f_A^{(m)}(W_{n,0})$, $n \in \mathbb{N}$, i.e.

$$\lim_{M\to\infty} \sup_{n\in\mathbb{N}} E\left[|f_A^{(m)}(W_{n,0})| \mathbb{1}_{(M,\infty)}(|f_A(W_{n,0})|) \mid \|X_0\| > u_n\right] = 0.$$

This is implied by the Lyapunov condition

$$\sup_{n\in\mathbb{N}} E\left[|f_A^{(m)}(W_{n,0})|^{1+\eta} \mid \|X_0\| > u_n\right] < \infty$$

for some $\eta > 0$. Due to the Minkowski inequality this condition is satisfied, if

$$\sup_{n\in\mathbb{N}} E\left[\left|\mathbb{1}_{\{\|X_0\|>u_n\}} \sum_{h=-m}^{m} \frac{\log(\|X_{n,h}\|)\|X_{n,h}\|^\alpha}{\sum_{k=-m}^{m}\|X_{n,k}\|^\alpha}\right.\right.$$
$$\left.\left. \times \left(\mathbb{1}_{\{h\in H_{n,i}\}}\mathbb{1}_A\left(\frac{X_{h+i}}{\|X_h\|}\right) + \mathbb{1}_{\{h\in H_{n,i}^C\}}\mathbb{1}_A(0)\right)\right|^{1+\eta} \middle| \|X_0\| > u_n\right] < \infty \tag{5.7.26}$$

and

$$\sup_{n\in\mathbb{N}} E\left[\left|\mathbb{1}_{\{\|X_0\|>u_n\}} \sum_{h=-m}^{m} \frac{\|X_{n,h}\|^\alpha \sum_{k=-m}^{m}\log(\|X_{n,k}\|)\|X_{n,k}\|^\alpha}{(\sum_{k=-m}^{m}\|X_{n,k}\|^\alpha)^2}\right.\right.$$
$$\left.\left. \times \left(\mathbb{1}_{\{h\in H_{n,i}\}}\mathbb{1}_A\left(\frac{X_{h+i}}{\|X_h\|}\right) + \mathbb{1}_{\{h\in H_{n,i}^C\}}\mathbb{1}_A(0)\right)\right|^{1+\eta} \middle| \|X_0\| > u_n\right] < \infty. \tag{5.7.27}$$

An application of the Hölder inequality for sums yields for the term under the expectation in (5.7.26)

$$\Big|\sum_{h=-m}^{m} \frac{\log(\|X_{n,h}\|)\|X_{n,h}\|^\alpha}{\sum_{k=-m}^{m}\|X_{n,k}\|^\alpha} \left(\mathbb{1}_{\{h\in H_{n,i}\}}\mathbb{1}_A\left(\frac{X_{h+i}}{\|X_h\|}\right) + \mathbb{1}_{\{h\in H_{n,i}^C\}}\mathbb{1}_A(0)\right)\Big|^{1+\eta}$$
$$\leq \left(\sum_{h=-m}^{m} \frac{|\log(\|X_{n,h}\|)|\|X_{n,h}\|^\alpha}{\sum_{k=-m}^{m}\|X_{n,k}\|^\alpha}\right)^{1+\eta}$$

$$
\begin{aligned}
&= \left(\sum_{h=-m}^{m} \frac{|\log(\|X_{n,h}\|)| \|X_{n,h}\|^{\alpha/(1+\eta)} \|X_{n,h}\|^{\eta\alpha/(1+\eta)}}{\sum_{k=-m}^{m} \|X_{n,k}\|^{\alpha}} \right)^{1+\eta} \\
&\leq \left(\frac{\left(\sum_{h=-m}^{m} |\log(\|X_{n,h}\|)|^{1+\eta} \|X_{n,h}\|^{\alpha}\right)^{1/(1+\eta)} \left(\sum_{k=-m}^{m} \|X_{n,k}\|^{\alpha}\right)^{\eta/(1+\eta)}}{\sum_{k=-m}^{m} \|X_{n,k}\|^{\alpha}} \right)^{1+\eta} \\
&= \frac{\sum_{h=-m}^{m} |\log(\|X_{n,h}\|)|^{1+\eta} \|X_{n,h}\|^{\alpha}}{\sum_{k=-m}^{m} \|X_{n,k}\|^{\alpha}}. \qquad (5.7.28)
\end{aligned}
$$

Moreover, for the term under the expectation in (5.7.27) similar arguments as leading to (5.7.28) yield

$$
\begin{aligned}
&\Big| \sum_{h=-m}^{m} \frac{\|X_{n,h}\|^{\alpha} \sum_{k=-m}^{m} \log(\|X_{n,k}\|) \|X_{n,k}\|^{\alpha}}{(\sum_{k=-m}^{m} \|X_{n,k}\|^{\alpha})^2} \left(\mathbb{1}_{\{h \in H_{n,i}\}} \mathbb{1}_A \left(\frac{X_{h+i}}{\|X_h\|} \right) + \mathbb{1}_{\{h \in H_{n,i}^C\}} \mathbb{1}_A(0) \right) \Big|^{1+\eta} \\
&\leq \left(\sum_{h=-m}^{m} \frac{\|X_{n,h}\|^{\alpha} \sum_{k=-m}^{m} |\log(\|X_{n,k}\|)| \|X_{n,k}\|^{\alpha}}{(\sum_{k=-m}^{m} \|X_{n,k}\|^{\alpha})^2} \right)^{1+\eta} \\
&= \left(\frac{\sum_{k=-m}^{m} |\log(\|X_{n,k}\|)| \|X_{n,k}\|^{\alpha}}{\sum_{k=-m}^{m} \|X_{n,k}\|^{\alpha}} \right)^{1+\eta} \\
&\leq \frac{\sum_{h=-m}^{m} |\log(\|X_{n,h}\|)|^{1+\eta} \|X_{n,h}\|^{\alpha}}{\sum_{k=-m}^{m} \|X_{n,k}\|^{\alpha}}.
\end{aligned}
$$

Therefore, to establish (5.7.26) and (5.7.27) it suffices to show

$$
\sup_{n \in \mathbb{N}} E \left[\frac{\sum_{h=-m}^{m} |\log(\|X_{n,h}\|)|^{1+\eta} \|X_{n,h}\|^{\alpha}}{\sum_{k=-m}^{m} \|X_{n,k}\|^{\alpha}} \Big| \|X_0\| > u_n \right] < \infty \qquad (5.7.29)
$$

for some $0 < \eta < 1$. Moreover, a decomposition of each summand in the numerator according $\{\|X_h\| \geq u_n\} = \{\|X_{n,h}\| \geq 1\}$ and $\{\|X_h\| < u_n\} = \{\|X_{n,h}\| < 1\}$ leads to the upper bounds

$$
\begin{aligned}
&\frac{\sum_{h=-m}^{m} |\log(\|X_{n,h}\|)|^{1+\eta} \|X_{n,h}\|^{\alpha} \mathbb{1}_{\{\|X_h\| \geq u_n\}}}{\sum_{k=-m}^{m} \|X_{n,k}\|^{\alpha}} = \frac{\sum_{h=-m}^{m} \left(\log^+(\|X_{n,h}\|)\right)^{1+\eta} \|X_{n,h}\|^{\alpha}}{\sum_{k=-m}^{m} \|X_{n,k}\|^{\alpha}} \\
&\leq \sup_{-m \leq h \leq m} \left(\log^+(\|X_{n,h}\|)\right)^{1+\eta} \frac{\sum_{h=-m}^{m} \|X_{n,h}\|^{\alpha}}{\sum_{k=-m}^{m} \|X_{n,k}\|^{\alpha}} \leq \sum_{h=-m}^{m} \left(\log^+(\|X_{n,h}\|)\right)^{1+\eta}
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{1}_{\{\|X_0\| > u_n\}} \frac{\sum_{h=-m}^{m} |\log(\|X_{n,h}\|)|^{1+\eta} \|X_{n,h}\|^{\alpha} \mathbb{1}_{\{\|X_h\| < u_n\}}}{\sum_{k=-m}^{m} \|X_{n,k}\|^{\alpha}} \\
&\leq \mathbb{1}_{\{\|X_0\| > u_n\}} \sum_{h=-m}^{m} |\log(\|X_{n,h}\|)|^{1+\eta} \|X_{n,h}\|^{\alpha} \mathbb{1}_{\{\|X_h\| > u_n\}} \\
&= \mathbb{1}_{\{\|X_0\| > u_n\}} \sum_{h=-m}^{m} \left(\log^-(\|X_{n,h}\|)\right)^{1+\eta} \|X_{n,h}\|^{\alpha}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&E\left[\frac{\sum_{h=-m}^{m}|\log(\|X_{n,h}\|)|^{1+\eta}\|X_{n,h}\|^\alpha}{\sum_{k=-m}^{m}\|X_{n,k}\|^\alpha}\Big|\|X_0\|>u_n\right]\\
&\le E\left[\sum_{h=-m}^{m}\left(\log^+(\|X_{n,h}\|)\right)^{1+\eta}+\sum_{h=-m}^{m}\left(\log^-(\|X_{n,h}\|)\right)^{1+\eta}\|X_{n,h}\|^\alpha\Big|\|X_0\|>u_n\right].
\end{aligned}
$$

According to (5.5.6), where only regular variation was used, we obtain for $h \in \mathbb{Z}$ and sufficiently large n

$$
E\Big[(\|X_{n,h}\|)^q \;\Big|\; \|X_0\|>u_n\Big] \le \int_1^\infty (1+\varepsilon)x^{-(\alpha-\varepsilon)/q}dx + 2 < \infty
$$

for any $q \in (0,\alpha)$. Moreover, $\log^+(x)^{1+\eta} \le c_q x^q$ for all $q \in (0,\alpha)$ and some suitable constant $c_q > 0$. Therefore, it follows for the finite sum

$$
\sum_{h=-m}^{m} E\left[\left(\log^+(\|X_{n,h}\|)\right)^2\Big|\|X_0\|>u_n\right] \le c_q \sum_{h=-m}^{m} E\left[\|X_{n,h}\|^q\Big|\|X_0\|>u_n\right] = \mathrm{O}(1)
$$

(cf. Kulik and Soulier (2020), Section 2.3.3, for a similar result). Moreover, $|\log(x)|^{1+\eta}x^\alpha$ is a bounded function for $x<1$, i.e $\log^-(x)^{1+\eta}x^\alpha < c$ for some $c>0$ which is why

$$
E\left[\sum_{h=-m}^{m}\left(\log^-(\|X_{n,h}\|)\right)^{1+\eta}\|X_{n,h}\|^\alpha\Big|\|X_0\|>u_n\right] \le (2m+1)c = \mathrm{O}(1).
$$

Combining these last two bounds yields (5.7.29). All in all, we have uniform integrability of $v_n^{-1}f_A^{(m)}(W_{n,0})$, $n \in \mathbb{N}$. Therefore, the weak convergence discussed previously implies the convergence of the expected value. □

Lemma 5.7.2. *Suppose the conditions (PR), (P0) and (PC) are satisfied, then*

$$
d_{A,n}^{(m)} \to d_A^{(m)}
$$

as $n \to \infty$ in probability for all $A \in \mathcal{A}$ and $m \in \mathbb{N}$, for random variables $d_{A,n}^{(m)}$ defined in (5.7.19) *and for the constants $d_A^{(m)}$ from Lemma 5.7.1.*

Proof. We rearrange the sum in $d_{A,n}^{(m)}$ by grouping r_n consecutive summands together and then split the sum into two sums each including every second block of length r_n:

$$
d_n^{(m)}(A) = \frac{1}{nv_n}\sum_{l=1}^{2}\sum_{j=1}^{m_{n,i}}\sum_{t=1}^{r_n} f_A^{(m)}(W_{n,(2(j-1)+l-1)r_n+t}) + \frac{1}{nv_n}\sum_{t=m_nr_n+1}^{n} f_A^{(m)}(W_{n,t}), \quad (5.7.30)
$$

where $m_n = \lfloor n/r_n \rfloor$ and $m_{n,l} = \lfloor m_n/2 \rfloor + \mathbb{1}_{\{\lfloor m_n/2 \rfloor 2 + l \le m_n\}} \sim n/(2r_n)$.

Define $F_{A,n}^{(m)}(l,j) = (r_n v_n)^{-1} \sum_{t=1}^{r_n} f_A^{(m)}(W_{n,(2(j-1)+l-1)r_n+t})$. We first prove that

$$E[(F_{A,n}^{(m)}(1,1))^2] = \frac{1}{(r_n v_n)^2} E\left[\left(\sum_{t=1}^{r_n} f_A^{(m)}(W_{n,t})\right)^2\right] = \mathrm{o}\left(\frac{n}{r_n}\right). \tag{5.7.31}$$

One has

$$\begin{aligned}
&E\left[\left(\sum_{t=1}^{r_n} f_A^{(m)}(W_{n,t})\right)^2\right] \\
&= E\Bigg[\Bigg(\sum_{t=1}^{r_n} \mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h=-m}^{m} \Bigg(\frac{\log(\|X_{n,t+h}\|)\|X_{n,t+h}\|^\alpha}{\sum_{k=-m}^{m}\|X_{n,t+k}\|^\alpha} \\
&\qquad - \frac{\|X_{n,t+h}\|^\alpha \sum_{k=-m}^{m}\log(\|X_{n,t+k}\|)\|X_{n,t+k}\|^\alpha}{(\sum_{k=-m}^{m}\|X_{n,t+k}\|^\alpha)^2}\Bigg) \\
&\qquad \times \left(\mathbb{1}_{\{h\in H_{n,i}\}}\mathbb{1}_A\left(\frac{X_{t+h+i}}{\|X_{t+h}\|}\right) + \mathbb{1}_{\{h\in H_{n,i}^C\}}\mathbb{1}_A(0)\right)\Bigg)^2\Bigg] \\
&\le E\Bigg[\Bigg(\sum_{t=1}^{r_n} \mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h=-m}^{m} \Bigg(\frac{|\log(\|X_{n,t+h}\|)|\|X_{n,t+h}\|^\alpha}{\sum_{k=-m}^{m}\|X_{n,t+k}\|^\alpha} \\
&\qquad + \frac{\|X_{n,t+h}\|^\alpha \sum_{k=-m}^{m}|\log(\|X_{n,t+k}\|)|\|X_{n,t+k}\|^\alpha}{(\sum_{k=-m}^{m}\|X_{n,t+k}\|^\alpha)^2}\Bigg)\Bigg)^2\Bigg] \\
&\le 4E\left[\left(\sum_{t=1}^{r_n} \mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h=-m}^{m} \frac{|\log(\|X_{n,t+h}\|)|\|X_{n,t+h}\|^\alpha}{\sum_{k=-m}^{m}\|X_{n,t+k}\|^\alpha}\right)^2\right] \\
&\le 8E\left[\left(\sum_{t=1}^{r_n} \mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h=-m}^{m} \frac{|\log(\|X_{n,t+h}\|)|\|X_{n,t+h}\|^\alpha}{\sum_{k=-m}^{m}\|X_{n,t+k}\|^\alpha}\mathbb{1}_{\{\|X_{t+h}\|\ge u_n\}}\right)^2\right] \\
&\quad + 8E\left[\left(\sum_{t=1}^{r_n} \mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h=-m}^{m} \frac{|\log(\|X_{n,t+h}\|)|\|X_{n,t+h}\|^\alpha}{\sum_{k=-m}^{m}\|X_{n,t+k}\|^\alpha}\mathbb{1}_{\{\|X_{t+h}\|<u_n\}}\right)^2\right] \\
&=: 8(I_I + I_{II}).
\end{aligned}$$

For the second term I_{II} it holds

$$\begin{aligned}
\frac{I_{II}\cdot r_n}{n(r_n v_n)^2} &= \frac{r_n}{n(r_n v_n)^2} E\left[\left(\sum_{t=1}^{r_n} \mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h=-m}^{m} \frac{\log^-(\|X_{n,t+h}\|)\|X_{n,t+h}\|^\alpha}{\sum_{k=-m}^{m}\|X_{n,t+k}\|^\alpha}\right)^2\right] \\
&\le \frac{r_n}{n(r_n v_n)^2} E\left[\left(\sum_{t=1}^{r_n}(2m+1)c\mathbb{1}_{\{\|X_t\|>u_n\}}\right)^2\right] \\
&= \mathrm{O}\left(\frac{r_n^2 v_n}{n(r_n v_n)^2}\right) = \mathrm{O}\left(\frac{1}{n v_n}\right) = \mathrm{o}(1)
\end{aligned}$$

for some $c > 0$. In the second step we applied that the denominator is at least 1 by the indicator $\mathbb{1}_{\{\|X_t\|>u_n\}}$ and that $|\log(x)|x$ is a bounded function for $x \in [0,1]$, i.e. $\log(x)^- x^\alpha \le c$ for some $c > 0$. The third step holds because of (5.7.10). Using stationarity as well as the Cauchy-Schwarz inequality both for expectations and sums, one obtains for the first term I_I

$$\begin{aligned}
I_I &= \sum_{t=1}^{r_n}\sum_{s=1}^{r_n} E\Bigg[\Bigg(\mathbb{1}_{\{\|X_t\|>u_n\}}\sum_{h=-m}^{m}\frac{\log^+(\|X_{n,t+h}\|)\|X_{n,t+h}\|^\alpha}{\sum_{k=-m}^{m}\|X_{n,t+k}\|^\alpha}\Bigg)\\
&\qquad\qquad\times\Bigg(\mathbb{1}_{\{\|X_s\|>u_n\}}\sum_{h=-m}^{m}\frac{\log^+(\|X_{n,s+h}\|)\|X_{n,s+h}\|^\alpha}{\sum_{k=-m}^{m}\|X_{n,s+k}\|^\alpha}\Bigg)\Bigg]\\
&\le \sum_{t=1}^{r_n}\sum_{s=1}^{r_n}\Bigg(E\Bigg[\Bigg(\mathbb{1}_{\{\|X_t\|>u_n\}}\sum_{h=-m}^{m}\frac{\log^+(\|X_{n,t+h}\|)\|X_{n,t+h}\|^\alpha}{\sum_{k=-m}^{m}\|X_{n,t+k}\|^\alpha}\Bigg)^2\Bigg]\\
&\qquad\qquad\times E\Bigg[\Bigg(\mathbb{1}_{\{\|X_s\|>u_n\}}\sum_{h=-m}^{m}\frac{\log^+(\|X_{n,s+h}\|)\|X_{n,s+h}\|^\alpha}{\sum_{k=-m}^{m}\|X_{n,s+k}\|^\alpha}\Bigg)^2\Bigg]\Bigg)^{1/2}\\
&= r_n^2 E\Bigg[\Bigg(\mathbb{1}_{\{\|X_0\|>u_n\}}\sum_{h=-m}^{m}\frac{\log^+(\|X_{n,h}\|)\|X_{n,h}\|^\alpha}{\sum_{k=-m}^{m}\|X_{n,k}\|^\alpha}\Bigg)^2\Bigg]\\
&= r_n^2 v_n E\Bigg[\Bigg(\sum_{h=-m}^{m}\frac{\log^+(\|X_{n,h}\|)\|X_{n,h}\|^\alpha}{\sum_{k=-m}^{m}\|X_{n,k}\|^\alpha}\mathbb{1}_{\{\|X_h\|>u_n\}}\Bigg)^2\Bigg|\|X_0\|>u_n\Bigg]\\
&\le r_n^2 v_n E\Bigg[\Bigg(\frac{\big(\sum_{h=-m}^{mn}(\log^+(\|X_{n,h}\|))^2\|X_{n,h}\|^\alpha\big)^{1/2}\big(\sum_{k=-m}^{m}\|X_{n,k}\|^\alpha\big)^{1/2}}{\sum_{k=-m}^{m}\|X_{n,k}\|^\alpha}\Bigg)^2\Bigg|\|X_0\|>u_n\Bigg]\\
&= r_n^2 v_n E\Bigg[\frac{\sum_{h=-m}^{m}(\log^+(\|X_{n,h}\|))^2\|X_{n,h}\|^\alpha}{\sum_{k=-m}^{m}\|X_{n,k}\|^\alpha}\Bigg|\|X_0\|>u_n\Bigg]\\
&\le r_n^2 v_n \sum_{h=-m}^{m} E\Big[\big(\log^+(\|X_{n,h}\|)\big)^2\Big|\|X_0\|>u_n\Big].
\end{aligned}$$

Due to the regular variation of the time series $(X_t)_{t\in\mathbb{Z}}$, we obtain using (5.5.6)

$$E\Big[(\|X_{n,h}\|)^q \mid \|X_0\|>u_n\Big] = \mathrm{O}(1)$$

for any $q\in(0,\alpha)$. Moreover, $\log^+(x)^2\le c_q x^q$ for all $q\in(0,\alpha)$ and some suitable constant $c_q>0$. Therefore, it follows for the finite sum

$$\begin{aligned}
I_I &\le r_n^2 v_n \sum_{h=-m}^{m} E\Big[\big(\log^+(\|X_{n,h}\|)\big)^2 \Big| \|X_0\|>u_n\Big]\\
&\le r_n^2 v_n c_q \sum_{h=-m}^{m} E\Big[\|X_{n,h}\|^q \Big| \|X_0\|>u_n\Big] = \mathrm{O}(r_n^2 v_n) = \mathrm{o}(n r_n v_n^2),
\end{aligned}$$

where we also applied $r_n=\mathrm{o}(\sqrt{n v_n})$. Combining this with the bound for I_{II} we conclude

$$\frac{r_n}{n}E[F_{A,n}^{(m)}(1,1)^2] = \frac{r_n}{n(r_n v_n)^2}I_I + \frac{r_n}{n(r_n v_n)^2}I_{II} = \mathrm{o}(1)$$

and (5.7.31) holds.

For the last sum in (5.7.30) one has

$$\Big|\frac{1}{nv_n}\sum_{t=m_nr_n+1}^{n} f_A^{(m)}(W_{n,t})\Big| \le \frac{1}{nv_n}\sum_{t=n-r_n+1}^{n} |f_A^{(m)}(W_{n,t})| \overset{d}{=} \frac{r_n}{n}\frac{1}{r_nv_n}\sum_{t=1}^{r_n}|f_A^{(m)}(W_{n,t})|.$$

By Chebyschef's inequality and (5.7.31) he have

$$\begin{aligned} P\left(\frac{r_n}{n}\frac{1}{r_nv_n}\sum_{t=1}^{r_n}|f_A^{(m)}(W_{n,t})| > \varepsilon\right) &\le \frac{E\left[\left((r_nv_n)^{-1}\sum_{t=1}^{r_n}|f_A^{(m)}(W_{n,t})|\right)^2\right]}{\varepsilon}\frac{r_n^2}{n^2} \\ &= E[(F_{A,n}^{(m)}(1,1))^2]\frac{r_n^2}{\varepsilon^2n^2} = \mathrm{o}\left(\frac{r_n}{n}\right) \to 0 \end{aligned}$$

for all $\varepsilon > 0$. Hence, the last sum in (5.7.30) converges to 0 in probability.
Observe that $F_{A,n}^{(m)}(l,j)$ is measurable w.r.t. $(X_{(2(j-1)+l-1)r_n+1-s_n}, ..., X_{(2(j-1)+l)r_n+s_n})$. For fixed $l \in \{1,2\}$ and different $j \in \{1,...,m_{n,l}\}$ the random variables $F_{A,n}^{(m)}(l,j)$ are measurable functions of the X_t which are separated in time by $r_n - 2s_n - 1$ observations. Denote with $F_{A,n}^{(m*)}(l,j)$, $j = 1,...,m_{n,l}$, independent copies of $F_{A,n}^{(m)}(l,j)$. Then, due to Eberlein's inequality (Eberlein, 1984), we obtain

$$\|P^{(F_{A,n}^{(m*)}(l,j))_{j=1,\dots,m_{n,l}}} - P^{(F_{A,n}^{(m)}(l,j))_{j=1,\dots,m_{n,l}}}\|_{TV} \le m_{n,l}\beta^X_{n,r_n-2s_n-1} \to 0$$

for $l = 1,2$. The convergence $n/r_n\beta^X_{n,r_n-2s_n-1} \to 0$ holds due to the mixing condition in (P0), since $r_n - 2s_n - 1 \ge l_n - s_n$ holds for n large enough by $s_n \le l_n = \mathrm{o}(r_n)$.
Thus, $r_n/n\sum_{j=1}^{m_{n,l}} F_A^{(m)}(l,j)$ converges in $\mathbb{R}$ if and only if $r_n/n\sum_{j=1}^{m_{n,l}} F_{A,n}^{(m*)}(l,j)$ converges and in case of convergence the limits coincide. By construction $F_{A,n}^{(m*)}(l,j)$, $j = 1,...,m_{n,l}$, $l = 1,2$ are iid. Moreover, since by definition of $F_{A,n}^{(m)}(l,j)$

$$d_{A,n}^{(m)} = \sum_{l=1}^{2}\frac{r_n}{n}\sum_{j=1}^{m_{n,l}} F_{A,n}^{(m)}(l,j) + \mathrm{o}_P(1),$$

the convergence $d_{A,n}^{(m)} \to d_A^{(m)}$ holds if and only if $r_n/n\sum_{j=1}^{m_{n,l}} F_{A,n}^{(m*)}(l,j) \to d_A^{(m)}/2$ in probability for $l \in \{1,2\}$.
By Chebyshev's inequality, (5.7.31) and for all $\varepsilon > 0$

$$\begin{aligned} &P\left(\Big|\frac{r_n}{n}\sum_{j=1}^{m_{n,l}} F_{A,n}^{(m*)}(l,j) - \frac{d_A^{(m)}}{2}\Big| > \varepsilon\right) \\ &\le P\left(\Big|\frac{r_n}{n}\sum_{j=1}^{m_{n,l}} F_{A,n}^{(m*)}(l,j) - E\Big[\frac{r_n}{n}\sum_{j=1}^{m_{n,l}} F_{A,n}^{(m*)}(l,j)\Big]\Big| + \Big|E\Big[\frac{r_n}{n}\sum_{j=1}^{m_{n,l}} F_{A,n}^{(m*)}(l,j)\Big] - \frac{d_A^{(m)}}{2}\Big| > \frac{\varepsilon}{2}\right) \\ &\le P\left(\frac{r_n}{n}\Big|\sum_{j=1}^{m_{n,l}} F_{A,n}^{(m*)}(l,j) - E\Big[\sum_{j=1}^{m_{n,l}} F_{A,n}^{(m*)}(l,j)\Big]\Big| > \frac{\varepsilon}{2}\right) \end{aligned}$$

$$+ P\left(\left|E\Big[\frac{r_n}{n}\sum_{j=1}^{m_{n,l}} F_{A,n}^{(m*)}(l,j)\Big] - \frac{d_A^{(m)}}{2}\right| > \frac{\varepsilon}{2}\right)$$
$$\le \frac{4r_n^2}{n^2\varepsilon^2} Var\left(\sum_{j=1}^{m_{n,l}} F_{A,n}^{(m*)}(l,j)\right) + P\left(\left|E\Big[\frac{r_n}{n} m_{n,l} F_{A,n}^{(m)}(1,1)\Big] - \frac{d_A^{(m)}}{2}\right| > \frac{\varepsilon}{2}\right)$$
$$\le \frac{4r_n^2}{n^2\varepsilon^2} m_{n,l} E[(F_{A,n}^{(m)}(1,1))^2] + P\left(\left|\frac{r_n m_{n,l}}{n v_n} E[f_A^{(m)}(W_{n,0})] - \frac{d_A^{(m)}}{2}\right| > \frac{\varepsilon}{2}\right)$$
$$= \mathrm{o}\left(\frac{r_n^2}{n^2}\frac{n}{r_n}\frac{n}{r_n}\right) + \mathrm{o}(1) = \mathrm{o}(1).$$

Here, we also applied Lemma 5.7.1, which is why the second probability is of order o(1), since $m_{n,l} \sim n/(2r_n)$ and thereby

$$E\left[\frac{r_n}{n} m_{n,l} F_{A,n}^{(m)}(1,1)\right] = \frac{r_n m_{n,l}}{n v_n} E[f_A^{(m)}(W_{n,0})] \to d_A^{(m)}/2.$$

Thus, $r_n/n \sum_{j=1}^{m_{n,l}} F_{A,n}^{(m*)}(l,j) \to d_A^{(m)}/2$ in probability. Therefore, $d_{A,n}^{(m)} \to 2(d_A^{(m)}/2) = d_A^{(m)}$ in probability for each $A \in \mathcal{A}$. □

Lemma 5.7.3. *Suppose (PR), (P0), (PC) and (PA) are satisfied. Then,* $\sup_{A\in\mathcal{A}} |d_A^{(m)}| < \infty$ *and*

$$\sup_{A\in\mathcal{A}} |d_{A,n}^{(m)} - d_A^{(m)}| = o_P(1)$$

for all $m \in \mathbb{N}$, *where* $d_{A,n}^{(m)}$ *is defined in* (5.7.19) *and* $d_A^{(m)}$ *is given in Lemma 5.7.1.*

Proof. Fix $m \in \mathbb{N}$. We already know from Lemma 5.7.2 that the convergence $d_{A,n}^{(m)} \to d_A^{(m)}$ holds in probability and pointwise for all $A \in \mathcal{A}$. Denote

$$f_{A,I}((w_h)_{h\in\mathbb{Z}}) = \mathbb{1}_{\{\|w_0\|>1\}} \sum_{|h|\le m} \frac{\log^+(\|w_h\|)\|w_h\|^\alpha}{\sum_{|k|\le m}\|w_k\|^\alpha} \mathbb{1}_A\left(\frac{w_{h+i}}{\|w_h\|}\right),$$
$$f_{A,II}((w_h)_{h\in\mathbb{Z}}) = \mathbb{1}_{\{\|w_0\|>1\}} \sum_{|h|\le m} \frac{\log^-(\|w_h\|)\|w_h\|^\alpha}{\sum_{|k|\le m}\|w_k\|^\alpha} \mathbb{1}_A\left(\frac{w_{h+i}}{\|w_h\|}\right),$$
$$f_{A,III}((w_h)_{h\in\mathbb{Z}}) = \mathbb{1}_{\{\|w_0\|>1\}} \sum_{|h|\le m} \frac{\|w_h\|^\alpha \sum_{|k|\le m}\log^+(\|w_k\|)\|w_k\|^\alpha}{(\sum_{|k|\le m}\|w_k\|^\alpha)^2} \mathbb{1}_A\left(\frac{w_{h+i}}{\|w_h\|}\right),$$
$$f_{A,IV}((w_h)_{h\in\mathbb{Z}}) = \mathbb{1}_{\{\|w_0\|>1\}} \sum_{|h|\le m} \frac{\|w_h\|^\alpha \sum_{|k|\le m}\log^-(\|w_k\|)\|w_k\|^\alpha}{(\sum_{|k|\le m}\|w_k\|^\alpha)^2} \mathbb{1}_A\left(\frac{w_{h+i}}{\|w_h\|}\right).$$

Note that these functions depends on $m \in \mathbb{N}$, however to ease the notation we suppress this m. The families of random variables $v_n^{-1} f_{A,\sharp}(W_{n,0})$, $n \in \mathbb{N}$, are uniformly integrable due to (5.7.26) and (5.7.27) in the proof of Lemma 5.7.1. Therefore, with the same arguments as for $d_{A,n}^{(m)}$ in the proof of Lemma 5.7.2 it follows

$$d_{A,n}^{\sharp} := \frac{1}{n v_n}\sum_{t=1}^{n} f_{A,\sharp} \xrightarrow{P} \lim_{n\to\infty} \frac{1}{v_n} E\left[f_{A,\sharp}(W_{n,0})\right] = E\left[f_{A,\sharp}((Y_t)_{t\in\mathbb{Z}})\right] =: d_A^{\sharp} < \infty$$

in probability for all $A \in \mathcal{A}$ and $\sharp \in \{I, II, III, IV\}$ and the limit on the right hand side exists. Indeed, by the same arguments this even holds for $A \in \mathcal{A} \cup \tilde{\mathcal{A}}$ with $\tilde{\mathcal{A}} = \{A_t^- | t \in [0, 1+\iota]^q\}$. Moreover, by definition $f_A^{(m)} = f_{A,I} - f_{A,II} - f_{A,III} + f_{A,IV}$ and $d_A^{(m)} = d_A^I - d_A^{II} - d_A^{III} + d_A^{IV}$. In particular, $d_{\mathbb{R}^d}^\sharp < \infty$ for $\sharp \in \{I, II, III, IV\}$ and this implies

$$\sup_{A\in\mathcal{A}} |d_A^{(m)}| \leq \sum_{\sharp\in\{I,II,III,IV\}} \sup_{A\in\mathcal{A}} |d_A^\sharp| \leq \sum_{\sharp\in\{I,II,III,IV\}} |d_{\mathbb{R}^d}^\sharp| < \infty,$$

which shows the first assertion $\sup_{A\in\mathcal{A}} |d_A^{(m)}| < \infty$.
By the decomposition of $d_{n,\sharp}$ and $d_\sharp$ we obtain

$$P\left(\sup_{A\in\mathcal{A}} |d_{A,n}^{(m)} - d_A^{(m)}| > \varepsilon\right) \leq \sum_{\sharp\in\{I,II,III,IV\}} P\left(\sup_{A\in\mathcal{A}} |d_{A,n}^\sharp - d_A^\sharp| > \frac{\varepsilon}{4}\right). \tag{5.7.32}$$

Thus, in order to show that $d_{A,n}^{(m)} \to d_A^{(m)}$ in probability uniformly for all $A \in \mathcal{A}$ it is enough to show that $d_{A,n}^\sharp \to d_A^\sharp$ in probability uniformly for all $A \in \mathcal{A}$ and for all $\sharp \in \{I, II, III, IV\}$, i.e. $\sup_{A\in\mathcal{A}} |d_{A,n}^\sharp - d_A^\sharp| = o_P(1)$. Using some bracketing arguments, we will show this in the next step.
Fix some $\varepsilon > 0$ and for the beginning $k \in \{1, ..., q\}$. By assumption (PA) (ii) the map $[0,1] \ni t \mapsto A_{t^{(k)}}$ is non-decreasing. Therefore, $[0,1] \ni t \mapsto E[f_{A_{t^{(k)}},\sharp}((Y_t)_{t\in\mathbb{Z}})]$ is non-decreasing since the summands in $f_{A,\sharp}$ all have a constant positive sign. Thus, $E[f_{A_{t^{(k)}},\sharp}((Y_t)_{t\in\mathbb{Z}})]$, $t \in [0,1]$ is linearly ordered. With condition (PA) (v) and the same arguments as leading to (5.7.11), this map is also right-continuous. The measure generated by this non-decreasing and right-continuous function is finite since $d_{\mathbb{R}^d}^\sharp < \infty$. We have $\|f_{\mathbb{R}^d,\sharp}\|_{L_1(P^Y)} = d_{\mathbb{R}^d}^\sharp < \infty$, where $\|\cdot\|_{L_1(P^Y)}$ denotes the L_1-norm corresponding to the probability measure $p^Y = P^{(Y_t)_{t\in\mathbb{Z}}}$.
Therefore, with the same arguments as in the proof of Proposition 5.2.7, there exist a finite $J_k \in \mathbb{N}$ and $0 =: t_{k,0} < t_{k,1} < ... < t_{k,J_k} := 1$ such that

$$E\left[f_{\bigcup_{s<t_{k,j}} A_{s^{(k)}},\sharp}((Y_t)_{t\in\mathbb{Z}}) - f_{A_{t_{k,j-1}^{(k)}},\sharp}((Y_t)_{t\in\mathbb{Z}})\right] = \|f_{\bigcup_{s<t_{k,j}} A_{s^{(k)}},\sharp} - f_{A_{t_{k,j-1}^{(k)}},\sharp}\|_{L_1(P^Y)} \leq \varepsilon \tag{5.7.33}$$

for all $1 \leq j \leq J_k$. As in the proof of Proposition 5.2.7 one could choose $t_{k,j}$ iteratively by

$$t_{k,j} = \inf\left\{s \in (t_{k,j-1}, 1] : E\Big[f_{A_{s^{(k)}},\sharp}((Y_t)_{t\in\mathbb{Z}}) - f_{A_{t_{k,j-1}^{(k)}},\sharp}((Y_t)_{t\in\mathbb{Z}})\Big] > \varepsilon\right\}.$$

Then, define the partition $\mathcal{T}_k = \{[t_{k,j-1}, t_{k,j}) | 1 \leq j \leq J_k\} \cup \{\{1\}\}$ of $[0,1]$. The basic idea of this construction is, that the indexes $t_{k,j}$ split up the mass of $E[f_{\mathbb{R}^d,\sharp}((Y_t)_{t\in\mathbb{Z}})] = \|f_{\mathbb{R}^d,\sharp}\|_{L_1(P^Y)}$ into finite many ε-brackets which cover $(f_{A,\sharp})_{A\in\mathcal{A}}$. This construction is done for all $1 \leq k \leq q$.

Using this construction we can define brackets for $(f_{A,\sharp})_{A\in\mathcal{A}}$ as

$$[T_1, ..., T_q] = \{f_{A_{s_1,...,s_q},\sharp} | s_k \in T_k \forall 1 \le k \le q\}$$

for all $t_k \in \mathcal{T}_k$, $1 \le k \le q$. By construction of the indexes these are $q \cdot \varepsilon$-brackets w.r.t. $L_1(P^Y)$-distance. Recall the notation $\underline{A}_T$, $\bar{A}_T$ from (5.7.12). Then we have for all $s, t \in T = \times_{k=1}^q T_k$

$$\begin{aligned}
&\|f_{A_s,\sharp} - f_{A_t,\sharp}\|_{L_1(P^Y)} = E\left[f_{A_s,\sharp}((Y_h)_{h\in\mathbb{Z}}) - f_{A_t,\sharp}((Y_h)_{h\in\mathbb{Z}})\right] \\
&\le E\left[f_{\bar{A}_T,\sharp}((Y_h)_{h\in\mathbb{Z}}) - f_{\underline{A}_T,\sharp}((Y_h)_{h\in\mathbb{Z}})\right] \\
&\le \sum_{k=1}^{q} E\left[f_{\bigcup_{r\in T_k} A_r,\sharp}((Y_h)_{h\in\mathbb{Z}}) - f_{A_{(\min T)^{(k)}},\sharp}((Y_h)_{h\in\mathbb{Z}})\right] \le q \cdot \varepsilon,
\end{aligned}$$

where the second step holds due to condition (PA) (ii) and the last inequality directly follows from (5.7.33).
Moreover, these brackets cover $(f_{A,\sharp})_{A\in\mathcal{A}}$ since $\mathcal{T}_k$ is a partition of $[0,1]$, $1 \le k \le q$. These are $\prod_{k=1}^q (J_k + 1) < \infty$ many $q\varepsilon$-brackets for $(f_{A,\sharp})_{A\in\mathcal{A}}$ for arbitrary $\varepsilon > 0$ and $\sharp \in \{I, II, III, IV\}$. Thus, $(f_{A,\sharp})_{A\in\mathcal{A}}$ is covered by finitely many $q\varepsilon$-brackets w.r.t. the $L_1(P^Y)$-norm. Therefore,

$$N_{[\cdot]}(\varepsilon, (f_{A,\sharp})_{A\in\mathcal{A}}, L_1(P^Y)) < \infty$$

for all $\varepsilon > 0$. Here $N_{[\cdot]}$ denotes the bracketing number of $(f_{A,\sharp})_{A\in\mathcal{A}}$ (see also condition (D2) in Section 3.1.2).
Since $d^{\sharp}_{A,n}$ converges pointwise for all $A \in \mathcal{A} \cup \tilde{\mathcal{A}}$, it follows with the same proof as for Theorem 2.4.1 in Van der Vaart and Wellner (1996) that $\sup_{A\in\mathcal{A}} |d^{\sharp}_{A,n} - d^{\sharp}_A| = o_P(1)$ for all $\sharp \in \{I, II, III, IV\}$. Thus, with (5.7.32) it follows $d^{(m)}_{A,n} \xrightarrow{P} d^{(m)}_A$ uniformly for all $A \in \mathcal{A}$. This concludes the proof. □

The following lemma basically deals with the rest term $I^{(m,R)}$ in the main proof.

Lemma 5.7.4. *(i) Suppose (PR), (P0), (PP), (PT), (PP1), (PM) (ii) and (iii) are satisfied. Then,*

$$\lim_{m\to\infty} \limsup_{n\to\infty} \sup_{A\in\mathcal{A}} E[|d^{(m)}_{A,n} - d_{A,n}|] = 0.$$

(ii) Suppose (PR), (P0) and (PM) (ii) are satisfied. Then,

$$\lim_{m\to\infty} \sup_{A\in\mathcal{A}} |d^{(m)}_A - d_A| = 0.$$

Proof. First note that due to stationarity

$$E[|d^{(m)}_{A,n} - d_{A,n}|] = \frac{1}{nv_n} E[|\sum_{t=1}^{n} (f^{(m)}_A(W_{n,t}) - f_A(W_{n,t}))|]$$

$$\leq \frac{1}{nv_n}\sum_{t=1}^{n} E[|f_A^{(m)}(W_{n,t}) - f_A(W_{n,t})|] = E[|f_A^{(m)}(W_{n,0}) - f_A(W_{n,0})| \mid \|X_0\| > u_n].$$

For n sufficiently large, such that $m+|i| \leq s_n$ (this is only needed to simplify the notation with the indicators a little bit), one has

$$\begin{aligned}
&E[|f_A^{(m)}(W_{n,0}) - f_A(W_{n,0})| \mid \|X_0\| > u_n] \\
&\leq E\Big[\Big|\sum_{|h|\leq m} \frac{\log(\|X_{n,h}\|)\|X_{n,h}\|^\alpha}{\sum_{|k|\leq m}\|X_{n,k}\|^\alpha}\mathbb{1}_A\Big(\frac{X_{n,h+i}}{\|X_{n,h}\|}\Big) \\
&\quad - \sum_{|h|\leq s_n} \frac{\log(\|X_{n,h}\|)\|X_{n,h}\|^\alpha}{\sum_{|k|\leq s_n}\|X_{n,k}\|^\alpha}\Big(\mathbb{1}_{\{h\in H_n\}}\mathbb{1}_A\Big(\frac{X_{t+h+i}}{\|X_{t+h}\|}\Big) + \mathbb{1}_{\{h\in H_n^C\}}\mathbb{1}_A(0)\Big)\Big| \,\Big|\, \|X_0\| > u_n\Big] \\
&\quad + E\Big[\Big|\frac{\sum_{|h|\leq m}\|X_{n,h}\|^\alpha \sum_{|k|\leq m}\log(\|X_{n,k}\|)\|X_{n,k}\|^\alpha \mathbb{1}_A\Big(\frac{X_{n,h+i}}{\|X_{n,h}\|}\Big)}{(\sum_{|k|\leq m}\|X_{n,k}\|^\alpha)^2} \\
&\quad - \frac{\sum_{|h|\leq s_n}\|X_{n,h}\|^\alpha \sum_{|k|\leq s_n}\log(\|X_{n,k}\|)\|X_{n,k}\|^\alpha}{(\sum_{|k|\leq s_n}\|X_{n,k}\|^\alpha)^2} \\
&\qquad\qquad \times \Big(\mathbb{1}_{\{h\in H_{n,i}\}}\mathbb{1}_A\Big(\frac{X_{t+h+i}}{\|X_{t+h}\|}\Big) + \mathbb{1}_{\{h\in H_{n,i}^C\}}\mathbb{1}_A(0)\Big)\Big| \,\Big|\, \|X_0\| > u_n\Big] \\
&:= T_1 + T_2.
\end{aligned}$$

Expanding the fractions to the same denominator, adding a 0 in the middle, using (5.7.2) and the triangular inequality, taking the absolute value of each summand and bounding the indicators by 1 yields

$$\begin{aligned}
T_1 = E\Big[\Big|&\frac{1}{\sum_{|k|\leq m}\|X_{n,k}\|^\alpha \sum_{|k|\leq s_n}\|X_{n,k}\|^\alpha} \\
&\times \Big(\sum_{|h|\leq m}\log(\|X_{n,h}\|)\|X_{n,h}\|^\alpha \mathbb{1}_A\Big(\frac{X_{n,h+i}}{\|X_{n,h}\|}\Big)\sum_{|k|\leq s_n}\|X_{n,k}\|^\alpha \\
&\quad - \sum_{|h|\leq m}\log(\|X_{n,h}\|)\|X_{n,h}\|^\alpha \mathbb{1}_A\Big(\frac{X_{n,h+i}}{\|X_{n,h}\|}\Big)\sum_{|k|\leq m}\|X_{n,k}\|^\alpha \\
&\quad + \sum_{|h|\leq m}\log(\|X_{n,h}\|)\|X_{n,h}\|^\alpha \mathbb{1}_A\Big(\frac{X_{n,h+i}}{\|X_{n,h}\|}\Big)\sum_{|k|\leq m}\|X_{n,k}\|^\alpha \\
&\quad - \sum_{|h|\leq s_n}\log(\|X_{n,h}\|)\|X_{n,h}\|^\alpha \sum_{|k|\leq m}\|X_{n,k}\|^\alpha \\
&\qquad\qquad \times \Big(\mathbb{1}_{\{h\in H_{n,i}\}}\mathbb{1}_A\Big(\frac{X_{t+h+i}}{\|X_{t+h}\|}\Big) + \mathbb{1}_{\{h\in H_{n,i}^C\}}\mathbb{1}_A(0)\Big)\Big)\Big| \,\Big|\, \|X_0\| > u_n\Big] \\
= E\Big[\Big|&\frac{1}{\sum_{|k|\leq m}\|X_{n,k}\|^\alpha \sum_{|k|\leq s_n}\|X_{n,k}\|^\alpha} \\
&\times \Big(\sum_{|h|\leq m}\log(\|X_{n,h}\|)\|X_{n,h}\|^\alpha \mathbb{1}_A\Big(\frac{w_{h+i}}{\|w_h\|}\Big)\sum_{m<|k|\leq s_n}\|X_{n,k}\|^\alpha \\
&\quad - \sum_{m<|h|\leq s_n}\log(\|X_{n,h}\|)\|X_{n,h}\|^\alpha \sum_{|k|\leq m}\|X_{n,k}\|^\alpha
\end{aligned}$$

$$\times\Big(\mathbb{1}_{\{h\in H_{n,i}\}}\mathbb{1}_A\Big(\frac{X_{t+h+i}}{\|X_{t+h}\|}\Big)+\mathbb{1}_{\{h\in H_{n,i}^C\}}\mathbb{1}_A(0)\Big)\Big)\Big|\ \Big|\ \|X_0\|>u_n\Big]$$

$$\le E\Bigg[\frac{\sum_{|h|\le m}|\log(\|X_{n,h}\|)|\|X_{n,h}\|^\alpha\mathbb{1}_A\Big(\frac{X_{n,h+i}}{\|X_{n,h}\|}\Big)\sum_{m<|k|\le s_n}\|X_{n,k}\|^\alpha}{\sum_{|k|\le m}\|X_{n,k}\|^\alpha\sum_{|k|\le s_n}\|X_{n,k}\|^\alpha}\ \Big|\ \|X_0\|>u_n\Bigg]$$
$$+E\Bigg[\sum_{m<|h|\le s_n}\frac{|\log(\|X_{n,h}\|)|\|X_{n,h}\|^\alpha\sum_{|k|\le m}\|X_{n,k}\|^\alpha}{\sum_{|k|\le m}\|X_{n,k}\|^\alpha\sum_{|k|\le s_n}\|X_{n,k}\|^\alpha}$$
$$\times\Big(\mathbb{1}_{\{h\in H_{n,i}\}}\mathbb{1}_A\Big(\frac{X_{t+h+i}}{\|X_{t+h}\|}\Big)+\mathbb{1}_{\{h\in H_{n,i}^C\}}\mathbb{1}_A(0)\Big)\ \Big|\ \|X_0\|>u_n\Bigg]$$
$$\le E\left[\frac{\sum_{|h|\le m}|\log(\|X_{n,h}\|)|\|X_{n,h}\|^\alpha\sum_{m<|k|\le s_n}\|X_{n,k}\|^\alpha}{\sum_{|k|\le m}\|X_{n,k}\|^\alpha\sum_{|k|\le s_n}\|X_{n,k}\|^\alpha}\Big|\|X_0\|>u_n\right]$$
$$+E\left[\frac{\sum_{m<|h|\le s_n}|\log(\|X_{n,h}\|)|\|X_{n,h}\|^\alpha}{\sum_{|k|\le s_n}\|X_{n,k}\|^\alpha}\Big|\|X_0\|>u_n\right]$$
$$=:T_{1,1}+T_{1,2}.$$

The term $T_{1,1}$ can be bounded by the Hölder inequality and in the second step we again use the Hölder inequality for sums (similar as in (5.7.28)):

$$T_{1,1}\le E\left[\left(\frac{\sum_{|h|\le m}|\log(\|X_{n,h}\|)|\|X_{n,h}\|^\alpha}{\sum_{|k|\le m}\|X_{n,k}\|^\alpha}\right)^{1+\delta}\ \Big|\ \|X_0\|>u_n\right]^{1/(1+\delta)}$$
$$\times E\left[\left(\frac{\sum_{m<|k|\le s_n}\|X_{n,k}\|^\alpha}{\sum_{|k|\le s_n}\|X_{n,k}\|^\alpha}\right)^{(1+\delta)/\delta}\ \Big|\ \|X_0\|>u_n\right]^{\delta/(1+\delta)}$$
$$\le E\left[\frac{\sum_{|h|\le m}|\log(\|X_{n,h}\|)|^{1+\delta}\|X_{n,h}\|^\alpha}{\sum_{|k|\le m}\|X_{n,k}\|^\alpha}\ \Big|\ \|X_0\|>u_n\right]^{1/(1+\delta)}$$
$$\times E\left[\frac{\sum_{m<|k|\le s_n}\|X_{n,k}\|^\alpha}{\sum_{|k|\le s_n}\|X_{n,k}\|^\alpha}\ \Big|\ \|X_0\|>u_n\right]^{\delta/(1+\delta)}.$$

By the definition of the tail process and the uniform integrability shown in (5.7.29), the first expectation converges to $E\left[\sum_{|h|\le m}|\log(\|Y_h\|)|^{1+\delta}\|Y_h\|^\alpha(\sum_{|k|\le m}\|Y_k\|^\alpha)^{-1}\right]^{1/(1+\delta)}$ for all $m\in\mathbb{N}$. These expectations are uniformly bounded for all $m\in\mathbb{N}$ by Condition (PM) (ii) and Lemma 5.3.2 part (i). Thus, the first expectation is bounded for $m\to\infty$.
The second expectation is bounded by (5.7.4) for sufficiently large n, which converges to 0 as $n\to\infty$ and then $m\to\infty$ because of condition (PP) and (PT). Thus, we obtain $\lim_{m\to\infty}\limsup_{n\to\infty}T_{1,1}=0$.
To deal with $T_{1,2}$ note that

$$E\left[\frac{\sum_{m<|h|\le s_n}\log^+(\|X_{n,h}\|)\|X_{n,h}\|^\alpha}{\sum_{|k|\le s_n}\|X_{n,k}\|^\alpha}\ \Big|\ \|X_0\|>u_n\right]$$
$$\le E\left[\sup_{m<|h|\le s_n}\log^+(\|X_{n,h}\|)\Big|\|X_0\|>u_n\right]$$

$$\leq \sum_{m<|h|\leq s_n} E\Big[\max(\log^+(\|X_{n,h}\|), \mathbb{1}_{\{\|X_{n,h}\|>1\}}) \max(\log^+(\|X_{n,0}\|), 1) \mid \|X_0\| > u_n\Big]$$
$$\leq 2 \sum_{m<h\leq s_n} e_n'(h)$$

by condition (PP1) and $\lim_{m\to\infty} \limsup_{n\to\infty} \sum_{m<h\leq s_n} e_n'(h) = 0$. For the negative part of the logarithm note that by condition (PM) (iii)

$$\lim_{m\to\infty} \limsup_{n\to\infty} E\left[\frac{\sum_{m<|h|\leq s_n} \log^-(\|X_{n,h}\|)\|X_{n,h}\|^\alpha}{\sum_{|k|\leq s_n} \|X_{n,k}\|^\alpha} \,\middle|\, \|X_0\| > u_n\right] = 0.$$

Thus, $\lim_{m\to\infty} \limsup_{n\to\infty} T_{1,2} = 0$ and, thereby, $\lim_{m\to\infty} \limsup_{n\to\infty} T_1 = 0$.
Next, we want to show the same for T_2. Expanding the fractions to the same numerator, using (5.7.2) twice, in particular $(\sum_{|h|\leq s_n} a_h)^2 - (\sum_{|h|\leq m} a_h)^2 \leq 2\sum_{|h|\leq s_n} a_h \sum_{m<|k|\leq s_n} a_k$, applying the triangular inequality, taking the absolute value of each summand and bounding the indicators by 1 yields

$$\begin{aligned}
&T_2\\
&= E\Bigg[\Bigg|\frac{1}{(\sum_{|k|\leq s_n} \|X_{n,k}\|^\alpha)^2(\sum_{|k|\leq m} \|X_{n,k}\|^\alpha)^2}\\
&\quad\times \Big(\sum_{|h|\leq m} \|X_{n,h}\|^\alpha \sum_{|k|\leq m} \log(\|X_{n,k}\|)\|X_{n,k}\|^\alpha \mathbb{1}_A\Big(\frac{X_{n,h+i}}{\|X_{n,h}\|}\Big)\Big(\sum_{|h|\leq s_n} \|X_{n,h}\|^\alpha\Big)^2\\
&\quad - \sum_{|h|\leq m} \|X_{n,h}\|^\alpha \sum_{|k|\leq m} \log(\|X_{n,k}\|)\|X_{n,k}\|^\alpha \mathbb{1}_A\Big(\frac{X_{n,h+i}}{\|X_{n,h}\|}\Big)\Big(\sum_{|h|\leq m} \|X_{n,h}\|^\alpha\Big)^2\\
&\quad + \sum_{|h|\leq m} \|X_{n,h}\|^\alpha \sum_{|k|\leq m} \log(\|X_{n,k}\|)\|X_{n,k}\|^\alpha \mathbb{1}_A\Big(\frac{X_{n,h+i}}{\|X_{n,h}\|}\Big)\Big(\sum_{|h|\leq m} \|X_{n,h}\|^\alpha\Big)^2\\
&\quad - \sum_{|h|\leq s_n} \|X_{n,h}\|^\alpha \sum_{|k|\leq s_n} \log(\|X_{n,k}\|)\|X_{n,k}\|^\alpha\\
&\qquad \times \Big(\mathbb{1}_{\{h\in H_{n,i}\}}\mathbb{1}_A\Big(\frac{X_{t+h+i}}{\|X_{t+h}\|}\Big) + \mathbb{1}_{\{h\in H_{n,i}^C\}}\mathbb{1}_A(0)\Big)\Big(\sum_{|h|\leq m} \|X_{n,h}\|^\alpha\Big)^2\Big)\Bigg| \,\Bigg|\, \|X_0\| > u_n\Bigg]\\
&= E\Bigg[\Bigg|\frac{1}{(\sum_{|k|\leq s_n} \|X_{n,k}\|^\alpha)^2(\sum_{|k|\leq m} \|X_{n,k}\|^\alpha)^2}\\
&\quad\times \Big(\sum_{|h|\leq m} \|X_{n,h}\|^\alpha \sum_{|k|\leq m} \log(\|X_{n,k}\|)\|X_{n,k}\|^\alpha \mathbb{1}_A\Big(\frac{X_{n,h+i}}{\|X_{n,h}\|}\Big)\\
&\qquad\qquad \times \Big(\Big(\sum_{|h|\leq s_n} \|X_{n,h}\|^\alpha\Big)^2 - \Big(\sum_{|h|\leq m} \|X_{n,h}\|^\alpha\Big)^2\Big)\\
&\quad + \Big(\sum_{|h|\leq m} \|X_{n,h}\|^\alpha \sum_{|k|\leq m} \log(\|X_{n,k}\|)\|X_{n,k}\|^\alpha \mathbb{1}_A\Big(\frac{X_{n,h+i}}{\|X_{n,h}\|}\Big)\\
&\quad - \sum_{|h|\leq s_n} \|X_{n,h}\|^\alpha \sum_{|k|\leq s_n} \log(\|X_{n,k}\|)\|X_{n,k}\|^\alpha\\
&\qquad \times \Big(\mathbb{1}_{\{h\in H_{n,i}\}}\mathbb{1}_A\Big(\frac{X_{t+h+i}}{\|X_{t+h}\|}\Big) + \mathbb{1}_{\{h\in H_{n,i}^C\}}\mathbb{1}_A(0)\Big)\Big)\Big(\sum_{|h|\leq m} \|X_{n,h}\|^\alpha\Big)^2\Big)\Bigg| \,\Bigg|\, \|X_0\| > u_n\Bigg]
\end{aligned}$$

$$
\begin{aligned}
&\leq E\bigg[\frac{1}{(\sum_{|k|\leq s_n}\|X_{n,k}\|^\alpha)^2(\sum_{|k|\leq m}\|X_{n,k}\|^\alpha)^2}\Big(\sum_{|h|\leq m}\|X_{n,h}\|^\alpha\sum_{|k|\leq m}|\log(\|X_{n,k}\|)|\|X_{n,k}\|^\alpha\\
&\qquad\times\Big(\sum_{|h|\leq s_n}\|X_{n,h}\|^\alpha\sum_{m<|k|\leq s_n}\|X_{n,k}\|^\alpha+\sum_{m<|h|\leq s_n}\|X_{n,h}\|^\alpha\sum_{|k|\leq m}\|X_{n,k}\|^\alpha\Big)\\
&\quad+\Big|\sum_{|h|\leq m}\|X_{n,h}\|^\alpha\sum_{|k|\leq m}\log(\|X_{n,k}\|)\|X_{n,k}\|^\alpha\mathbb{1}_A\Big(\frac{X_{n,h+i}}{\|X_{n,h}\|}\Big)\\
&\quad-\sum_{|h|\leq s_n}\|X_{n,h}\|^\alpha\sum_{|k|\leq s_n}\log(\|X_{n,k}\|)\|X_{n,k}\|^\alpha\Big(\mathbb{1}_{\{h\in H_{n,i}\}}\mathbb{1}_A\Big(\frac{X_{t+h+i}}{\|X_{t+h}\|}\Big)+\mathbb{1}_{\{h\in H_{n,i}^C\}}\mathbb{1}_A(0)\Big)\Big|\\
&\qquad\times\Big(\sum_{|h|\leq m}\|X_{n,h}\|^\alpha\Big)^2\Big)\;\Big|\;\|X_0\|>u_n\bigg]\\
&\leq E\bigg[\frac{\sum_{|k|\leq m}|\log(\|X_{n,k}\|)|\|X_{n,k}\|^\alpha 2\sum_{m<|h|\leq s_n}\|X_{n,h}\|^\alpha}{\sum_{|k|\leq m}\|X_{n,k}\|^\alpha\sum_{|k|\leq s_n}\|X_{n,k}\|^\alpha}\;\Big|\;\|X_0\|>u_n\bigg]\\
&\quad+E\bigg[\sum_{|h|\leq s_n}\frac{\|X_{n,h}\|^\alpha\sum_{m<|k|\leq s_n}|\log(\|X_{n,k}\|)|\|X_{n,k}\|^\alpha}{(\sum_{|k|\leq s_n}\|X_{n,k}\|^\alpha)^2}\\
&\qquad\times\Big(\mathbb{1}_{\{h\in H_{n,i}\}}\mathbb{1}_A\Big(\frac{X_{t+h+i}}{\|X_{t+h}\|}\Big)+\mathbb{1}_{\{h\in H_{n,i}^C\}}\mathbb{1}_A(0)\Big)\;\Big|\;\|X_0\|>u_n\bigg]\\
&\quad+E\bigg[\sum_{m<|h|\leq s_n}\frac{\|X_{n,h}\|^\alpha\sum_{|k|\leq m}|\log(\|X_{n,k}\|)|\|X_{n,k}\|^\alpha}{(\sum_{|k|\leq s_n}\|X_{n,k}\|^\alpha)^2}\\
&\qquad\times\Big(\mathbb{1}_{\{h\in H_{n,i}\}}\mathbb{1}_A\Big(\frac{X_{t+h+i}}{\|X_{t+h}\|}\Big)+\mathbb{1}_{\{h\in H_{n,i}^C\}}\mathbb{1}_A(0)\Big)\;\Big|\;\|X_0\|>u_n\bigg]\\
&\leq 2E\bigg[\frac{\sum_{|k|\leq m}|\log(\|X_{n,k}\|)|\|X_{n,k}\|^\alpha\sum_{m<|h|\leq s_n}\|X_{n,h}\|^\alpha}{\sum_{|k|\leq m}\|X_{n,k}\|^\alpha\sum_{|k|\leq s_n}\|X_{n,k}\|^\alpha}\;\Big|\;\|X_0\|>u_n\bigg]\\
&\quad+E\bigg[\frac{\sum_{m<|k|\leq s_n}|\log(\|X_{n,k}\|)|\|X_{n,k}\|^\alpha}{\sum_{|k|\leq s_n}\|X_{n,k}\|^\alpha}\;\Big|\;\|X_0\|>u_n\bigg]\\
&\quad+E\bigg[\frac{\sum_{m<|h|\leq s_n}\|X_{n,h}\|^\alpha\sum_{|k|\leq m}|\log(\|X_{n,k}\|)|\|X_{n,k}\|^\alpha}{(\sum_{|k|\leq s_n}\|X_{n,k}\|^\alpha)^2}\;\Big|\;\|X_0\|>u_n\bigg]\\
&\leq 3E\bigg[\frac{\sum_{|k|\leq m}|\log(\|X_{n,k}\|)|\|X_{n,k}\|^\alpha\sum_{m<|h|\leq s_n}\|X_{n,h}\|^\alpha}{\sum_{|k|\leq m}\|X_{n,k}\|^\alpha\sum_{|k|\leq s_n}\|X_{n,k}\|^\alpha}\;\Big|\;\|X_0\|>u_n\bigg]\\
&\quad+E\bigg[\frac{\sum_{m<|k|\leq s_n}|\log(\|X_{n,k}\|)|\|X_{n,k}\|^\alpha}{\sum_{|k|\leq s_n}\|X_{n,k}\|^\alpha}\;\Big|\;\|X_0\|>u_n\bigg]\\
&=3T_{1,1}+T_{1,2}.
\end{aligned}
$$

Thus, $\lim_{m\to\infty}\limsup_{n\to\infty}T_2=0$ and, therefore,

$$
\lim_{m\to\infty}\limsup_{n\to\infty}E[|f_A^{(m)}(W_{n,0})-f_A(W_{n,0})|\mid\|X_0\|>u_n]\leq\lim_{m\to\infty}\limsup_{n\to\infty}(4T_{1,1}+2T_{1,2})=0.
$$

This concludes the proof of part (i).

For part (ii), first note that by using $Y_0\overset{d}{=}\|Y_0\|\Theta_0$ with Θ_0 and $\|Y_0\|\sim Par(\alpha)$ independent, we obtain

$$
\begin{aligned}
&E[f_A((Y_h)_{h\in\mathbb{Z}})]\\
&= E\left[\sum_{h\in\mathbb{Z}}\left(\frac{\log(\|Y_h\|)\|Y_h\|^\alpha}{\|Y\|_\alpha^\alpha} - \frac{\|Y_h\|^\alpha \sum_{k\in\mathbb{Z}}\log(\|Y_k\|)\|Y_k\|^\alpha}{\|Y\|_\alpha^{2\alpha}}\right)\mathbb{1}_A\left(\frac{Y_{h+i}}{\|Y_h\|}\right)\right]\\
&= E\left[\sum_{h\in\mathbb{Z}}\left(\frac{\log(\|\Theta_h\|\|Y_0\|)\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha} - \frac{\|\Theta_h\|^\alpha \sum_{k\in\mathbb{Z}}\log(\|\Theta_k\|\|Y_0\|)\|\Theta_k\|^\alpha}{\|\Theta\|_\alpha^{2\alpha}}\right)\mathbb{1}_A\left(\frac{\Theta_{h+i}}{\|\Theta_h\|}\right)\right]\\
&= E\left[\sum_{h\in\mathbb{Z}}\left(\frac{\log(\|\Theta_h\|)\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha} + \log(\|Y_0\|)\frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha} - \frac{\|\Theta_h\|^\alpha \sum_{k\in\mathbb{Z}}\log(\|\Theta_k\|)\|\Theta_k\|^\alpha}{\|\Theta\|_\alpha^{2\alpha}}\right.\right.\\
&\qquad\qquad\left.\left. - \log(\|Y_0\|)\frac{\|\Theta_h\|^\alpha \sum_{k\in\mathbb{Z}}\|\Theta_k\|^\alpha}{\|\Theta\|_\alpha^{2\alpha}}\right)\mathbb{1}_A\left(\frac{\Theta_{h+i}}{\|\Theta_h\|}\right)\right]\\
&= E\left[\sum_{h\in\mathbb{Z}}\left(\frac{\log(\|\Theta_h\|)\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha} - \frac{\|\Theta_h\|^\alpha \sum_{k\in\mathbb{Z}}\log(\|\Theta_k\|)\|\Theta_k\|^\alpha}{\|\Theta\|_\alpha^{2\alpha}}\right)\mathbb{1}_A\left(\frac{\Theta_{h+i}}{\|\Theta_h\|}\right)\right]\\
&= E\left[\sum_{h\in\mathbb{Z}}\frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha}\left(\log(\|\Theta_h\|)\frac{\sum_{k\in\mathbb{Z}}\|\Theta_k\|^\alpha}{\|\Theta\|_\alpha^\alpha} - \frac{\sum_{k\in\mathbb{Z}}\log(\|\Theta_k\|)\|\Theta_k\|^\alpha}{\|\Theta\|_\alpha^\alpha}\right)\mathbb{1}_A\left(\frac{\Theta_{h+i}}{\|\Theta_h\|}\right)\right]\\
&= E\left[\sum_{h\in\mathbb{Z}}\frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha}\left(-\frac{\sum_{k\in\mathbb{Z}}\log(\|\Theta_k\|/\|\Theta_h\|)(\|\Theta_k\|/\|\Theta_h\|)^\alpha}{\|\Theta/\|\Theta_h\|\|_\alpha^\alpha}\right)\mathbb{1}_A\left(\frac{\Theta_{h+i}}{\|\Theta_h\|}\right)\right]\\
&= E\left[-\frac{\sum_{k\in\mathbb{Z}}\log(\|\Theta_k\|)\|\Theta_k\|^\alpha}{\|\Theta\|_\alpha^\alpha}\mathbb{1}_A(\Theta_i)\right] = d_A.
\end{aligned}
$$

In the penultimate step we used the RS-transformation for $(\Theta_t)_{t\in\mathbb{Z}}$ and the invariance of the distribution of $(\Theta_t)_{t\in\mathbb{Z}}$ under the RS-transformation.
Thus, it remains to show

$$
\begin{aligned}
\lim_{m\to\infty}|d_A^{(m)} - d_A| &= \lim_{m\to\infty}|E[f_A^{(m)}((Y_h)_{h\in\mathbb{Z}}) - f_A((Y_h)_{h\in\mathbb{Z}})]|\\
&\le \lim_{m\to\infty} E[|f_A^{(m)}((Y_h)_{h\in\mathbb{Z}}) - f_A((Y_h)_{h\in\mathbb{Z}})|] = 0.
\end{aligned}
$$

By exactly the same arguments as leading to $E[f_A^{(m)}(W_{n,0}) - f_A(W_{n,0})| \mid \|X_0\| > u_n] \le (4T_{1,1} + 2T_{1,2})$ in part (i) of this proof, one also achieves

$$
\begin{aligned}
&E[|f_A^{(m)}((Y_h)_{h\in\mathbb{Z}}) - f_A((Y_h)_{h\in\mathbb{Z}})|]\\
&\le 4E\left[\frac{\sum_{|k|\le m}|\log(\|Y_k\|)|\|Y_k\|^\alpha \sum_{|h|>m}\|Y_h\|^\alpha}{\sum_{|k|\le m}\|Y_k\|^\alpha \sum_{k\in\mathbb{Z}}\|Y_k\|^\alpha}\right]\\
&\qquad + 2E\left[\frac{\sum_{|k|>m}|\log(\|Y_k\|)|\|Y_k\|^\alpha}{\sum_{k\in\mathbb{Z}}\|Y_k\|^\alpha}\right]\\
&=: 4T_1^Y + 2T_2^Y.
\end{aligned}
$$

Using the Lemma of Fatou for T_2^Y yields

$$E\Bigg[\frac{\sum_{|k|>m} |\log(\|Y_k\|)|\|Y_k\|^\alpha}{\sum_{k\in\mathbb{Z}} \|Y_k\|^\alpha}\Bigg] = E\Bigg[\lim_{M\to\infty} \frac{\sum_{m<|k|<M} |\log(\|Y_k\|)|\|Y_k\|^\alpha}{\sum_{k\in\mathbb{Z}} \|Y_k\|^\alpha}\Bigg]$$
$$\leq \liminf_{M\to\infty} E\Bigg[\frac{\sum_{m<|k|<M} |\log(\|Y_k\|)|\|Y_k\|^\alpha}{\sum_{k\in\mathbb{Z}} \|Y_k\|^\alpha}\Bigg]$$
$$\leq \limsup_{M\to\infty} E\Bigg[\frac{\sum_{|k|<M} |\log(\|Y_k\|)|\|Y_k\|^\alpha}{\sum_{|k|<M} \|Y_k\|^\alpha}\Bigg] < \infty$$

by Jensen's inequality and Condition (PM) (ii) for all $m \in \mathbb{N}$. Thus, by monotone convergence it directly follows $\lim_{m\to\infty} T_2^Y = 0$. By the Hölder inequality T_1^Y can be bounded similar as before by

$$T_1^Y \leq E\left[\frac{\sum_{|h|\leq m} |\log(\|Y_h\|)|^{1+\delta}\|Y_h\|^\alpha}{\sum_{|k|\leq m} \|Y_k\|^\alpha}\right]^{1/(1+\delta)} E\left[\frac{\sum_{|k|>m} \|Y_k\|^\alpha}{\|Y\|_\alpha^\alpha}\right]^{\delta/(1+\delta)}.$$

Again, by dominated convergence (the fractions are bounded by 1) the second expectation converges to 0 as $m \to \infty$. The first expectation is bounded by condition (PM) (ii) and Lemma 5.3.2 (i), and hence $\lim_{m\to\infty} T_1^Y = 0$. All in all, this proves

$$\lim_{m\to\infty} E|[f_A^{(m)}((Y_h)_{h\in\mathbb{Z}}) - f_A((Y_h)_{h\in\mathbb{Z}})|] \leq \lim_{m\to\infty} 4T_1^Y + 2T_2^Y = 0,$$

which concludes the proof. □

The previous lemmas were concerned with terms occurring in $I(A)$ in the main proof of Theorem 5.3.1. The next two lemmas deal with $II(A)$.

Lemma 5.7.5. *Let $m \in \mathbb{N}$, $a = (a_1, ..., a_m) \in [0,1]^m$ and set $0\log(0) = 0\log^2(0) := 0$.*

(i) The function $M : [0,1]^m \to \mathbb{R}$ with

$$M(a) := \frac{\sum_{k=1}^m \log^2(a_k)a_k}{1 + \sum_{k=1}^m a_k}$$

is bounded by $\sup_{a\in[0,1]^m} M(a) = O\left(\log^2(m)\right)$ as $m \to \infty$.

(ii) The function $M_1 : [0,1]^m \to \mathbb{R}$ with

$$M_1(a) = \frac{\sum_{k=1}^m |\log(a_k)|a_k}{1 + \sum_{k=1}^m a_k}$$

is bounded by $\sup_{a\in[0,1]^m} M_1(a) = O(\log(m))$ as $m \to \infty$.

Proof. It is easy to see that the function $\mathbb{R} \to \mathbb{R}$, $x \mapsto \log^2(x)x$ has a local maximum in $(0,1)$ at $x = e^{-2}$ and is decreasing on $[e^{-2}, 1]$. Therefore, if $a_h > e^{-2}$ for some $h \in \{1, ..., m\}$, then replacing a_h by e^{-2} increases the numerator of $M(a)$ and decreases the

denominator of $M(a)$ and thus increases $M(a)$. Hence, any point of maximum of M must belong to $[0, e^{-2}]^m$.
The partial derivatives of M are given by

$$\frac{\partial}{\partial a_h} M(a) = \frac{\log^2(a_h) + 2\log(a_h)}{1 + \sum_{k=1}^m a_k} - \frac{\sum_{k=1}^m \log^2(a_k) a_k}{(1 + \sum_{k=1}^m a_k)^2}$$

for all $h = 1, ..., m$. It directly holds $\partial/(\partial a_h) M(a) \to \infty$ as $a_h \downarrow 0$, since $\log(a_h)^2 \to \infty$. Hence, any point of maximum of M must belong to $(0, e^{-2}]^m$. Since the point of maximum is in the interior of $[0,1]^m$ the partial derivatives of M must vanish at the point of maximum, i.e. the point of maximum solves the system of equations

$$\frac{\partial}{\partial a_h} M(a) = 0, \quad \forall h \in \{1, ..., m\}. \tag{5.7.34}$$

This leads to $\partial/(\partial a_h) M = \partial/(\partial a_1) M$ for all $h \in \{1, ..., m\}$, which yields

$$\log^2(a_h) + 2\log(a_h) - \frac{\sum_{k=1}^m \log^2(a_k) a_k}{1 + \sum_{k=1}^m a_k} = \log^2(a_1) + 2\log(a_1) - \frac{\sum_{k=1}^m \log^2(a_k) a_k}{1 + \sum_{k=1}^m a_k}$$

which is equivalent to

$$\log^2(a_h) + 2\log(a_h) = \log^2(a_1) + 2\log(a_1),$$

which in turn is equivalent to

$$(\log(a_h) + 1)^2 = (\log(a_1) + 1)^2.$$

This equation has the solutions $a_h = a_1$ or $a_h = \exp(-\log(a_1) - 2) = e^{-2}/a_1$. In order to fulfill the restriction $a_h < 1$, the second solution is only a solution for our problem if $a_1 > e^{-2}$. The equation $a_h = e^{-2}/a_1$ is equivalent to $a_1 = e^{-2}/a_h$, i.e. also $a_h > e^{-2}$ is a necessary condition for $a_1 < 1$ for all $h = 1, ..., m$. This contradicts the above result that the point of maximum satisfies $a_1, a_h \leq e^{-2}$. Thus, for the maximum of M the system of equations (5.7.34) leads to $a_1 = a_h$ for all $h \in \{1, ..., m\}$, i.e. all coordinates of the point of maximum must be the same. Inserting this in (5.7.34) leads to the simplified problem

$$\begin{aligned}
&\frac{\log^2(a_1) + 2\log(a_1)}{1 + ma_1} - \frac{m\log^2(a_1) a_1}{(1 + ma_1)^2} = 0 \\
&\Leftrightarrow (\log^2(a_1) + 2\log(a_1))(1 + ma_1) - m\log^2(a_1) a_1 = 0 \\
&\Leftrightarrow \log^2(a_1) + 2\log(a_1) + m\log^2(a_1) a_1 + 2\log(a_1) m a_1 - m\log^2(a_1) a_1 = 0 \\
&\Leftrightarrow |\log(a_1)| = 2(1 + ma_1).
\end{aligned} \tag{5.7.35}$$

By the mean value theorem and by monotonicity, this equation has a unique solution a_1^*. To check that this solution indeed maximizes $M = m\log^2(a_1) a_1/(1 + ma_1)$ one could observe that $M \to 0$ as $a_1 \to 0$ or $a_1 \to 1$. Thus, a_1^* must be the point of maximum. Due to the form of (5.7.35), the solution a_1^* of (5.7.35) is of smaller order than $m^{\delta-1}$ and of larger order than $m^{-\delta-1}$ as $m \to \infty$ for all $\delta > 0$. To check this simply insert $m^{\delta-1}$ and $m^{-\delta-1}$ in (5.7.35) and consider the order of right and left hand side of the equation as $m \to \infty$, note that $\log(m) = \mathrm{o}(m^{\delta})$ and $m^{-\delta} = \mathrm{o}(\log(m))$. In particular it follows $|\log(a_1^*)| \sim \log(m)$ and thus

$$a_1^* \sim \frac{1}{2}\frac{\log(m)}{m}$$

as $m \to \infty$. Therefore, we can bound M by

$$\begin{aligned}\sup_{a\in[0,1]^m} M(a) \leq M(a_1^*, ..., a_1^*) &= \frac{m\log^2(a_1^*)a_1^*}{1+ma_1^*} = \frac{m4(1+ma_1^*)^2 a_1^*}{1+ma_1^*} = 4ma_1^*(1+ma_1^*)\\ &\sim 2\log(m)(1+\frac{1}{2}\log(m)) \sim \log^2(m)\end{aligned}$$

as $m \to \infty$.

Next we turn to part (ii). The arguments are along the same lines as for part (i). The partial derivatives of M_1 are given by

$$\frac{\partial}{\partial a_h}M_1(a) = \frac{-\log(a_h)-1}{1+\sum_{k=1}^m a_k} + \frac{\sum_{k=1}^m |\log(a_k)|a_k}{(1+\sum_{k=1}^m a_k)^2}$$

for all $h = 1, ..., m$. Here again one has $\partial/(\partial a_h)M_1(a) \to \infty$ as $a_h \to 0$ and $|\log(x)|x \to 0$ as $x \to 1$. Moreover, $x \mapsto x|\log(x)|$ has a local maximum at $x = e^{-1}$. Therefore, the point of maximum of M_1 belongs to $(0, e^{-1}]^m$. This point of maximum solves the system of equations $\partial/(\partial a_h)M_1(a) = 0$ for all $h \in \{1, ..., m\}$. Note that for $a_h < 1$ one has $|\log(a_h)| = -\log(a_h)$. Equating the equation for a_h and a_1 yields

$$-\log(a_h) - 1 = -\log(a_1) - 1,$$

i.e. $a_h = a_1$ for all $h = 1, ..., m$. Inserting this in the equation for $h = 1$ yields

$$\begin{aligned}&-\log(a_1) - 1 - \frac{m\log(a_1)a_1}{1+ma_1} = 0\\ &\Leftrightarrow -\log(a_1) - 1 - m\log(a_1)a_1 - ma_1 + m\log(a_1)a_1 = 0\\ &\Leftrightarrow |\log(a_1)| = -\log(a_1) = 1 + ma_1. \qquad (5.7.36)\end{aligned}$$

This equation has a unique solution $a_1^* \in (0, 1)$, which follows from the mean value theorem and monotonicity. To check that this solution is indeed a point of maximum, note that $M_1(a_1, ..., a_1) = m|\log(a_1)|a_1/(1 + ma_1)$ converges to 0 as $a_1 \to 0$ or $a_1 \to 1$. Due to the form of (5.7.36), the solution of (5.7.36) is of smaller order than $m^{\delta-1}$ and of larger order than $m^{-\delta-1}$ as $m \to \infty$ for all $\delta > 0$. In particular, $|\log(a_1^*)| \sim \log(m)$ and a_1^* behaves like

$$a_1^* \sim \frac{\log(m)}{m}$$

as $m \to \infty$. Hence, an upper bound of M_1 is

$$\sup_{a\in[0,1]^m} M_1(a) \leq M_1(a_1^*, ..., a_1^*) = \frac{m|\log(a_1^*)|a_1^*}{1+ma_1^*} = \frac{m(1+ma_1^*)a_1^*}{1+ma_1^*} = ma_1^* \sim \log(m)$$

as $m \to \infty$. □

Lemma 5.7.6. *Suppose that the conditions (PR), (P0), (PP), (PC), (PP1), (PB$_\alpha$), and (PM) (i) are satisfied and* $\log(n)^4 = o(nv_n)$ *holds. Then, the term* $II(A)$ *defined in* (5.7.18) *is uniform negligible, i.e.*

$$\sup_{A\in\mathcal{A}} \sqrt{nv_n} II(A) = o_P(1).$$

Proof. Note that $2^{-1}nv_n(\hat{\alpha}_n - \alpha)^2 \to 2^{-1}Z_\alpha^2$ weakly by (5.7.17) and the continuous mapping theorem. Furthermore, $nv_n/(\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}) \to 1$ in probability due to the weak convergence in Proposition 5.2.7 for the set $\mathbb{R}^d$. Therefore, it remains to check that the term

$$\begin{aligned}
&\sqrt{nv_n}\frac{2\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}}{nv_n(nv_n(\hat{\alpha}_n-\alpha)^2)}|II(A)| \\
&= \frac{1}{(nv_n)^{3/2}}\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}\sum_{h=-s_n}^{s_n}\Bigg(\frac{\log^2(\|X_{n,t+h}\|)\|X_{n,t+h}\|^{\bar{\alpha}}}{\sum_{k=-s_n}^{s_n}\|X_{n,t+k}\|^{\bar{\alpha}}} \\
&\qquad -2\frac{\log(\|X_{n,t+h}\|)\|X_{n,t+h}\|^{\bar{\alpha}}\sum_{k=-s_n}^{s_n}\log(\|X_{n,t+k}\|)\|X_{n,t+k}\|^{\bar{\alpha}}}{(\sum_{k=-s_n}^{s_n}\|X_{n,t+k}\|^{\bar{\alpha}})^2} \\
&\qquad -\frac{\|X_{n,t+h}\|^{\bar{\alpha}}\sum_{k=-s_n}^{s_n}\log^2(\|X_{n,t+k}\|)\|X_{n,t+k}\|^{\bar{\alpha}}}{(\sum_{k=-s_n}^{s_n}\|X_{n,t+k}\|^{\bar{\alpha}})^2} \\
&\qquad +2\,\frac{\|X_{n,t+h}\|^{\bar{\alpha}}(\sum_{k=-s_n}^{s_n}\log(\|X_{n,t+k}\|)\|X_{n,t+k}\|^{\bar{\alpha}})^2}{(\sum_{k=-s_n}^{s_n}\|X_{n,t+k}\|^{\bar{\alpha}})^3}\Bigg) \\
&\qquad \times\left(\mathbb{1}_{\{h\in H_{n,i}\}}\mathbb{1}_A\left(\frac{X_{t+h+i}}{\|X_{t+h}\|}\right)+\mathbb{1}_{\{h\in H_{n,i}^C\}}\mathbb{1}_A(0)\right)
\end{aligned} \tag{5.7.37}$$

converges to 0 in probability uniformly for all $A \in \mathcal{A}$.

The absolute value of (5.7.37) can be bound from above by taking the absolute value of each summand and bounding the sum of the indicators with 1. Involving this, it now results in the upper bound

$$\begin{aligned}
&\frac{1}{(nv_n)^{3/2}}\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}\sum_{h=-s_n}^{s_n}\Bigg(\frac{\log^2(\|X_{n,t+h}\|)\|X_{n,t+h}\|^{\bar{\alpha}}}{\sum_{k=-s_n}^{s_n}\|X_{n,t+k}\|^{\bar{\alpha}}} \\
&\qquad +2\frac{\log(\|X_{n,t+h}\|)\|X_{n,t+h}\|^{\bar{\alpha}}\sum_{k=-s_n}^{s_n}\log(\|X_{n,t+k}\|)\|X_{n,t+k}\|^{\bar{\alpha}}}{(\sum_{k=-s_n}^{s_n}\|X_{n,t+k}\|^{\bar{\alpha}})^2}
\end{aligned}$$

$$
\begin{aligned}
&\quad + \frac{\|X_{n,t+h}\|^{\bar\alpha} \sum_{k=-s_n}^{s_n} \log^2(\|X_{n,t+k}\|)\|X_{n,t+k}\|^{\bar\alpha}}{(\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar\alpha})^2} \\
&\quad +2 \; \frac{\|X_{n,t+h}\|^{\bar\alpha}(\sum_{k=-s_n}^{s_n} \log(\|X_{n,t+k}\|)\|X_{n,t+k}\|^{\bar\alpha})^2}{(\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar\alpha})^3}\Bigg) \\
&= \frac{1}{(nv_n)^{3/2}} \sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}} 2 \sum_{h=-s_n}^{s_n} \frac{\log^2(\|X_{n,t+h}\|)\|X_{n,t+h}\|^{\bar\alpha}}{\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar\alpha}} \\
&\quad + \frac{1}{(nv_n)^{3/2}} \sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}} 4 \frac{(\sum_{h=-s_n}^{s_n} |\log(\|X_{n,t+h}\|)|\|X_{n,t+h}\|^{\bar\alpha})^2}{(\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar\alpha})^2} =: 2T_1 + 4T_2. \qquad (5.7.38)
\end{aligned}
$$

Thus, it suffices to show, that this expression converges to 0. Since this is an upper bound independent of A, this would imply the uniform convergence to 0 of (5.7.37). To show the convergence to 0 of (5.7.38) we split both sums in two summands, which we consider individually. More specific, we will distinguish between the cases whether $\|X_{t+h}\|$ exceeds u_n or not, i.e. we consider the sums individually on the sets $\{\|X_{t+h}\| \geq u_n\}$ and $\{\|X_{t+h}\| < u_n\}$.
Define the set $B_n := \left\{\max_{1\leq t\leq n} \|X_{n,t}\| > n^{2/\alpha}\right\}$. Using the regular variation of $\|X_0\|$ to bound the survival function yields for all $\varepsilon \in (0, 1/2)$ and sufficiently large n

$$
\begin{aligned}
P(B_n) &\leq nP\left(\|X_{n,0}\| > n^{2/\alpha}\right) \leq nP\left(\|X_0\| > n^{2/\alpha}\right) = \mathrm{o}\left(n(n^{2/\alpha})^{-\alpha(1-\varepsilon)}\right) \\
&= \mathrm{o}\left(n^{-1+2\varepsilon}\right) = \mathrm{o}(1).
\end{aligned}
$$

On the set B_n^C it holds

$$
0 \leq \log\left(\|X_{n,t}\|\right) \mathbb{1}_{\{\|X_t\|\geq u_n\}} \leq \frac{2}{\alpha}\log(n)
$$

for all $1 \leq t \leq n$. Thus, on B_n^C (which hold with probability tending to 1) it holds

$$
\begin{aligned}
&\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h=-s_n}^{s_n} \frac{\log^2(\|X_{n,t+h}\|)\|X_{n,t+h}\|^{\bar\alpha}\mathbb{1}_{\{\|X_{t+h}\|>u_n\}}}{\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar\alpha}} \\
&\leq \sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}} \frac{4}{\alpha^2}\log^2(n) \frac{\sum_{h=-s_n}^{s_n} \|X_{n,t+h}\|^{\bar\alpha}}{\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar\alpha}} \\
&= \frac{4}{\alpha^2}\log^2(n) \sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}} = \sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}} \mathrm{o}(\sqrt{nv_n}) = \mathrm{o}_P((nv_n)^{3/2}) \qquad (5.7.39)
\end{aligned}
$$

since $\log^2(n) = \mathrm{o}(\sqrt{nv_n})$ by assumption and $(nv_n)^{-1} \sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}} = \mathrm{O}_P(1)$ due to Proposition 5.2.7 with $\mathbb{R}^d$. Similarly,

$$
\begin{aligned}
&\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}} \frac{(\sum_{k=-s_n}^{s_n} |\log(\|X_{n,t+k}\|)|\|X_{n,t+k}\|^{\bar\alpha}\mathbb{1}_{\{\|X_{t+k}\|\geq u_n\}})^2}{(\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar\alpha})^2} \\
&\leq \sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}} \frac{(2\log(n)/\alpha \sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar\alpha})^2}{(\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar\alpha})^2}
\end{aligned}
$$

$$= \frac{4}{\alpha^2} \log^2(n) \sum_{t=1}^{n} \mathbb{1}_{\{\|X_t\|>u_n\}} = \sum_{t=1}^{n} \mathbb{1}_{\{\|X_t\|>u_n\}} \mathrm{o}(\sqrt{nv_n}) = \mathrm{o}_P((nv_n)^{3/2}). \tag{5.7.40}$$

Thus, (5.7.38) with the additional indicator $\mathbb{1}_{\{\|X_{t+h}\|\geq u_n\}}$ in the numerator inserted is of order $\mathrm{o}_P(1)$.

In the next step, we consider the same terms as before with the additional indicator $\mathbb{1}_{\{\|X_{t+k}\|<u_n\}}$. Again we start with the first summand T_1 of (5.7.38). One has

$$\begin{aligned}
&\sum_{t=1}^{n} \mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h=-s_n}^{s_n} \frac{\log^2(\|X_{n,t+h}\|)\|X_{n,t+h}\|^{\bar{\alpha}} \mathbb{1}_{\{\|X_{t+h}\|<u_n\}}}{\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar{\alpha}}} \\
&\leq \sum_{t=1}^{n} \mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h=-s_n,h\neq 0}^{s_n} \frac{\log^2(\|X_{n,t+h}\|)\|X_{n,t+h}\|^{\bar{\alpha}} \mathbb{1}_{\{\|X_{t+h}\|<u_n\}}}{1+\sum_{k=-s_n,k\neq 0}^{s_n} \|X_{n,t+k}\|^{\bar{\alpha}}}.
\end{aligned}$$

This holds, since for $k=0$ the summand in the denominator is greater or equal to 1, due to $\mathbb{1}_{\{\|X_t\|>u_n\}}$, and the summand for $h=0$ in the second sum is 0. If $\bar{\alpha}>0$ the problem to bound this term is related to the maximization of the function $M:[0,1]^m \to \mathbb{R}$ with

$$M(a_1,...,a_m) = \frac{\sum_{k=1}^{m} \log^2(a_k)a_k}{1+\sum_{k=1}^{m} a_k}$$

under the restriction $0 \leq a_k < 1$ for all $k=1,...,m$ and for some $m \in \mathbb{N}$. Note that $\alpha > 0$ and $\hat{\alpha} > 0$ because of the definition of $\hat{\alpha}$, i.e. it holds $\bar{\alpha} > 0$.

An application of the upper bound of M established in Lemma 5.7.5, part (i), with $a_k = \|X_{n,t+k}\|^{\bar{\alpha}}$ yields

$$\begin{aligned}
&\sum_{t=1}^{n} \mathbb{1}_{\{\|X_t\|>u_n\}} \sum_{h=-s_n,h\neq 0}^{s_n} \frac{\log^2(\|X_{n,t+h}\|)\|X_{n,t+h}\|^{\bar{\alpha}} \mathbb{1}_{\{\|X_{t+h}\|<u_n\}}}{1+\sum_{k=-s_n,k\neq 0}^{s_n} \|X_{n,t+k}\|^{\bar{\alpha}}} \\
&= \sum_{t=1}^{n} \mathbb{1}_{\{\|X_t\|>u_n\}} \frac{1}{\bar{\alpha}^2} \sum_{h=-s_n,h\neq 0}^{s_n} \frac{\log^2(\|X_{n,t+h}\|^{\bar{\alpha}})\|X_{n,t+h}\|^{\bar{\alpha}} \mathbb{1}_{\{\|X_{t+h}\|<u_n\}}}{1+\sum_{k=-s_n,k\neq 0}^{s_n} \|X_{n,t+k}\|^{\bar{\alpha}}} \\
&= \sum_{t=1}^{n} \mathbb{1}_{\{\|X_t\|>u_n\}} \mathrm{O}(\log^2(2s_n)) \\
&= \sum_{t=1}^{n} \mathbb{1}_{\{\|X_t\|>u_n\}} \mathrm{O}(\log^2(n)) = \sum_{t=1}^{n} \mathbb{1}_{\{\|X_t\|>u_n\}} \mathrm{o}(\sqrt{nv_n}) = \mathrm{o}_P((nv_n)^{3/2}).
\end{aligned}$$

This last bound together with (5.7.39) implies

$$\begin{aligned}
T_1 &= \frac{1}{(nv_n)^{3/2}} \sum_{t=1}^{n} \mathbb{1}_{\{\|X_t\|>u_n\}} \left[\sum_{h=-s_n}^{s_n} \frac{\log^2(\|X_{n,t+h}\|)\|X_{n,t+h}\|^{\bar{\alpha}} \mathbb{1}_{\{\|X_{t+h}\|>u_n\}}}{\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar{\alpha}}} \right] \\
&\quad + \frac{1}{(nv_n)^{3/2}} \sum_{t=1}^{n} \mathbb{1}_{\{\|X_t\|>u_n\}} \left[\sum_{h=-s_n}^{s_n} \frac{\log^2(\|X_{n,t+h}\|)\|X_{n,t+h}\|^{\bar{\alpha}} \mathbb{1}_{\{\|X_{t+h}\|\leq u_n\}}}{\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar{\alpha}}} \right] \\
&= \frac{1}{(nv_n)^{3/2}} \mathrm{o}_P((nv_n)^{3/2}) = \mathrm{o}_P(1).
\end{aligned} \tag{5.7.41}$$

Now, we will establish an upper bound of the same order for T_2. Using the upper bound established in Lemma 5.7.5, part (ii), with $a_k = \|X_{n,t+k}\|^{\bar{\alpha}}$ yields

$$
\begin{aligned}
&\sum_{t=1}^{n} \mathbb{1}_{\{\|X_t\|>u_n\}} \left(\frac{\sum_{k=-s_n}^{s_n} |\log(\|X_{n,t+k}\|)| \|X_{n,t+k}\|^{\bar{\alpha}} \mathbb{1}_{\{\|X_{t+k}\|<u_n\}}}{\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar{\alpha}}} \right)^2 \\
&\leq \sum_{t=1}^{n} \mathbb{1}_{\{\|X_t\|>u_n\}} \left(\frac{\sum_{k=-s_n,k\neq 0}^{s_n} |\log(\|X_{n,t+k}\|)| \|X_{n,t+k}\|^{\bar{\alpha}} \mathbb{1}_{\{\|X_{t+k}\|<u_n\}}}{1+\sum_{k=-s_n,k\neq 0}^{s_n} \|X_{n,t+k}\|^{\bar{\alpha}}} \right)^2 \\
&= \sum_{t=1}^{n} \mathbb{1}_{\{\|X_t\|>u_n\}} \frac{1}{\bar{\alpha}^2} \left(\frac{\sum_{k=-s_n,k\neq 0}^{s_n} |\log(\|X_{n,t+k}\|^{\bar{\alpha}})| \|X_{n,t+k}\|^{\bar{\alpha}} \mathbb{1}_{\{\|X_{t+k}\|<u_n\}}}{1+\sum_{k=-s_n,k\neq 0}^{s_n} \|X_{n,t+k}\|^{\bar{\alpha}}} \right)^2 \\
&= \sum_{t=1}^{n} \mathbb{1}_{\{\|X_t\|>u_n\}} \mathrm{O}(\log(2s_n))^2 = \sum_{t=1}^{n} \mathbb{1}_{\{\|X_t\|>u_n\}} \mathrm{o}(\sqrt{nv_n}) = \mathrm{o}_P((nv_n)^{3/2}).
\end{aligned}
$$

By combining this last bound with (5.7.40) we conclude

$$
\begin{aligned}
T_2 &\leq \frac{2}{(nv_n)^{3/2}} \sum_{t=1}^{n} \mathbb{1}_{\{\|X_t\|>u_n\}} \frac{(\sum_{h=-s_n}^{s_n} |\log(\|X_{n,t+h}\|)| \|X_{n,t+h}\|^{\bar{\alpha}} \mathbb{1}_{\{\|X_{t+h}\|<u_n\}})^2}{(\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar{\alpha}})^2} \\
&\quad + \frac{2}{(nv_n)^{3/2}} \sum_{t=1}^{n} \mathbb{1}_{\{\|X_t\|>u_n\}} \frac{(\sum_{h=-s_n}^{s_n} |\log(\|X_{n,t+h}\|)| \|X_{n,t+h}\|^{\bar{\alpha}} \mathbb{1}_{\{\|X_{t+h}\|\geq u_n\}})^2}{(\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar{\alpha}})^2} \\
&= \frac{1}{(nv_n)^{3/2}} \mathrm{o}_P((nv_n)^{3/2}) = \mathrm{o}_P(1). \qquad (5.7.42)
\end{aligned}
$$

The convergences (5.7.41) and (5.7.42) show that (5.7.38) converges to 0 in probability, and, therefore, (5.7.37) converges to 0 in probability uniformly for all $A \in \mathcal{A}$. This concludes the proof. □

In the last lemma of this section the covariances between $(Z(A))_{A\in\mathcal{A}}$ and $Z(\phi)$ as well as the variance of $Z(\phi)$ are calculated. This lemma proves parts of condition (C) for the application of Theorem 3.1.10 in the proof of Theorem 5.3.1. Note that $m_n/(p_n b_n(g_A)^2) = m_n/(p_n b_n(\phi)^2) = m_n/(nv_n) \asymp 1/(r_n v_n)$, which is why the covariance in the following lemma is standardized with $(r_n v_n)^{-1}$.

Lemma 5.7.7. *Suppose the conditions (PR), (P0), (PP), (PT), (PC) and, (PP1) are satisfied. Then,*

(i) $$\frac{1}{r_n v_n} Var\Big(\sum_{t=1}^{r_n} \phi(W_{n,t}) \Big) = \alpha^{-1} \sum_{k\in\mathbb{Z}} E\left[(1 \wedge \|\Theta_k\|^{\alpha})(|\log(\|\Theta_k\|)| + 2\alpha^{-1}) \right],$$

(ii) for all $A \in \mathcal{A}$

$$
\begin{aligned}
&\frac{1}{r_n v_n} Cov\Big(\sum_{j=1}^{r_n} \phi(W_{n,j}), \sum_{t=1}^{r_n} g_A(W_{n,t}) \Big) \\
&\quad = \sum_{k\in\mathbb{Z}} E\Bigg[\sum_{h\in\mathbb{Z}} \frac{\|\Theta_h\|^{\alpha}}{\|\Theta\|_{\alpha}^{\alpha}} \mathbb{1}_A\Big(\frac{\Theta_{h+i}}{\|\Theta_h\|} \Big) (1 \wedge \|\Theta_k\|^{\alpha})(\log(\|\Theta_k\| \vee 1) + \alpha^{-1}) \Bigg].
\end{aligned}
$$

Proof. The proof uses analogous arguments as used in the proof of Lemma 5.2.4. We start with the proof of (i).
In the third step we will apply $E\Big[\log^+(\|X_{n,0}\|) \mid \|X_0\| > u_n\Big] = \mathrm{O}(1)$ which holds due to regular variation and (5.5.6) (see also Kulik and Soulier (2020), Section 2.3.3). By stationarity we obtain

$$\begin{aligned}
&\frac{1}{r_n v_n} Var\left(\sum_{t=1}^{r_n} \phi(W_{n,t})\right) \\
&= \frac{1}{r_n v_n} E\left[\sum_{t=1}^{r_n} \phi(W_{n,t}) \sum_{j=1}^{r_n} \phi(W_{n,j})\right] - \frac{1}{r_n v_n} E\left[\sum_{j=1}^{r_n} \phi(W_{n,j})\right]^2 \\
&= \frac{1}{v_n} \sum_{k=-r_n+1}^{r_n-1} \Big(1 - \frac{|k|}{r_n}\Big) E\left[\phi(W_{n,0})\phi(W_{n,k})\right] \\
&\qquad + \frac{(r_n v_n)^2}{r_n v_n} \Big(E\Big[\log^+(\|X_{n,0}\|) \mid \|X_0\| > u_n\Big]\Big)^2 \\
&= \sum_{k=-r_n+1}^{r_n-1} \Big(1 - \frac{|k|}{r_n}\Big) E\Big[\log^+(\|X_{n,0}\|) \log^+(\|X_{n,k}\|) \mid \|X_0\| > u_n\Big] + \mathrm{O}(r_n v_n) \\
&\to \sum_{k \in \mathbb{Z}} \lim_{n\to\infty} \Big(1 - \frac{|k|}{r_n}\Big) E\Big[\log^+(\|X_{n,0}\|) \log^+(\|X_{n,k}\|) \mid \|X_0\| > u_n\Big],
\end{aligned}$$

where we applied Pratt's Lemma in the last step, which allows the interchange of limes and sum. This lemma can be applied due to condition (PP1) and

$$\begin{aligned}
&\Big(1 - \frac{|k|}{r_n}\Big) E\Big[\log^+(\|X_{n,0}\|) \log^+(\|X_{n,k}\|) \mid \|X_0\| > u_n\Big] \\
&\le E\Big[\max(\log^+(\|X_{n,0}\|), 1) \max(\log^+\Big(\|X_{n,|k|}\|\Big), \mathbb{1}_{\{\|X_{|k|}\| > u_n\}}) \mid \|X_0\| > u_n\Big] \\
&\le e_n'(|k|).
\end{aligned}$$

The definition of the tail process together with the continuous mapping theorem imply the weak convergence

$$\mathcal{L}\left(\log^+\Big(\frac{\|X_0\|}{u_n}\Big) \log^+\Big(\frac{\|X_k\|}{u_n}\Big) \Big| \|X_0\| > u_n\right) \to \mathcal{L}(\log^+(\|Y_0\|) \log^+(\|Y_k\|))$$

for all $k \in \mathbb{Z}$. From the Cauchy Schwartz inequality and (5.5.6) we conclude

$$E[\|X_{n,k}\|^q \|X_{n,0}\|^q \mid \|X_0\| > u_n] = \mathrm{O}(1)$$

for all $k \in \mathbb{Z}$ and $q \in (0, \alpha/2)$. Since $\log^+(x)^{1+\delta} \le c_q x^q$ for some $c_q > 0$, this implies

$$\sup_{n \in \mathbb{N}} E\left[\Big(\log^+(\|X_{n,0}\|) \log^+(\|X_{n,k}\|)\Big)^{1+\delta} \Big| \|X_0\| > u_n\right] < \infty.$$

This uniform moment bound in turn implies uniform integrability of random variables with the distribution of $\log^+(\|X_{n,0}\|)\log^+(\|X_{n,k}\|)$ given $\|X_0\| > u_n$, $n \in \mathbb{N}$. Thus, the weak convergence stated before implies

$$E\left[\log^+(\|X_{n,0}\|)\log^+(\|X_{n,k}\|) \mid \|X_0\| > u_n\right] \to E\left[\log^+(Y_0)\log^+(Y_k)\right]$$

for all $k \in \mathbb{Z}$. This shows

$$\begin{aligned}
&\sum_{k\in\mathbb{Z}} \lim_{n\to\infty}\left(1-\frac{|k|}{r_n}\right)E\left[\log^+(\|X_{n,0}\|)\log^+(\|X_{n,k}\|) \mid \|X_0\| > u_n\right] \\
&= \sum_{k\in\mathbb{Z}} E\left[\log^+(\|Y_0\|)\log^+(\|Y_k\|)\right].
\end{aligned}$$

Moreover,

$$\begin{aligned}
&E\left[\log^+(\|Y_0\|)\log^+(\|Y_k\|)\right] \\
&= E\left[\log^+(\|Y_0\|)\log^+(\|\Theta_k\|\|Y_0\|)\right] \\
&= \int\int_1^\infty \log(y)\log(\|\theta_k\|y)\mathbb{1}_{\{y>\|\theta_k\|^{-1}\}}P^{\|Y_0\|}(dy)P^{\Theta}(d\theta) \\
&= \int\int_{1\vee\|\theta_k\|^{-1}}^\infty \log(y)\log(\|\theta_k\|y)\alpha y^{-\alpha-1}dyP^{\Theta}(d\theta) \\
&= \int\left[-\alpha^{-2}y^{-\alpha}(\alpha\log(\|\theta_k\|y)+\alpha\log(y)(\alpha\log(\|\Theta_k\|y)+1)+2)\right]_{y=1\vee\|\theta_k\|^{-1}}^\infty P^{\Theta}(d\theta) \\
&= E\left[\alpha^{-2}(1\wedge\|\Theta_k\|^\alpha)(\alpha\log(\|\Theta_k\|\vee 1)+\alpha\log(1\vee\|\Theta_k\|^{-1})(\alpha\log(1\wedge\|\Theta_k\|)+1)+2)\right] \\
&= \alpha^{-1}E\left[(1\wedge\|\Theta_k\|^\alpha)(\log(\|\Theta_k\|\vee 1)+\log(1\vee\|\Theta_k\|^{-1})\right. \\
&\qquad \left.+\alpha\log(1\vee\|\Theta_k\|^{-1})\alpha\log(1\wedge\|\Theta_k\|)+2\alpha^{-1})\right] \\
&= \alpha^{-1}E\left[(1\wedge\|\Theta_k\|^\alpha)(\log((\|\Theta_k\|\vee 1)\cdot(1\vee\|\Theta_k\|^{-1}))+2\alpha^{-1})\right] \\
&= \alpha^{-1}E\left[(1\wedge\|\Theta_k\|^\alpha)(|\log(\|\Theta_k\|)|+2\alpha^{-1})\right],
\end{aligned}$$

where we applied $Y_k = \Theta_k\|Y_0\|$ with Θ_k and $\|Y_0\|$independent and $\|Y_0\|$ is Par(α)-distributed, i.e. $P(\|Y_0\| > y) = y^{-\alpha}\wedge 1$. This proves the assertion (i).
For part (ii) first note that $E[g_A(W_{n,0})] \le r_n v_n$ and $E\left[\log^+(\|X_{n,0}\|) \mid \|X_0\| > u_n\right] = \mathrm{O}(1)$. By stationarity

$$\begin{aligned}
&\frac{1}{r_n v_n}Cov\left(\sum_{j=1}^{r_n}\phi(W_{n,j}),\sum_{t=1}^{r_n}g_A(W_{n,t})\right) \\
&= \frac{1}{r_n v_n}E\left[\sum_{t=1}^{r_n}g_A(W_{n,t})\sum_{j=1}^{r_n}\phi(W_{n,j})\right] - \frac{1}{r_n v_n}E\left[\sum_{j=1}^{r_n}\phi(W_{n,j})\right]E\left[\sum_{t=1}^{r_n}g_A(W_{n,t})\right] \\
&= \frac{1}{v_n}\sum_{k=-r_n+1}^{r_n-1}\left(1-\frac{|k|}{r_n}\right)E\left[\phi(W_{n,k})g_A(W_{n,0})\right] \\
&\quad - \frac{r_n v_n}{r_n v_n}E\left[\log^+(\|X_{n,0}\|) \mid \|X_0\| > u_n\right]E\left[g_A(W_{n,0})\right]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=-r_n+1}^{r_n-1} \Big(1 - \frac{|k|}{r_n}\Big) E\left[\log^+(\|X_{n,k}\|)\, g_A(W_{n,0}) \mid \|X_0\| > u_n\right] + \mathrm{O}(r_n v_n) \\
&\to \sum_{k\in\mathbb{Z}} \lim_{n\to\infty} \Big(1 - \frac{|k|}{r_n}\Big) E\left[\log^+(\|X_{n,k}\|)\, g_A(W_{n,0}) \mid \|X_0\| > u_n\right],
\end{aligned}
$$

provided the limit exists, where again the last step holds due to Pratt's Lemma. This lemma can be applied because of condition (PP1) and

$$
\begin{aligned}
&\Big(1 - \frac{k}{r_n}\Big) E\left[\log^+(\|X_{n,k}\|)\, g_A(W_{n,0}) \mid \|X_0\| > u_n\right] \\
&\le E\left[\max(\log^+(\|X_{n,0}\|), 1) \max(\log^+(\|X_{n,|k|}\|), \mathbb{1}_{\{\|X_{|k|}\|>u_n\}}) \mid \|X_0\| > u_n\right] \\
&\le e_n'(|k|)
\end{aligned}
$$

for all $k \in \mathbb{Z}$, since $g_A(W_{n,0}) \le \mathbb{1}_{\{\|X_0\|>u_n\}}$ and for $k < 0$ stationarity was applied. Now, again we use approximating arguments as in the proof of Proposition 5.2.1. Define $g_A^{(m)}$ as in (5.7.1) and recall that this function is continuous. The weak convergence defining the tail process (2.1.1) together with the continuous mapping theorem imply the weak convergence

$$
\begin{aligned}
&\mathcal{L}\left(\log^+(\|X_{n,j}\|)\, g_A^{(m)}(W_{n,k}) \mid \|X_0\| > u_n\right) \\
&\to \mathcal{L}\left(\log^+(\|Y_j\|)\, \mathbb{1}_{\{\|Y_k\|>1\}} \sum_{|h|\le m} \frac{\|Y_h\|^\alpha}{\sum_{|k|\le m}\|Y_k\|^\alpha} \mathbb{1}_A\left(\frac{Y_{h+i}}{\|Y_h\|}\right)\right)
\end{aligned}
$$

for all $j, k \in \mathbb{Z}$. The P^Y-a.s. continuity of the function applied follows from the continuity of $\log^+(x)$ and $g_A^{(m)}$.

Since $\log^+(x) \le c_q x^q$ for $q \in (0, \alpha)$ and some constant $c_q > 0$, the regular variation of the time series and (5.5.6) imply

$$
\begin{aligned}
&\sup_{n\in\mathbb{N}} E\left[\left(g_A^{(m)}(W_{n,j}) \log^+(\|X_{n,k}\|)\right)^{1+\eta} \Big| \|X_0\| > u_n\right] \\
&\le \sup_{n\in\mathbb{N}} E\left[\left(\log^+(\|X_{n,k}\|)\right)^{1+\eta} \Big| \|X_0\| > u_n\right] < \infty. \qquad (5.7.43)
\end{aligned}
$$

This uniform moment bound in turn implies the uniform integrability of random variables with the distribution of $g_A^{(m)}(W_{n,j}) \log^+(\|X_{n,k}\|)$ given $\|X_0\| > u_n$, $n \in \mathbb{N}$. Thus, the weak convergence stated before implies

$$
\begin{aligned}
&E\left[g_A^{(m)}(W_{n,0}) \log^+(\|X_{n,k}\|) \mid \|X_0\| > u_n\right] \\
&\to E\left[\sum_{|h|\le m} \frac{\|Y_h\|^\alpha}{\sum_{|k|\le m}\|Y_k\|^\alpha} \mathbb{1}_A\left(\frac{Y_{h+i}}{\|Y_h\|}\right) \log^+(\|Y_k\|)\right]
\end{aligned}
$$

for all $m \in \mathbb{N}$. Moreover, due to the uniform integrability condition, it also follows

$E[\log^+(\|Y_k\|)] < \infty$ for all $k \in \mathbb{Z}$, and, therefore, dominated convergence implies

$$\lim_{m\to\infty} E\left[\sum_{|h|\le m} \frac{\|Y_h\|^\alpha}{\sum_{|k|\le m}\|Y_k\|^\alpha} \mathbb{1}_A\left(\frac{Y_{h+i}}{\|Y_h\|}\right)\log^+(\|Y_k\|)\right]$$
$$= E\left[\sum_{h\in\mathbb{Z}} \frac{\|Y_h\|^\alpha}{\|Y\|_\alpha^\alpha} \mathbb{1}_A\left(\frac{Y_{h+i}}{\|Y_h\|}\right)\log^+(\|Y_k\|)\right].$$

In addition,

$$|E\Big[g_A(W_{n,0})\log^+(\|X_{n,k}\|) - g_A^{(m)}(W_{n,0})\log^+(\|X_{n,k}\|) \mid \|X_0\| > u_n\Big]|$$
$$\le E\Big[|g_A(W_{n,0}) - g_A^{(m)}(W_{n,0})|\log^+(\|X_{n,k}\|) \mid \|X_0\| > u_n\Big]$$
$$\le E\Big[|g_A(W_{n,0}) - g_A^{(m)}(W_{n,0})|^{(1+\delta)/\delta} \mid \|X_0\| > u_n\Big]^{\delta/(1+\delta)}$$
$$\times E\Big[\log^+(\|X_{n,k}\|)^{1+\delta} \mid \|X_0\| > u_n\Big]^{1/(1+\delta)}$$
$$\le E\Big[2^{1/\delta}|g_A(W_{n,0}) - g_A^{(m)}(W_{n,0})| \mid \|X_0\| > u_n\Big]^{\delta/(1+\delta)}$$
$$\times E\Big[\log^+(\|X_{n,k}\|)^{1+\delta} \mid \|X_0\| > u_n\Big]^{1/(1+\delta)}.$$

While the second expectation is bounded due to (5.7.43), the first expectation converges to 0 as $n \to \infty$ and then $m \to \infty$ by (5.7.5). Thus,

$$\lim_{m\to\infty}\limsup_{n\to\infty} |E\Big[g_A(W_{n,0})\log^+(\|X_{n,k}\|) - g_A^{(m)}(W_{n,0})\log^+(\|X_{n,k}\|)\Big| \|X_0\| > u_n\Big]| = 0$$

and therefore

$$E\Big[g_A(W_{n,0})\log^+(\|X_{n,k}\|)\Big| \|X_0\| > u_n\Big] = \lim_{m\to\infty} E\Big[g_A^{(m)}(W_{n,0})\log^+(\|X_{n,k}\|)\Big| \|X_0\| > u_n\Big]$$
$$\xrightarrow[n\to\infty]{} \lim_{m\to\infty} E\Big[g_A^{(m)}((Y_h)_{h\in\mathbb{Z}})\log^+(\|Y_k\|)\Big] = E\Big[g_A((Y_h)_{h\in\mathbb{Z}})\log^+(\|Y_k\|)\Big]$$

for all $k \in \mathbb{Z}$. All in all,

$$\sum_{k\in\mathbb{Z}}\lim_{n\to\infty}\Big(1-\frac{|k|}{r_n}\Big)E\left[\log^+(\|X_{n,k}\|)\, g_A(W_{n,0})\Big|\|X_0\| > u_n\right]$$
$$= \sum_{k\in\mathbb{Z}} E\left[\sum_{h\in\mathbb{Z}} \frac{\|Y_h\|^\alpha}{\|Y\|_\alpha^\alpha} \mathbb{1}_A\left(\frac{Y_{h+i}}{\|Y_h\|}\right)\log^+(\|Y_k\|)\mathbb{1}_{\{\|Y_0\|>1\}}\right].$$

Moreover, using the Par(α)-distribution of $\|Y_0\|$ and $Y_k \overset{d}{=} \Theta_k\|Y_0\|$ with $\|Y_0\|$ independent of Θ_k, we obtain

$$E\left[\sum_{h\in\mathbb{Z}} \frac{\|Y_h\|^\alpha}{\|Y\|_\alpha^\alpha} \mathbb{1}_A\left(\frac{Y_{h+i}}{\|Y_h\|}\right)\log^+(\|Y_k\|)\mathbb{1}_{\{\|Y_0\|>1\}}\right]$$

$$
\begin{aligned}
&= E\left[\sum_{h\in\mathbb{Z}} \frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbb{1}_A\left(\frac{\Theta_{h+i}}{\|\Theta_h\|}\right) \log^+(\|\Theta_k\|\|Y_0\|)\right] \\
&= \int \int_{1\vee\|\theta_k\|^{-1}}^{\infty} \sum_{h\in\mathbb{Z}} \frac{\|\theta_h\|^\alpha}{\|\theta\|_\alpha^\alpha} \mathbb{1}_A\left(\frac{\theta_{h+i}}{\|\theta_h\|}\right) \log(\|\theta_k\| y)\alpha y^{-\alpha-1} dy P^\Theta(d\theta) \\
&= \int \left[-\sum_{h\in\mathbb{Z}} \frac{\|\theta_h\|^\alpha}{\|\theta\|_\alpha^\alpha} \mathbb{1}_A\left(\frac{\theta_{h+i}}{\|\theta_h\|}\right) \alpha^{-1} y^{-\alpha}(\alpha \log(\|\theta_k\| y) + 1)\right]_{y=1\vee\|\theta_k\|^{-1}}^{\infty} P^\Theta(d\theta) \\
&= E\left[\sum_{h\in\mathbb{Z}} \frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbb{1}_A\left(\frac{\Theta_{h+i}}{\|\Theta_h\|}\right) (1\wedge \|\Theta_k\|^\alpha)(\log(\|\Theta_k\|\vee 1) + \alpha^{-1})\right],
\end{aligned}
$$

which proves assertion (ii). □

5.7.5 Proofs for Sections 5.3 and 5.4

Apart from Theorem 5.3.1, which was proven in the previous section, in Section 5.3 we have only to prove Lemma 5.3.2 about the strengthened Condition (PM1).

Proof of Lemma 5.3.2. We start with assertion (i): since $Y_h \stackrel{d}{=} \Theta_h\|Y_0\|$ and $(a+b)^{1+\delta} \leq 2^{1+\delta}(a^{1+\delta} + b^{1+\delta})$ for $a, b, \delta > 0$, we directly conclude

$$
\begin{aligned}
&E\left[\frac{\sum_{|h|\leq m} |\log\|Y_h\||^{1+\delta}\|Y_h\|^\alpha}{\sum_{|k|\leq m}\|Y_k\|^\alpha}\right] = E\left[\frac{\sum_{|h|\leq m} |\log\|\Theta_h\| + \log\|Y_0\||^{1+\delta}\|\Theta_h\|^\alpha}{\sum_{|k|\leq m}\|\Theta_k\|^\alpha}\right] \\
&\leq 2^{1+\delta}\left(E\left[\frac{\sum_{|h|\leq m} |\log\|\Theta_h\||^{1+\delta}\|\Theta_h\|^\alpha}{\sum_{|k|\leq m}\|\Theta_k\|^\alpha}\right] + E\left[\frac{\sum_{|h|\leq m} |\log\|Y_0\||^{1+\delta}\|\Theta_h\|^\alpha}{\sum_{|k|\leq m}\|\Theta_k\|^\alpha}\right]\right) \\
&= 2^{1+\delta}\left(E\left[\frac{\sum_{|h|\leq m} |\log\|\Theta_h\||^{1+\delta}\|\Theta_h\|^\alpha}{\sum_{|k|\leq m}\|\Theta_k\|^\alpha}\right] + E\Big[|\log\|Y_0\||^{1+\delta}\Big]\right) < \infty.
\end{aligned}
$$

Here, the first expectation in the last line is finite due to (PM) (ii). The second expectation is finite due to regular variation of the time series, since $(\log^+ x)^{1+\delta} = o(x^{\alpha-\varepsilon})$ for some $\varepsilon > 0$, i.e $(\log^+ x)^{1+\delta} \leq cx^{\alpha-\varepsilon}$ for some $c > 0$, and $\|Y_0\| \geq 1$ a.s. (for the proof of finiteness of the expectation recall the definition of the tail process and see also the proof for uniform integrability in (5.5.6), alternatively see also Kulik and Soulier (2020), Section 2.3.3). Thus, Condition (PM) (ii) implies (5.3.2).
Conversely, assuming (5.3.2) yields

$$
\begin{aligned}
&E\left[\frac{\sum_{|h|\leq m} |\log\|\Theta_h\||^{1+\delta}\|\Theta_h\|^\alpha}{\sum_{|k|\leq m}\|\Theta_k\|^\alpha}\right] \\
&= E\left[\frac{\sum_{|h|\leq m} |\log\|Y_h\| - \log\|Y_0\||^{1+\delta}\|Y_h\|^\alpha}{\sum_{|k|\leq m}\|Y_k\|^\alpha}\right] \\
&\leq 2^{1+\delta}\left(E\left[\frac{\sum_{|h|\leq m} |\log\|Y_h\||^{1+\delta}\|Y_h\|^\alpha}{\sum_{|k|\leq m}\|Y_k\|^\alpha}\right] + E\left[\frac{\sum_{|h|\leq m} |\log\|Y_0\||^{1+\delta}\|Y_h\|^\alpha}{\sum_{|k|\leq m}\|Y_k\|^\alpha}\right]\right) \\
&\leq 2^{1+\delta}\left(E\left[\frac{\sum_{|h|\leq m} |\log\|Y_h\||^{1+\delta}\|Y_h\|^\alpha}{\sum_{|k|\leq m}\|Y_k\|^\alpha}\right] + E\Big[|\log\|Y_0\||^{1+\delta}\Big]\right) < \infty.
\end{aligned}
$$

Thus, condition (PM) (ii) is equivalent to (5.3.2). Next we turn to assertion (ii). If the family of random variables

$$\frac{1}{v_n}\mathbb{1}_{\{\|X_0\|>u_n\}}\sum_{|h|\le(s_n\wedge m)}\frac{|\log(\|X_h\|/u_n)|^{1+\delta}\|X_h\|^\alpha}{\sum_{|k|\le(s_n\wedge m)}\|X_k\|^\alpha},\ n,m\in\mathbb{N}$$

is uniform integrable for some $\delta>0$, then the weak convergence defining the tail process and dominated convergence imply

$$\begin{aligned}&\limsup_{m\to\infty}E\Big[\sum_{|h|\le m}\frac{|\log(\|Y_h\|)|^{1+\delta}\|Y_h\|^\alpha}{\sum_{|k|\le m}\|Y_k\|^\alpha}\Big]\\&\limsup_{m\to\infty}\lim_{n\to\infty}E\Big[\sum_{|h|\le(s_n\wedge m)}\frac{|\log(\|X_h\|/u_n)|^{1+\delta}\|X_h\|^\alpha}{\sum_{|k|\le(s_n\wedge m)}\|X_k\|^\alpha}\,\Big|\,\|X_0\|>u_n\Big]\\&\le\sup_{m\in\mathbb{N}}\sup_{n\in\mathbb{N}}E\Big[\sum_{|h|\le(s_n\wedge m)}\frac{|\log(\|X_h\|/u_n)|^{1+\delta}\|X_h\|^\alpha}{\sum_{|k|\le(s_n\wedge m)}\|X_k\|^\alpha}\,\Big|\,\|X_0\|>u_n\Big]<\infty,\end{aligned}$$

i.e. (PM) (ii) is satisfied. The uniform integrability is implied by the uniform moment bound

$$\sup_{m\in\mathbb{N}}\sup_{n\in\mathbb{N}}E\Bigg[\Bigg(\frac{\sum_{|h|\le(s_n\wedge m)}|\log(\|X_{n,h}\|)|^{1+\delta}\|X_{n,h}\|^\alpha}{\sum_{|k|\le(s_n\wedge m)}\|X_{n,k}\|^\alpha}\Bigg)^{1+\eta}\,\Bigg|\,\|X_0\|>u_n\Bigg]<\infty$$

for some $\eta>0$. Applying the Hölder inequality for sums yields, similar to the arguments in (5.7.28),

$$\Bigg(\frac{\sum_{|h|\le(s_n\wedge m)}|\log(\|X_{n,h}\|)|^{1+\delta}\|X_{n,h}\|^\alpha}{\sum_{|k|\le(s_n\wedge m)}\|X_{n,k}\|^\alpha}\Bigg)^{1+\eta}\le\frac{\sum_{|h|\le(s_n\wedge m)}|\log(\|X_{n,h}\|)|^{1+\delta'}\|X_{n,h}\|^\alpha}{\sum_{|k|\le(s_n\wedge m)}\|X_{n,k}\|^\alpha},$$

with $\delta'=(1+\delta)(1+\eta)-1>0$. Therefore, it suffices to show

$$\sup_{m\in\mathbb{N}}\sup_{n\in\mathbb{N}}E\left[\frac{\sum_{|h|\le(s_n\wedge m)}|\log(\|X_{n,h}\|)|^{1+\eta'}\|X_{n,h}\|^\alpha}{\sum_{|k|\le(s_n\wedge m)}\|X_{n,k}\|^\alpha}\,\Bigg|\,\|X_0\|>u_n\right]<\infty\tag{5.7.44}$$

for some $\eta'>0$. Note that similar to the calculations in the proof of Lemma 5.7.1 following equation (5.7.29) one has

$$\begin{aligned}&E\left[\frac{\sum_{|h|\le(s_n\wedge m)}|\log(\|X_{n,h}\|)|^{1+\eta'}\|X_{n,h}\|^\alpha}{\sum_{|k|\le(s_n\wedge m)}\|X_{n,k}\|^\alpha}\Bigg|\|X_0\|>u_n\right]\\&\le E\left[\sup_{-s_n\le h\le s_n}\Big(\log^+(\|X_{n,h}\|)\Big)^{1+\eta'}+\sum_{h=-s_n}^{s_n}\Big(\log^-(\|X_{n,h}\|)\Big)^{1+\eta'}\|X_{n,h}\|^\alpha\Bigg|\|X_0\|>u_n\right]\end{aligned}$$

and hence Conditions (PM1) (i) and (ii) (with $\eta'=\delta$) imply the Lyapunov-type condition (5.7.44). Thus, condition (PM) (ii) holds. This concludes the proof. □

Next, we turn to the only proof for Section 5.4, namely the asymptotic normality of the projection based estimator for multiple time points. This proof is quite short, since the proof is basically the same as for Theorem 5.2.9 and Theorem 5.3.1 and here we only emphasis the few differences.

Proof of Theorem 5.4.1. One has $\hat{p}^M_{n,A} = T_{n,g_{A,M}}/T_{n,g_{(\mathbb{R}^d)^M,M}}$ and the proof for the first assertion regarding the asymptotic normality of $\hat{p}^M_{n,A}$ is the same as the proof of Theorem 5.2.9. And also the proof of the asymptotic normality of $\hat{\hat{p}}^M_{n,A}$ works along the same line as the lengthy proof of Theorem 5.3.1, including the proofs of Lemma 5.2.4, Proposition 5.2.5, Proposition 5.2.7 and Lemmas 5.7.1-5.7.7. The only difference is that one uses the $[0,1]$-valued function $g_{A,M}$ instead of g_A. In the proofs just mentioned, one only has to replace the indicators $\mathbb{1}_A(X_{t+h+i}/\|X_t+h\|)$ by $\mathbb{1}_A(X_{t+h+i_1}/\|X_{t+h}\|,...,X_{t+h+i_M}/\|X_{t+h}\|)$. Where conditions depending on the time point i and set A where used, one now has to use the modified condition as stated in the assertion.

For the process convergence note that the arguments for the bracketing remain the same as in the proof of Proposition 5.2.7, since we use the same assumptions there, in particular the family of sets $\mathcal{A}$ has the same structure as before.

We omit the details since all arguments are the same, but the notation becomes much more messy. $\square$

Chapter 6

Outlook

In this final chapter, we consider shortly two possible further research areas which could follow-up this thesis. One additional open field for future research are bootstrap techniques to estimate the asymptotic variance of $\hat{\hat{p}}_{n,A}$, as mentioned on page 151. This problem has already been addressed there and will not be further elaborated here.

Sliding blocks in POT and block maxima setting

One aspect for future research might be the efficiency advantage of sliding blocks estimators. For our peak-over-threshold approach, we have shown in Theorem 3.3.6 that sliding blocks estimators always have an asymptotic variance less than or equal to the asymptotic variance of the corresponding disjoint blocks estimator. For the examples of the extremal index and the stop-loss index in Section 4, the two variances are equal. The same holds for more general cluster indexes under suitable conditions, as shown by Cissokho and Kulik (2021). Robert et al. (2009) proved that the sliding blocks version of their estimator strictly outperforms the disjoint version, however, they used a smaller threshold u_n with $r_n P(X_0 > u_n) \to \tau > 0$. For the block maxima approach, Zou et al. (2021) also show that the sliding blocks estimator is at least as efficient as its disjoint counterpart. In that paper and also in Bücher and Segers (2018*a*) and Bücher and Jennessen (2020*b*), examples where the asymptotic variance of the sliding blocks estimator is actually strictly smaller than the variance of the disjoint estimator are considered. This raises the question for the reason for these differences between the POT and the block maxima approach, in particular, whether this is only a random phenomenon or whether there is a deeper structural difference between these approaches. This difference could be related to the question which observations are considered as extreme and, therefore, how large u_n is chosen, since for the POT setting one usually uses $r_n P(X_0 > u_n) \to 0$ while in the block maxima framework one has $r_n P(X_0 > u_n) \sim 1$. Thus, here is an open question regarding the structural difference between the POT and the block maxima setting, which could explain the different behavior of sliding blocks statistics compared to disjoint blocks statistics. A similar open question was recently also formulated by Cissokho and Kulik

(2021), Section 5.3.

Projection based estimator for the extremal index

Another issue for future research could be the method of the projection based estimation for extremal dependency. In (5.1.3), a projection based estimator was defined for the whole distribution of the spectral tail process $(\Theta_t)_{|t|\leq s_n}$. In the course of Chapter 5, only exemplary estimators for $P(\Theta_i \in A)$ for $i \in \mathbb{Z}$ and Borel-sets $A \subset \mathbb{R}^d$ were discussed in greater detail. However, the approach for the projection-based estimator can also be used for other probabilities, parameters or indexes that depend on the spectral tail process $(\Theta_t)_{t\in\mathbb{Z}}$. Exemplary, we consider this for the extremal index θ introduced in Definition 4.2.1. (Even though there are already a lot of estimators for this index.)
Under the conditions (θ1) and (θP), (4.2.2) holds due to the discussion above. Thus, for a regularly varying time series $(X_t)_{t\in\mathbb{Z}}$, by the definition of the tail process $(Y_t)_{t\in\mathbb{Z}}$ and by Lemma 2.1.9

$$\theta = \lim_{n\to\infty} P(M_{1,s_n} \leq u_n \mid X_0 > u_n) = P\Big(\sup_{t\geq 1} Y_t \leq 1\Big)$$

holds with $M_{j,k} := \max(X_j, ..., X_k)$ and $s_nP(X_0 > u_n) \to 0$, $s_n \to \infty$. Direct calculations using $Y_t \overset{d}{=} \Theta_t\|Y_0\|$, the Par($\alpha$)-distribution of $\|Y_0\|$ and independence of $\|Y_0\|$ and $(\Theta_t)_{t\in\mathbb{Z}}$ yields

$$\begin{aligned}
\theta &= P\Big(\sup_{t\geq 1} Y_t \leq 1\Big) = E\Big[\mathbb{1}_{\{\sup_{t\geq 1}\Theta_t\|Y_0\|\leq 1\}}\Big] \\
&= E\Big[\int_1^\infty \mathbb{1}_{\{\sup_{t\geq 1}\Theta_t\leq y^{-1}\}} P^{\|Y_0\|}(dy)\Big] = E\bigg[\int_1^{(\sup_{t\geq 1}\Theta_t)^{-1}} \alpha y^{-\alpha-1} dy\bigg] \\
&= E\Big[[-y^{-\alpha}]_{y=1}^{(\sup_{t\geq 1}\Theta_t)^{-1}\vee 1}\Big] = E\bigg[-\bigg(\Big(\sup_{t\geq 1}\Theta_t\Big)^{-1}\vee 1\bigg)^{-\alpha} + 1\bigg] \\
&= 1 - E\bigg[\bigg(\Big(\sup_{t\geq 1}\Theta_t\Big)\wedge 1\bigg)^{\alpha}\bigg] = 1 - E\Big[\Big(\sup_{t\geq 1}\Theta_t^\alpha\Big)\wedge 1\Big].
\end{aligned} \tag{6.0.1}$$

An empirical version of this representation of the extremal index based on observations $X_1, ..., X_{n+s_n}$ leads to the estimator

$$1 - \frac{1}{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}}} \sum_{t=1}^n \mathbb{1}_{\{\|X_t\|>u_n\}} \bigg(\Big(\max_{1\leq j\leq s_n} \frac{X_{t+j}}{\|X_t\|}\Big)^\alpha \wedge 1\bigg).$$

If one uses the RS-transformation on this estimator or, equivalently, if one calculates the expectation in (6.0.1) by using the estimated measure from (5.1.3) instead of the true distribution of $(\Theta_t)_{t\in\mathbb{Z}}$, one is lead to the following projection based estimator for the extremal index:

$$\hat{\theta}_n^{RS} := 1 - \sum_{t=1}^{n} \frac{\mathbb{1}_{\{\|X_t\|>u_n\}}}{\sum_{s=1}^{n} \mathbb{1}_{\{\|X_s\|>u_n\}}} \sum_{h=-s_n}^{s_n} \frac{\|X_{t+h}\|^\alpha}{\sum_{k=-s_n}^{s_n} \|X_{t+k}\|^\alpha} \left(\left(\max_{\substack{j\geq 1\\ |h+j|\leq s_n}} \frac{X_{t+h+j}}{\|X_{t+h}\|} \right)^\alpha \wedge 1 \right).$$

Since α is typically unknown, one has to replace α by an estimator $\hat{\alpha}_n$, e.g. as given in (5.1.5).

The extremal index θ can be estimated by $\hat{\theta}_n^{RS}$ and the asymptotic normality can possibly be derived with the results of Section 3.2, due to its form of a sliding blocks estimator. The regular variation used here is an additional assumption compared to Section 4.2. This estimator $\hat{\theta}_n^{RS}$ is motivated by a general principle and is not as specifically designed for the estimation of the extremal index as the estimators in Section 4.2. However, the construction makes use of the fundamental properties of the spectral tail process, which possibly improves the estimation. Therefore, it would be interesting to investigate how well this projection based estimator $\hat{\theta}_n^{RS}$ performs compared to known estimators from the literature. However, (6.0.1) is just one representation of θ, other representations may lead to different projection based estimators for this index.

This is only an example how projection based estimators can be constructed for indexes that depend on the spectral tail process. The RS projection method provides a new tool for the construction of estimators for indexes characterizing extreme events, in particular extreme dependences. This tool can also be used for other interesting parameters.

Bibliography

Alsmeyer, G. (2003), 'On the harris recurrence of iterated random lipschitz functions and related convergence rate results', *Journal of Theoretical Probability* **16**(1), 217–247.

Basrak, B., Davis, R. A. and Mikosch, T. (2002), 'Regular variation of garch processes', *Stochastic Processes and Their Applications* **99**(1), 95–115.

Basrak, B., Planinić, H. and Soulier, P. (2018), 'An invariance principle for sums and record times of regularly varying stationary sequences', *Probability Theory and Related Fields* **172**(3-4), 869–914.

Basrak, B. and Segers, J. (2009), 'Regularly varying multivariate time series', *Stochastic Processes and Their Applications* **119**(4), 1055–1080.

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. L. (2004), *Statistics of Extremes: Theory and Applications*, John Wiley & Sons, Chichester, UK.

Berghaus, B. and Bücher, A. (2018), 'Weak convergence of a pseudo maximum likelihood estimator for the extremal index', *The Annals of Statistics* **46**(5), 2307–2335.

Billingsley, P. (1968), *Convergence of Probability Measures*, Wiley, New York.

Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1989), *Regular variation*, Vol. 27, Cambridge university press, Cambridge.

Brockwell, P. J. and Davis, R. A. (1991), *Time series: theory and methods*, Springer Science & Business Media, New York.

Bücher, A. and Jennessen, T. (2020*a*), 'Method of moments estimators for the extremal index of a stationary time series', *Electronic Journal of Statistics* **14**(2), 3103–3156.

Bücher, A. and Jennessen, T. (2020*b*), 'Statistical analysis for stationary time series at extreme levels: new estimators for the limiting cluster size distribution', *arXiv preprint:2011.04733* .

Bücher, A. and Segers, J. (2018*a*), 'Inference for heavy tailed stationary time series based on sliding blocks', *Electronic Journal of Statistics* **12**(1), 1098–1125.

Bücher, A. and Segers, J. (2018*b*), 'Maximum likelihood estimation for the fréchet distribution based on block maxima extracted from a time series', *Bernoulli* **24**(2), 1427–1462.

Bücher, A., Volgushev, S. and Zou, N. (2019), 'On second order conditions in the multivariate block maxima and peak over threshold method', *Journal of Multivariate Analysis* **173**, 604–619.

Bücher, A. and Zhou, C. (2018), 'A horse racing between the block maxima method and the peak-over-threshold approach', *arXiv preprint:1807.00282* .

Buraczewski, D., Damek, E., Mikosch, T. et al. (2016), *Stochastic models with power-law tails*, Springer, New York.

Cai, J. J. (2019), 'A nonparametric estimator of the extremal index', *arXiv preprint:1911.06674* .

Castillo, E. (2012), *Extreme value theory in engineering*, Academia Press, Boston.

Cissokho, Y. and Kulik, R. (2021), 'Estimation of cluster functionals for regularly varying time series: sliding blocks estimators', *Electronic Journal of Statistics* **15**(1), 2777–2831.

Coles, S. (2001), *An introduction to statistical modeling of extreme values*, Springer, London.

Davis, R. A., Drees, H., Segers, J. and Warchoł, M. (2018), 'Inference on the tail process with application to financial time series modeling', *Journal of Econometrics* **205**(2), 508–525.

Davis, R. A. and Hsing, T. (1995), 'Point process and partial sum convergence for weakly dependent random variables with infinite variance', *The Annals of Probability* pp. 879–917.

Davis, R. A. and Mikosch, T. (2009*a*), Extremes of stochastic volatility models, *in* 'Handbook of Financial Time Series', Springer, Berlin, pp. 355–364.

Davis, R. A., Mikosch, T. and Cribben, I. (2012), 'Towards estimating extremal serial dependence via the bootstrapped extremogram', *Journal of Econometrics* **170**(1), 142–152.

Davis, R. and Mikosch, T. (2009*b*), 'The extremogram: a correlogram for extreme events', *Bernoulli* **15**(4), 977–1009.

De Haan, L. and Ferreira, A. (2006), *Extreme value theory: an introduction*, Springer Science & Business Media, New York.

Dombry, C., Hashorva, E. and Soulier, P. (2018), ‘Tail measure and spectral tail process of regularly varying time series’, *Annals of Applied Probability* **28**(6), 3884–3921.

Douc, R., Moulines, E., Priouret, P. and Soulier, P. (2018), *Markov chains*, Springer, Berlin.

Doukhan, P. (1994), *Mixing: Properties and Examples*, Vol. 85 of *Lecture Notes in Statistics*, Springer, New York.

Drees, H. (2015), ‘Bootstrapping empirical processes of cluster functionals with application to extremograms’, *arXiv preprint:1511.00420* .

Drees, H., Janßen, A. and Neblung, S. (2021), ‘Cluster based inference for extremes of time series’, *Stochastic Processes and Their Applications* **142**, 1–33.

Drees, H. and Knezevic, M. (2020), ‘Peak-over-threshold estimators for spectral tail processes: random vs deterministic thresholds’, *Extremes* **23**(3), 465–491.

Drees, H. and Neblung, S. (2021), ‘Asymptotics for sliding blocks estimators of rare events’, *Bernoulli* **27**(2), 1239–1269.

Drees, H. and Rootzén, H. (2010), ‘Limit theorems for empirical processes of cluster functionals’, *The Annals of Statistics* **38**(4), 2145–2186.

Drees, H., Segers, J. and Warchoł, M. (2015), ‘Statistics for tail processes of markov chains’, *Extremes* **18**(3), 369–402.

Eberlein, E. (1984), ‘Weak convergence of partial sums of absolutely regular sequences’, *Statistics & Probability Letters* **2**(5), 291–293.

Ehlert, A., Fiebig, U.-R., Janßen, A. and Schlather, M. (2015), ‘Joint extremal behavior of hidden and observable time series with applications to garch processes’, *Extremes* **18**(1), 109–140.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (2013), *Modelling extremal events: for insurance and finance*, Vol. 33, Springer Science & Business Media, New York.

Ferro, C. A. and Segers, J. (2003), ‘Inference for clusters of extreme values’, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65**(2), 545–556.

Fils-Villetard, A., Guillou, A. and Segers, J. (2008), ‘Projection estimators of pickands dependence functions’, *Canadian Journal of Statistics* **36**(3), 369–382.

Finkenstädt, B. and Rootzén, H. (2003), *Extreme values in finance, telecommunications, and the environment*, CRC Press, New York.

Goldie, C. M. (1991), 'Implicit renewal theory and tails of solutions of random equations', *The Annals of Applied Probability* **1**(1), 126–166.

Hsing, T. (1991), 'Estimating the parameters of rare events', *Stochastic Processes and Their Applications* **37**(1), 117–139.

Hsing, T. (1993), 'Extremal index estimation for a weakly dependent stationary sequence', *The Annals of Statistics* **21**(4), 2043–2071.

Janßen, A. (2019), 'Spectral tail processes and max-stable approximations of multivariate regularly varying time series', *Stochastic Processes and Their Applications* **129**(6), 1993–2009.

Janssen, A. and Segers, J. (2014), 'Markov tail chains', *Journal of Applied Probability* **51**(4), 1133–1153.

Kesten, H. (1973), 'Random difference equations and renewal theory for products of random matrices', *Acta Mathematica* **131**, 207–248.

Kosorok, M. R. (2008), *Introduction to empirical processes and semiparametric inference.*, Springer, New York.

Kulik, R. and Soulier, P. (2020), *Heavy Tailed Time Series*, Springer, New York.

Kulik, R., Soulier, P. and Wintenberger, O. (2019), 'The tail empirical process of regularly varying functions of geometrically ergodic markov chains', *Stochastic Processes and Their Applications* **129**(11), 4209–4238.

Leadbetter, M. R. (1983), 'Extremes and local dependence in stationary sequences', *Probability Theory and Related Fields* **65**(2), 291–306.

Leadbetter, M., Weissman, I., De Haan, L. and Rootzén, H. (1989), 'On clustering of high values in statistically stationary series', *Proc. 4th Int. Meet. Statistical Climatology* **16**, 217–222.

Mammen, E., Marron, J., Turlach, B. and Wand, M. (2001), 'A general projection framework for constrained smoothing', *Statistical Science* **16**(3), 232–248.

Meyn, S. P. and Tweedie, R. L. (1992), 'Stability of markovian processes i: Criteria for discrete-time chains', *Advances in Applied Probability* pp. 542–574.

Meyn, S. P. and Tweedie, R. L. (1993), *Markov chains and stochastic stability*, Springer Science & Business Media, New York.

Mikosch, T. and Stărică, C. (2000), 'Limit theory for the sample autocorrelations and extremes of a GARCH (1, 1) process', *The Annals of Statistics* **28**(5), 1427–1451.

Mikosch, T. and Wintenberger, O. (2013), 'Precise large deviations for dependent regularly varying sequences', *Probability Theory and Related Fields* **156**(3-4), 851–887.

Mikosch, T. and Wintenberger, O. (2014), 'The cluster index of regularly varying sequences with applications to limit theory for functions of multivariate markov chains', *Probability Theory and Related Fields* **159**(1-2), 157–196.

Mikosch, T. and Zhao, Y. (2014), 'A fourier analysis of extreme events', *Bernoulli* **20**(2), 803–845.

Northrop, P. J. (2015), 'An efficient semiparametric maxima estimator of the extremal index', *Extremes* **18**(4), 585–603.

O'Brien, G. L. (1987), 'Extreme values for stationary and markov sequences', *The Annals of Probability* **15**(1), 281–291.

Oesting, M. and Schnurr, A. (2020), 'Ordinal patterns in clusters of subsequent extremes of regularly varying time series', *Extremes* **23**(4), 521–545.

Petrov, V. V. (1975), *Sums of Independent Random Variables*, Springer Science, Berlin.

Planinić, H. and Soulier, P. (2018), 'The tail process revisited', *Extremes* **21**(4), 551–579.

Pratt, J. W. (1960), 'On interchanging limits and integrals', *The Annals of Mathematical Statistics* **31**(1), 74–77.

Reiss, R.-D. and Thomas, M. (1997), *Statistical analysis of extreme values*, Vol. 2, Springer, Basel.

Resnick, S. I. (1987), *Extreme values, regular variation and point processes*, Springer, New York.

Resnick, S. I. (2007), *Heavy-tail phenomena: probabilistic and statistical modeling*, Springer Science & Business Media, New York.

Robert, C. Y., Segers, J. and Ferro, C. A. T. (2009), 'A sliding blocks estimator for the extremal index', *Electronic Journal of Statistics* **3**, 993–1020.

Segers, J. (2003), 'Functionals of clusters of extremes', *Advances in Applied Probability* **35**(4), 1028–1045.

Segers, J. (2005), 'Approximate distributions of clusters of extremes', *Statistics & Probability Letters* **74**(4), 330–336.

Smith, R. L. and Weissman, I. (1994), 'Estimating the extremal index', *Journal of the Royal Statistical Society: Series B* **56**(3), 515–528.

Sun, J. and Samorodnitsky, G. (2019), 'Multiple thresholds in extremal parameter estimation', *Extremes* **22**(2), 317–341.

Süveges, M. (2007), 'Likelihood estimation of the extremal index', *Extremes* **10**(1-2), 41–55.

Van der Vaart, A. W. and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, Springer, New York.

Vervaat, W. (1979), 'On a stochastic difference equation and a representation of non–negative infinitely divisible random variables', *Advances in Applied Probability* **11**(4), 750–783.

Volkonskii, V. and Rozanov, Y. A. (1959), 'Some limit theorems for random functions. i', *Theory of Probability & Its Applications* **4**(2), 178–197.

Weissman, I. and Novak, S. Y. (1998), 'On blocks and runs estimators of the extremal index', *Journal of Statistical Planning and Inference* **66**(2), 281–288.

Zou, N., Volgushev, S. and Bücher, A. (2021), 'Multiple block sizes and overlapping blocks for multivariate time series extremes', *The Annals of Statistics* **49**(1), 295–320.

Appendix A

Formalities

A.1 Abstract

For the understanding of the behavior of the extremes of a stationary time series $(X_t)_{t\in\mathbb{Z}}$, the analysis of the extremal dependence in time is of high importance. For quantities describing this temporal dependence of extreme events, block estimators are often used. Block estimators are defined as the average of values $g(W_{n,t})$ for some function g and $W_{n,t} := (X_{t+h}/u_n)_{0\le h\le s_n-1}$. Here, u_n is a deterministic threshold above which we describe observations X_t as extreme and s_n is the block length. An estimator for the extreme temporal dependence can be defined as an average over $g(W_{n,t})$, $1 \le t \le n - s_n + 1$, in which case they are called sliding blocks estimators. Alternatively, it can be defined as an average over $g(W_{n,(t-1)s_n+1})$, $1 \le t \le \lfloor n/s_n \rfloor$, in which case we obtain so-called disjoint blocks estimators.

The asymptotic analysis for disjoint blocks estimators can be performed using the central limit theorems of Drees and Rootzén (2010). For the analysis of sliding blocks estimators, a comparable tool is missing so far. In this thesis, a generalized functional limit theorem for suitable empirical processes is derived. As a special case, for the first time this allows a systematic asymptotic analysis of sliding blocks estimators. Specifically, the asymptotic normality of the standardized sliding blocks estimator is proved under weak conditions.

In general, both the sliding and the disjoint blocks estimator can be used for the same estimation problem. It has been conjectured in the literature that the sliding blocks estimator is more efficient and this has been shown concretely in some examples. In this thesis, we prove that the sliding blocks estimator in the POT setting never has a larger asymptotic variance than the disjoint blocks estimator.

Among the indexes describing specific aspects of the extremal dependence of time series are the so-called cluster indexes. In this thesis, we consider two cluster indexes: the well known extremal index and the newer stop-loss index. For both indexes, the asymptotic distributions of the estimation errors are derived on the basis of the general theory mentioned above and, for the family of stop-loss indexes, even process convergence is shown.

In each case, we consider a sliding blocks estimator, the associated disjoint blocks estimator and a runs estimator. With the unified framework used in this thesis, it is shown that all three estimators for the extremal index have the same asymptotic distribution - a fact that was not yet known in the literature. The asymptotic result for the sliding blocks estimator is shown for the first time in this work.

Under the assumption of regular variation, the spectral tail process $(\Theta_t)_{t\in\mathbb{Z}}$ describes the entire extremal dependence structure of a stationary time series. Thus, for the initial problem of describing the temporal dependence of extremes, the estimation of its distribution is of particular interest. In this thesis, a new type of estimator is proposed, which is based on an invariance principle of the distribution of the spectral tail process. This invariance principle can be used for the construction of estimators by means of a projection method. For the corresponding estimator of $P(\Theta_i \in A)$ with a Borel set A and a lag $i \in \mathbb{Z}$, the asymptotic normality is derived using the general results for sliding blocks estimators mentioned above. Asymptotic normality is proved for both a known and an estimated index of regular variation. The conditions required for these asymptotic results are all shown to be satisfied by the general example of solutions to stochastic recurrence equations. Simulation results show that this new projection based estimator mostly has smaller variance than estimators known from the literature. Moreover, this estimator also has the most stable performance in terms of the RMSE. Overall, the new estimator has some desirable properties that its predecessors from the literature do not possess.

A.2 Zusammenfassung

Für das Verständnis des Verhaltens extremer Beobachtungen einer stationären Zeitreihe $(X_t)_{t\in\mathbb{Z}}$ ist insbesondere die Analyse der extremalen Abhängigkeiten in der Zeit von hoher Bedeutung. Für Kennzahlen, die diese temporale Abhängigkeit extremer Ereignisse beschreiben, werden oft Block-Schätzer benutzt. Diese sind definiert als Durchschnitt der Werte $g(W_{n,t})$ für eine Funktion g und $W_{n,t} := (X_{t+h}/u_n)_{0\leq h\leq s_n-1}$. Dabei ist u_n eine deterministische Schranke, ab deren Überschreiten wir Beobachtungen X_t als extrem beschrieben und s_n die Blocklänge. Eine Teststatistik für die extremale Zeitabhängigkeit kann nun als Durchschnitt über $g(W_{n,t})$, $1 \leq t \leq n - s_n + 1$, definiert werden, dann liegen sogenannte sliding-Block-Schätzer vor. Alternativ kann sie als Durschnitt über $g(W_{n,(t-1)s_n+1})$, $1 \leq t \leq \lfloor n/s_n \rfloor$ gebildet werden, dann liegen sogenannte disjoint-Block-Schätzer vor.

Die asymptotische Analyse von disjoint-Block-Schätzern kann mithilfe der zentralen Grenzwertsätze von Drees and Rootzén (2010) durchgeführt werden. Für die Analyse von sliding-Block-Schätzern fehlte bisher ein vergleichbares Werkzeug. In dieser Arbeit wird ein verallgemeinerter funktionaler Grenzwertsatz für geeignete empirische Prozesse bewiesen. Als Spezialfall ermöglicht dieser erstmals eine systematische asymptotische Analyse von sliding-Block-Schätzern. Konkret wird unter schwachen Bedingungen die asymptotische Normalität des standardisierten sliding-Block-Schätzers hergeleitet.

In der Regel kann man sowohl den sliding- als auch den disjoint-Block-Schätzer für das selbe Schätzproblem verwenden. In der Literatur wurde vermutet, dass der sliding-Block-Schätzer effizienter ist, für einige Beispiele wurde dies konkret gezeigt. In dieser Arbeit wird beweisen, dass der sliding-Block-Schätzer im POT-Setting niemals eine größere asymptotische Varianz als der disjoint-Block-Schätzer hat.

Zu den Kennzahlen, welche spezifische Aspekte der extremalen Abhängigkeit von Zeitreihen beschreiben, gehören die sogenannten Cluster Indexe. In dieser Arbeit betrachten wir zwei Cluster Indexe: Den aus der Literatur wohlbekannten Extremal Index und den neueren Stop-loss Index. Für beide Indexe werden die asymptotischen Verteilungen der Schätzfehler auf Basis der zuvor erwähnten allgemeinen Theorie hergeleitet, wobei für den Stop-loss Index sogar Prozesskonvergenz gezeigt wird. Dabei betrachten wir jeweils einen sliding-Block-Schätzer, den zugehörigen disjoint-Block-Schätzer und einen Runs-Schätzer. Mit dem in dieser Arbeit verwendeten vereinheitlichten Rahmen wird gezeigt, dass alle drei Schätzer für den Extremal Index die gleiche asymptotische Verteilung haben - ein Umstand der in der Literatur noch nicht bekannt war. Das asymptotische Resultat für den sliding-Block-Schätzer wird in dieser Arbeit zum ersten Mal gezeigt.

Unter der Annahme der regulären Variation beschreibt der Tail-Spektralprozess $(\Theta_t)_{t\in\mathbb{Z}}$ die gesamte extremale Abhängigkeitsstruktur einer stationären Zeitreihe. Für das Ausgangsproblem der Beschreibung der temporalen Abhängigkeit von Extremwerten ist also insbesondere die Schätzung dieser Verteilung von Interesse. In dieser Arbeit wird ein neuer Typ von Schätzern vorgeschlagen, welche auf einem Invarianzprinzip der Verteilung des Tail-Spektralprozesses basieren. Dieses Invarianzprinzip kann mittels einer Projektionsmethode für die Konstruktion von Schätzern verwendet werden. Für den Schätzer von $P(\Theta_i \in A)$ für eine Borel-Menge A und ein Lag $i \in \mathbb{Z}$ wird in dieser Arbeit die asymptotische Normalität mit den zuvor genannten allgemeinen Resultaten für sliding-Block-Schätzer hergeleitet. Die asymptotische Normalität wird sowohl für einen bekannten als auch für einen geschätzten Index der regulären Variation bewiesen. Für die asymptotischen Resultate werden eine Reihe an Bedingungen benötigt, diese werden alle für das allgemeine Beispiel der Lösungen von stochastischen Rekurrenzgleichungen verifiziert. Simulationsergebnisse deuten darauf hin, dass dieser neue projektionsbasierte Schätzer im Vergleich zu aus der Literatur bekannten Schätzern zumeist eine kleinere Varianz aufweist. Darüber hinaus hat dieser Schätzer auch im Sinne des RMSE die stabilere Performance. Insgesamt hat der neue Schätzer einige wünschenswerte Eigenschaften, die seine Vorgänger aus der Literatur nicht besitzen.

A.3 Publications related to this dissertation

Extracts of the results of this dissertation have already been published in papers in collaboration with my supervisors Anja Janßen and Holger Drees.

- Drees and Neblung (2021) includes the abstract limit theorem and the sliding blocks limit theorem developed in Sections 3.1 and 3.2 as well as the comparison of disjoint and sliding blocks in Section 3.3. The results about the extremal index in Section 4.2 are also presented in shortened form in that paper.
- Drees et al. (2021) contains the projection based estimator motivated in Chapter 5, the corresponding asymptotic results from Sections 5.2 and 5.3 and parts of the simulation study from Section 5.6. The examples from Section 5.5 are also presented in shortened form in that paper.

A.4 Eidesstattliche Versicherung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Sebastian Neblung
Hamburg, den 05.05.2021