Modeling and Statistical Analysis of Extremes in Stationary Stochastic Processes

Bei der Fakultät Mathematik und Physik der Universität Stuttgart zur Erlangung der Würde eines Doktors der Naturwissenschaften (Dr. rer. nat.) eingereichte Abhandlung

vorgelegt von

Ioan Scheffel

aus Stuttgart

Institut für Stochastik und Anwendungen

Universität Stuttgart

2026

Der Mensch in Pracht, doch ohne Einsicht, er ist wie das Vieh, das verstummt.

Psalm 49,21

Acknowledgments

I am deeply grateful to Prof. Dr. Marco Oesting for providing me with the opportunity to pursue this doctorate, giving me the freedom to work on different topics, and frequently providing helpful feedback and advice. Apart from his expertise, I particularly valued his organizational talent and his ability to create a genuine sense of teamwork. My work with him has been both personally enjoyable and formative. I am very grateful to Prof. Dr. Gilles Stupfler for his substantial contribution to the supervision of this dissertation and for the care and precision with which he followed my work. I also thank him for welcoming me so generously during my several visits to Angers, which made each stay both pleasant and productive.

My closest colleagues, Max Thannheimer and Dr. Carolin Forster, have been by my side from the very beginning. I would like to address them in German: Max, es ist immer wieder erfreulich, wie du mich zu neuen Schandtaten beflügelst, und Carolin, danke für deinen kritischen Geist und deine Freundschaft. I am also grateful to everyone else at the ISA, especially those in the offices on Wankelstraße, for the warm and collegial atmosphere. I often came to the office mainly for your company. We talked a lot, drank specialty coffee together, and frequently played table football. This formed an important part of my social life outside my family and parish, and was great fun.

My thanks also go to the parish communities of St. Elisabeth in Geradstetten and St. Michael in Grunbach for welcoming me and for giving me the opportunity to play the organ during services. Since I had previously earned a degree in church music, I greatly valued the opportunity to continue playing for others. Both the music and the parish communities provided a welcome balance to my work on the dissertation. I am especially grateful to Heidrun Lutz for providing me with a flat in her house, perched on a hill above the Remstal.

Finally, I would like to thank my family, in particular my sister, Carla Scheffel, and my cousin, Ioana Marcu, whom I met frequently during this time, as well as the Romanian side of my family, with whom I spent the final stages of writing this thesis.

Declaration

I, *Ioan Scheffel*, declare herewith that this work with the title

Modeling and Statistical Analysis of Extremes in Stationary Stochastic Processes

is my own work. I declare that I am the sole author of this cumulative dissertation and that I have not reproduced, without acknowledgment, the work of another. I also declare that the electronic copy of this thesis coincides with the physical copies.

Place and Date: Stuttgart, August 14, 2026

Signature: ______________________________

Contents

Abstract

This thesis studies extreme value theory (EVT) for stochastic processes, with a particular focus on the asymptotic analysis and construction of stationary processes with nonstandard dependence structures. The research presented in this thesis focuses on two different types of stochastic processes: long memory linear time series, for which we develop central limit theory for threshold-based tail estimators, and maxima of time-changed Markovian particle systems, which we use to construct a new class of stationary max-infinitely divisible processes. Together, these address both the Peaks-over-Threshold and Block Maxima approaches to EVT.

For long memory linear time series, the main challenge is that these processes generally do not satisfy the mixing and anti-clustering conditions commonly imposed in EVT. Moreover, existing reduction principles for long memory processes are not designed for transformations that depend on increasing thresholds. In Chapter 5, we establish a reduction principle for such threshold-dependent transformations and use it to derive central limit theorems for exceedance-based statistics such as the Hill estimator. In Chapter 6, we extend the theory to the multivariate setting by means of a uniform reduction principle for serial tail dependence estimators. This yields central limit theory for extremogram-type estimators and allows for heavy-tailed innovations with possibly infinite variance, as well as for sample quantiles as thresholds.

Turning to the second class of processes, we consider in Chapter 7 maxima of stationary systems of randomly time-changed Lévy particles. Starting from an invariant process–measure pair underlying a Lévy–Brown–Resnick construction, we use a state-dependent random time change to generate an infinite-dimensional class of stationary max-infinitely divisible processes. The construction changes the dependence structure while preserving stationarity. The resulting processes are shown to lie in the max-domain of attraction of the original Lévy–Brown–Resnick process.

Overall, the thesis develops new methods for the asymptotic analysis of extremes and the construction of stationary processes with nonstandard dependence structures, combining techniques from long memory time series, Poisson particle systems, Markov processes, and EVT.

Zusammenfassung

Diese Dissertation untersucht die Extremwerttheorie (EVT) stationärer stochastischer Prozesse, mit besonderem Schwerpunkt auf der asymptotischen Analyse und Konstruktion stationärer Prozesse mit nichtstandardmäßigen Abhängigkeitsstrukturen. Die in dieser Dissertation vorgestellte Forschung konzentriert sich auf zwei verschiedene Arten stochastischer Prozesse: langzeitabhängige lineare Zeitreihen, für die wir zentrale Grenzwertsätze für schwellenwertbasierte Tail-Schätzer entwickeln, und Maxima zeitveränderter Markovscher Partikelsysteme, die wir verwenden, um eine neue Klasse stationärer max-unendlich-oft teilbarer Prozesse zu konstruieren. Zusammen decken diese Schwerpunkte sowohl die Peaks-over-Threshold- als auch die Block-Maxima-Ansätze der EVT ab.

Für lineare Zeitreihen mit Langzeitabhängigkeiten besteht die zentrale Herausforderung darin, dass diese Prozesse im Allgemeinen nicht die Mischungs- und Anti-Clustering-Bedingungen erfüllen, die in der EVT üblicherweise vorausgesetzt werden. Außerdem sind bestehende Reduktionsprinzipien für Prozesse mit Langzeitabhängigkeiten nicht für Transformationen ausgelegt, die von wachsenden Schwellenwerten abhängen. In Kapitel 5 entwickeln wir ein Reduktionsprinzip für solche schwellenwertabhängigen Transformationen und verwenden es, um zentrale Grenzwertsätze für Schwellenwertüberschreitungs-basierte Statistiken wie den Hill-Schätzer herzuleiten. In Kapitel 6 erweitern wir die Theorie mittels eines uniformen Reduktionsprinzips für serielle Tail-Dependence-Schätzer auf den multivariaten Fall. Dies liefert eine zentrale Grenzwerttheorie für extremogrammartige Schätzer und erlaubt heavy-tailed Innovationen mit möglicherweise unendlicher Varianz sowie Stichprobenquantile als Schwellenwerte.

In Kapitel 7 wenden wir uns der zweiten Klasse von Prozessen zu und betrachten Maxima stationärer Systeme zufällig zeitveränderter Lévy-Partikel. Ausgehend von einem invarianten Prozess-Maß-Paar, das einer Lévy-Brown-Resnick-Konstruktion zugrunde liegt, verwenden wir eine zustandsabhängige zufällige Zeitänderung, um eine unendlichdimensionale Klasse stationärer max-unendlich-oft teilbarer Prozesse zu erzeugen. Die Konstruktion verändert die Abhängigkeitsstruktur, während sie die Stationarität bewahrt. Es wird gezeigt, dass die resultierenden Prozesse im Max-Anziehungsbereich des ursprünglichen Lévy-Brown-Resnick-Prozesses liegen.

Insgesamt entwickelt die Dissertation neue Methoden für die asymptotische Analyse von Extremen und die Konstruktion stationärer Prozesse mit nichtstandardmäßigen Abhängigkeitsstrukturen, wobei Techniken aus der The-

orie langzeitabhängiger Zeitreihen, von Poisson-Partikelsystemen und Markov-Prozessen sowie aus der EVT kombiniert werden.

Part I

Introduction

De particularibus enim etiam philosophia esse non poterit.

For there cannot be philosophy concerning particulars.

Albertus Magnus, *De vegetabilibus et plantis*,
lib. VI, tract. I, cap. 1

Chapter 1

Introduction and Contribution of this Thesis

1.1 Introduction and Motivation

While, for most of history, rare phenomena were seen as marking the boundary of knowledge [Daston and Park, 1998], extreme value theory (EVT) tries to extend knowledge beyond the range of empirical observations. A first extrapolation approach was developed through the work of Fréchet [1927] and Fisher and Tippett [1928] on limiting laws for maxima of independent and identically distributed observations, and was then generalized by Gnedenko [1943]. Since then, EVT has become a standard framework for the statistical analysis and modeling of extreme events [Coles, 2001, Reiss and Thomas, 2007], with applications in engineering and environmental science [Castillo et al., 2005], insurance and finance [Embrechts et al., 1997], and telecommunications [Resnick, 2007].

The extension of EVT from independent to dependent observations began in the 1950s with the work of Watson [1954] on m-dependent stationary stochastic processes. Further developments were made by Berman [1964], Loynes [1965], and Cramér [1966], followed by the work of Leadbetter [1974], Leadbetter et al. [1983]. In these works, stationarity was the baseline to move beyond the theory of Fréchet, Fisher–Tippett, and Gnedenko, and has since remained the standard reference framework in dependent EVT, even as theory has expanded to nonstationary models.

Stationarity assumptions affect both model specification and the validity of statistical analysis [Brockwell and Davis, 1991, Leadbetter et al., 1983]. As a modeling principle, stationarity means that the random mechanism generating a stochastic process is stable over time. Variation in time is seen as random fluctuation rather than as evidence that the random mechanism itself is changing; see Figure 1.1 for an illustration. Hence, as a condition for statistical analysis, stationarity makes it meaningful to pool observations across time to estimate a common feature.

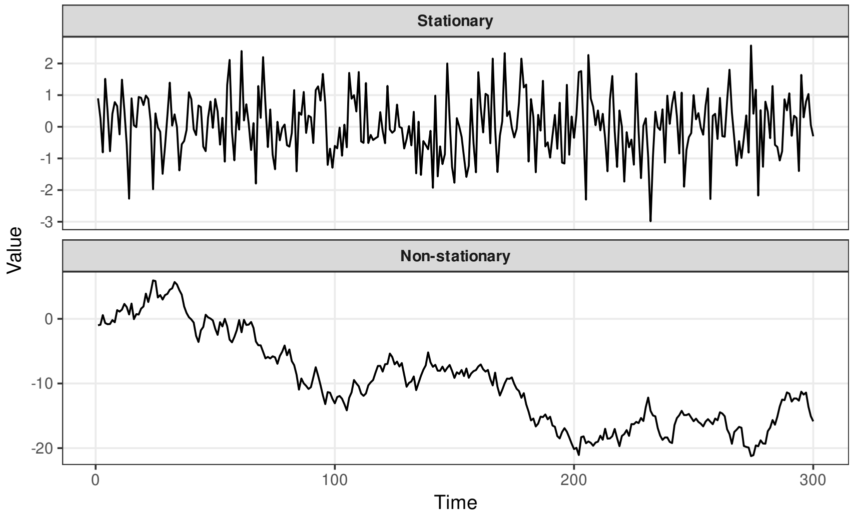

Figure 1.1: Visual comparison of a stationary autoregressive time series and a nonstationary random walk with drift. The stationary series fluctuates around a stable level, while the nonstationary series changes its level over time.

Among the simplest stationary models are linear time series of the form

$$X_t = \sum_{i=0}^{\infty} a_i \varepsilon_{t-i}, \qquad t \in \mathbb{Z},$$

where (a_i) is a deterministic sequence satisfying suitable convergence conditions and (ε_i) is a sequence of independent and identically distributed innovations [Brockwell and Davis, 1991]. In some stationary time series, observations separated by long time spans remain substantially dependent, a phenomenon that is called long-range dependence, or long memory [Beran, 1994]. In linear time series, the strength of dependence is governed by the decay of the coefficients (a_i). In standard long memory models these coefficients decay so slowly that $\sum_{i=0}^{\infty} |a_i| = \infty$. Consequently, many classical results based on short memory summability conditions no longer apply, and both the appropriate normalizations and the limiting distributions may change [Beran et al., 2013, Giraitis et al., 2012]. For extremes, the persistence and clustering of large observations introduce additional difficulties in the analysis of tail-based statistics [Davis and Mikosch, 2009, Kulik and Soulier, 2011].

While in a light-tailed setting with normally distributed innovations, high exceedances often appear relatively isolated, in a heavy-tailed setting, a large shock may produce several exceedances close to one another, leading to visible extremal clusters; see Figure 1.2 for an illustration.

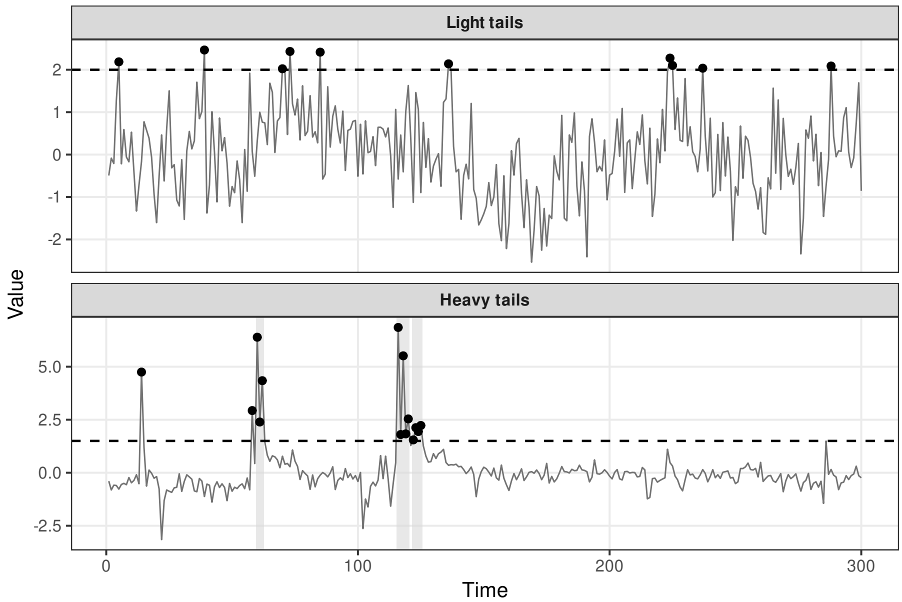

Figure 1.2: Peaks-over-Threshold comparison for a Gaussian ARFIMA path and an ARFIMA path with t-distributed innovations ($\nu = 3$). The dashed line indicates a common threshold on the standardized scale. Exceedances in the heavy-tailed case tend to occur in more pronounced clusters. In the heavy-tailed panel, shaded bands mark clusters of consecutive exceedances.

Although extremal clustering is well studied, much of the existing limit theory for stationary time series relies on anti-clustering conditions, which ensure that extremal dependence is asymptotically local and that extreme clusters do not persist over arbitrarily long time spans [Davis and Hsing, 1995, Basrak and Segers, 2009]. These conditions fail for the models considered in this thesis. Developing theoretical methods for the asymptotic analysis of this persistent extremal regime is therefore one of the central objectives of this thesis. Thus, this thesis lies at the intersection of EVT, which provides a framework for extrapolating beyond observed extremes, and stationarity, which offers a parsimonious model for dependent data. In addition to asymptotic theory for long memory linear time series, this thesis also includes work on continuous-time processes, in particular, maximum processes with a latent Markovian structure [Brown and Resnick, 1977].

Taken together, these settings allow us to address the two principal paradigms of EVT: the Block Maxima and Peaks-over-Threshold (PoT) approaches [Coles, 2001, Bücher and Zhou, 2021]. As a visual guide, the Block Maxima approach may be viewed as a horizontal reduction of the data: the observation period is divided into blocks, and only the maximum within each block is retained. The PoT approach may instead be viewed as a vertical reduction, retaining only observations that exceed a sufficiently high threshold. In asymptotic analysis, both approaches retain only a vanishing fraction of the available observations and therefore present related analytical difficulties. They are discussed in greater detail in the methodology section.

1.2 Contribution of this Thesis

In the first paper [Scheffel et al., 2026a], we study PoT statistics for long memory linear time series. Such statistics depend on high thresholds that grow with the sample size. This puts them outside the standard central limit theory for subordinated long memory linear processes, which usually treats fixed transformations of the data. It also puts them outside much of the existing EVT, since long memory linear processes do not satisfy the usual mixing assumptions. We prove a reduction principle for sample-size-dependent transformations of long memory linear time series. This shows that, under suitable assumptions, threshold-based partial sums have the same asymptotic behavior as the linear part of the underlying process, after a different normalization. As applications, we derive central limit theorems for exceedance counts and Hill-type estimators with deterministic and random thresholds. We treat heavy-tailed and light-tailed cases. Our results show that long memory changes the convergence rates and asymptotic scales of PoT estimators in various ways that do not occur for independent or short memory data.

In the second paper [Scheffel et al., 2026b], we study tail-dependence estimators for long memory linear time series with possibly infinite variance. For this, the available theory for long memory linear time series does not provide the required tools, since it is essentially based on one-dimensional reduction principles and does not cover the multivariate nature of dependence estimators. Conversely, the existing asymptotic theory for such dependence estimators is derived under mixing and anti-clustering assumptions. These assumptions fail for the long memory linear processes considered here. We close this gap by developing a multivariate uniform reduction principle for threshold-dependent indicator functions. This is not a routine extension of the univariate setting of the first paper, because the multivariate setting creates dependence between several overlapping linear processes, and the estimates needed to handle this dependence have to be developed separately. Using this reduction principle, we prove uniform central limit theorems for extremogram-type tail dependence estimators of long memory linear time series. While in Scheffel et al. [2026a] we made the transition from deterministic to random thresholds by proving a separate derandomization device, in this work the uniformity of the reduction principle makes this transition straightforward.

In the third paper [Scheffel, 2026], we study maxima of systems of randomly time-changed Lévy particles. Classical Brown-Resnick constructions show how stationary max-stable processes can arise even from nonstationary particle systems. Known examples of this type, however, give little flexibility beyond the max-stable setting, and the max-domains of attraction of such processes are still poorly understood. The main contribution is a construction principle for stationary max-infinitely divisible processes beyond max-stability. Starting from a Lévy-Brown-Resnick process, we perturb the underlying invariant process-measure pair by a state-dependent random time change. The perturbation is governed by an admissible mass function. Varying this function gives an infinite-dimensional class of stationary max-infinitely divisible processes and changes the dependence structure while preserving stationarity. We then prove that the constructed processes lie in the max-domain of attraction of the original Lévy-Brown-Resnick process. The paper therefore gives both new classes of stationary max-infinitely divisible processes and a concrete description of part of the domain of attraction of a given max-stable process. The argument combines potential theory for Markov processes, Poisson particle systems, and extreme value theory.

1.3 Structure of this Thesis

The remainder of this thesis is structured as follows: We provide theoretical and methodological background for the rest of the thesis in Chapter 2. We then discuss the main results of this thesis in more detail in Chapter 3. In Chapter 4, we clarify the contributions of the author of this thesis to the results. Chapter 5 contains the first paper on central limit theory for PoT estimators in long memory linear time series. Chapter 6 contains the second paper on central limit theory for serial tail dependence estimators in heavy-tailed long memory linear time series. Finally, Chapter 7 contains the third paper on maxima of systems of randomly time-changed Lévy particles.

Chapter 2

Theoretical and Methodological Background

2.1 Stochastic Processes

Throughout, $(\Omega, \mathcal{A}, \mathbf{P})$ denotes a probability space rich enough to support all random variables under consideration. For $p \geq 1$, we write $L^p(\mathbf{P})$ for the space of random variables with finite p-th moment. For a generic, absolutely continuous random variable Z, we let F_Z (resp. f_Z, q_Z) denote its cumulative distribution function (resp. probability density function, quantile function). We say that a stochastic process $(X_t)_{t \in T}$ is stationary if, for every $n \in \mathbb{N}$, $t_1, \ldots, t_n \in T$, and every h such that $t_1 + h, \ldots, t_n + h \in T$,

$$(X_{t_1}, \ldots, X_{t_n}) \qquad \text{has the same distribution as} \qquad (X_{t_1+h}, \ldots, X_{t_n+h}) .$$

2.1.1 Long memory linear time series

We consider linear processes

$$X_t \;=\; \sum_{j \in \mathbb{Z}} a_j \varepsilon_{t-j}, \qquad t \in \mathbb{Z} ,$$

where $(a_j)_{j \in \mathbb{Z}}$ is a real-valued coefficient sequence and $(\varepsilon_j)_{j \in \mathbb{Z}}$ are i.i.d. copies of a random variable ε. We call $(a_j)_{j \in \mathbb{Z}}$ the coefficient sequence and $(\varepsilon_j)_{j \in \mathbb{Z}}$ the innovation sequence of the linear time series $(X_t)_{t \in \mathbb{Z}}$. If $a_{-j} = 0$ for all $j \in \mathbb{N}$, we say that the linear time series is causal. A linear time series is well-defined whenever the random series on the right-hand side converges in some sense, for example, $\mathbf{P}$-almost surely. In the long memory setting considered below, the time series are always causal and the assumptions on the innovations and the decay of the coefficients ensure almost sure convergence [Durrett, 2019, Exercise 2.5.7]. Note that, if the linear process (X_t) exists, then

$$(X_{t_1}, \ldots, X_{t_n}) \;=\; \sum_{j \in \mathbb{Z}} a_j (\varepsilon_{t_1 - j}, \ldots, \varepsilon_{t_n - j})$$

has, for $h \in \mathbb{Z}$, the same distribution as

$$(X_{t_1+h}, \ldots, X_{t_n+h}) = \sum_{j\in\mathbb{Z}} a_j(\varepsilon_{t_1+h-j}, \ldots, \varepsilon_{t_n+h-j}),$$

by the i.i.d. assumption on the innovations, so that (X_t) is stationary.

We now make the long memory setting that we are going to consider more precise. Set

$$\nu = \nu_\varepsilon := \sup\{p > 1 \mid \|\varepsilon\|_{L^p(\mathbf{P})} < \infty\}.$$

We have $\nu = \infty$, for example, if ε is standard Gaussian. If the innovations are regularly varying (see Section 2.2.1.1 for a definition), then $-\nu$ is the index of regular variation. Next we state long memory assumptions on the coefficients (a_i) of the linear time series.

Assumption 1 (Long memory)**.** It holds that $\nu > 1$, and that the sequence $(n^{1-d}a_n)$ converges to a finite, nonzero limit c_a for some $d \in (0, 1 - 1/\alpha)$, where $\alpha := 2 \wedge \nu$.

Remark 2.1.1 (On Assumption 1)**.** Assume that ε is centered. Since $\alpha \in (1, 2]$, it holds for $\delta > 0$ sufficiently small that $\alpha - \delta \in (1, 2)$ and $(1 - d)(\alpha - \delta) > 1$, so that

$$\sum_{i=0}^{\infty} \mathbf{E}[|a_i \cdot \varepsilon_{t-i}|^{\alpha-\delta}] = \mathbf{E}[|\varepsilon|^{\alpha-\delta}] \sum_{i=0}^{\infty} |a_i|^{\alpha-\delta} \le C \cdot \sum_{i=1}^{\infty} i^{-(\alpha-\delta)(1-d)} < \infty,$$

and the series converges $\mathbf{P}$-almost surely [Durrett, 2019, Exercise 2.5.7]. The assumption imposes that the innovations have at least a finite moment of order larger than 1, and that the coefficient sequence (a_j), while forming a divergent series and thus granting the long memory property to (X_t), is able to accommodate potentially heavy tails of the innovations.

2.1.1.1 Reduction principle for subordinated linear processes

We follow Section 4.3 in Beran et al. [2013]. Let $G : \mathbb{R} \to \mathbb{R}$ be a measurable function and consider the subordinated partial sum

$$\sum_{t=1}^{n} (G(X_t) - \mathbf{E}[G(X_0)]) .$$

Such sums are relevant for statistical applications because many estimators can be expressed as empirical averages of (non-linear) transformations of the observations.

At a conceptual level, the reduction principle works as follows. For $k \in \mathbb{Z}$, let $\mathcal{F}_k = \sigma(\varepsilon_k, \varepsilon_{k-1}, \ldots)$ denote the past σ-algebra generated by the sequence (ε_t) up to index k. For $k \ge 0$, let $X_{t,k} := \sum_{j=0}^{k} a_j\varepsilon_{t-j}$ be the k-truncated time series, and write

$$G_k(y) := \mathbf{E}[G(X_{0,k} + y)] \qquad \text{and} \qquad G'_k(y) = \frac{\mathrm{d}}{\mathrm{d}y} G_k(y),$$

where we set $G_\infty(y) := \mathbf{E}[G(X_{0,\infty} + y)] = \mathbf{E}[G(X_0 + y)]$ for $k = \infty$ in the natural way, and make the convention $X_{t,-1} = 0$ and $G_{-1} = G$. Assuming that the functions G_k are differentiable for all $k \in \mathbb{N}_0 \cup \{\infty\}$, the reduction principle consists of showing that

$$
\begin{aligned}
&\sum_{t=1}^{n} (G(X_t) \ - \ \mathbf{E}[G(X_0)]) \\
&= \sum_{t=1}^{n}\sum_{k=0}^{\infty} \left(\mathbf{E}\left[G(X_t) \mid \mathcal{F}_{t-k}\right] \ - \ \mathbf{E}\left[G(X_t) \mid \mathcal{F}_{t-(k+1)}\right]\right) \\
&= \sum_{t=1}^{n}\sum_{k=0}^{\infty} (G_{k-1}(X_t - X_{t,k-1}) \ - \ G_k(X_t - X_{t,k})) \\
&\approx \sum_{t=1}^{n}\sum_{k=0}^{\infty} (G_{k}(X_t - X_{t,k-1}) \ - \ G_k(X_t - X_{t,k})) \\
&\approx \sum_{t=1}^{n}\sum_{k=0}^{\infty} a_k \varepsilon_{t-k} G_k'(X_t - X_{t,k}) \\
&= G_\infty'(0) \sum_{t=1}^{n} X_t \ + \ \sum_{t=1}^{n}\sum_{k=0}^{\infty} a_k \varepsilon_{t-k} \left(G_k'(X_t - X_{t,k}) \ - \ G_\infty'(0)\right) .
\end{aligned}
\tag{2.1}
$$

The first steps rely on a martingale-difference decomposition and on the independence structure of the linear time series, whereas the subsequent approximation steps require additional regularity assumptions and error bounds. The rigorous justification of these steps is the main content of the reduction principle, and the relations above should not be read as readily derived algebraic identities.

Through the reduction principle, the asymptotic profile turns out to be determined only by the linear term, for which central limit theory is readily available; see Theorem 2.1.2. In the first and second papers of this thesis [Scheffel et al., 2026a,b], we extend this reduction principle to so-called univariate PoT estimators, and to serial tail dependence estimators; see Section 2.2.1.2 for an overview of these estimators. The following assumptions are standard for the reduction principle and actually mild, because they allow for infinite variance of the innovations.

Assumption 2 (Regularity of the innovations)**.** Suppose that $F_\varepsilon \in C^2(\mathbb{R})$, and that ε has a symmetric distribution. Furthermore, assume that one of the following alternatives holds:

(i) (Case $\alpha = 2$) One has $\mathbf{E}[\varepsilon^2] < \infty$, and, for some $\delta > 0$, the characteristic function of ε is such that $(1 + |s|)^\delta \mathbf{E}[\exp(is\varepsilon)]$ is uniformly bounded in $s \in \mathbb{R}$.

(ii) (Case $\alpha \in (1, 2)$) Assume that there is a finite positive constant L such that $\lim_{x\to\infty} x^\alpha \mathbf{P}[\varepsilon > x] = \lim_{x\to\infty} x^\alpha \mathbf{P}[\varepsilon < -x] = L/2$, and suppose that the probability density function $f_\varepsilon \in C^1(\mathbb{R})$ satisfies:

$$\begin{aligned} |f_\varepsilon'(x)| &\lesssim (1+|x|)^{-\alpha} \qquad \text{for } x \in \mathbb{R}\,, \\ |f_\varepsilon'(x) - f_\varepsilon'(y)| &\lesssim |x-y| \cdot (1+|x|)^{-\alpha} \qquad \text{for } x, y \in \mathbb{R}\,, |x-y| < 1\,, \end{aligned}$$

where $\lesssim$ denotes inequality up to a multiplicative constant $C_\varepsilon > 0$ depending only on the distribution of ε.

Through the reduction principle, the asymptotic profile of partial sums of subordinated long memory linear time series is determined by the asymptotic behavior of linear partial sums, for which we now provide central limit theory.

Theorem 2.1.2 (Central limit theorems for partial sums)**.** *Let Assumptions 1 and 2 hold. Then*

$$n^{-d-1/\alpha} \sum_{t=1}^{n} X_t \xrightarrow{\mathrm{d}} Z_\alpha \qquad \text{as } n \to \infty\,,$$

where

$$Z_\alpha \quad \text{has distribution} \quad \begin{cases} \mathcal{N}(0, \sigma^2) & \text{if } \alpha = 2 \text{ with } \mathbf{E}[\varepsilon^2] < \infty, \\ S\alpha S(\eta) & \text{if } \alpha \in (1,2)\,, \end{cases}$$

with variance

$$\sigma^2 := c_a^2 \mathbf{E}[\varepsilon^2] \frac{B(1-2d, d)}{d(2d+1)} \quad \text{when } \mathbf{E}[\varepsilon^2] < \infty,$$

or scale

$$\eta := \frac{c_a}{d} \cdot \left(L \cdot \frac{\Gamma(2-\alpha)\cos(\pi\alpha/2)}{1-\alpha} \right)^{1/\alpha} \left(\int_{-\infty}^{1} \left((1-v)_+^d - (-v)_+^d \right)^\alpha \mathrm{d}v \right)^{1/\alpha} .$$

We define the distribution $S\alpha S(\eta)$ in the usual sense, having characteristic function $x \mapsto \exp(-\eta^\alpha |x|^\alpha)$.

For a proof of the last result we refer to Theorem 4.6 ($\alpha = 2$) and Theorem 4.17 ($\alpha \in (1,2)$) in Beran et al. [2013].

2.1.2 Particle Systems

In addition to long memory linear time series, we consider stochastic particle systems. Here, rather than studying a single stochastic process observed over time, we consider a random collection of particles whose positions evolve according to independent copies of a given stochastic process. We first introduce Poisson point processes, which will serve as the random initial configurations of these particle systems.

2.1.2.1 Poisson point process

Definition 2.1.3 (Poisson point process)**.** Let $(\mathcal{X}, \mathcal{C}, \mu)$ be a measure space satisfying $\{x\} \in \mathcal{C}$ for all $x \in \mathcal{X}$ and $\mu(\mathcal{X}) = \infty$. A random collection $\{U^{(i)}\}_{i \in \mathbb{N}}$ of points in $\mathcal{X}$ is called a Poisson point process having intensity measure μ if for every $m \in \mathbb{N}$ and corresponding disjoint sets $C_1, \ldots, C_m \in \mathcal{C}$ and non-negative (including $+\infty$) integers $r_1, \ldots, r_m$ it holds that

$$\mathbf{P}\left[\sum_{i \in \mathbb{N}} \mathbb{1}\{U^{(i)} \in C_j\} \;=\; r_j \qquad \text{for all } j \in \{1, \ldots, m\}\right] \;=\; \prod_{j=1}^{m} p(\mu(C_j), r_j)\,,$$

where

$$p(\lambda, k) \;=\; \begin{cases} \frac{\lambda^k \exp(-\lambda)}{k!} & \text{if } \lambda < \infty \text{ and } k < \infty \text{ and } (\lambda, k) \neq (0,0)\,, \\ 1 & \text{if } (\lambda, k) \in \{(0,0), (\infty, \infty)\}\,, \\ 0 & \text{elsewhere.} \end{cases}$$

The defining property of a Poisson point process is thus that the number of points falling into a measurable set C is Poisson distributed with mean $\mu(C)$, while the numbers of points falling into disjoint sets are independent. The intensity measure μ therefore determines both the expected number of points in a region and how these points are distributed across disjoint regions.

In the literature, a PPP is often represented as a random counting measure $N = \sum_{i \in \mathbb{N}} \delta_{U^{(i)}}$. Here, we instead take a particle-based point of view and work directly with the random collection of atoms $\{U^{(i)}\}_{i \in \mathbb{N}}$. This representation is more convenient for the particle systems considered below, where each point $U^{(i)}$ serves as the initial position of an individual particle.

Remark 2.1.4. A collection $\{U^{(i)}\}_{i \in \mathbb{N}}$, in contrast to a set $\{U^{(i)} \mid i \in \mathbb{N}\}$, can contain the same element more than once. The existence of a PPP with prescribed intensity measure is guaranteed if, for example, the intensity measure is σ-finite [Last and Penrose, 2017, Corollary 3.7].

Theorem 2.1.5 (Mapping theorem; see for example Theorem 24.16 in Klenke [2020])**.** *Let $(\mathcal{X}, \mathcal{C}, \mu)$ be a measure space satisfying $\{x\} \in \mathcal{C}$ for all $x \in \mathcal{X}$ and $\mu(\mathcal{X}) = \infty$, let $(\mathcal{Y}, \mathcal{D})$ be a measurable space, and let $T : (\mathcal{X}, \mathcal{C}) \to (\mathcal{Y}, \mathcal{D})$ be measurable. Then $\{T(U^{(i)})\}_{i \in \mathbb{N}}$ is a Poisson point process with intensity measure $\mu \circ T^{-1}$, that is, the pushforward of μ by T.*

Example 2.1.6 (Interpretation as counting process)**.** Let $(N_t)_{t \geq 0}$ be a unit Poisson process, that is, a Lévy process where the increments $N_t - N_s$, for $0 \leq s < t$, are Poisson distributed with rate $t - s$, and let $\{\Gamma_i\}_{i \in \mathbb{N}}$ be the jump times of the counting process. Then, for disjoint intervals $(a_j, b_j]$, $j = 1, \ldots, m$, and integers r_j,

$$\begin{aligned}
&\mathbf{P}\left[\sum_{i\in\mathbb{N}} \mathbb{1}\{\Gamma_i \in (a_j, b_j]\} \;=\; r_j \qquad \text{for all } j \in \{1,\dots,m\}\right] \\
&= \mathbf{P}\left[N_{b_j} - N_{a_j} \;=\; r_j \qquad \text{for all } j \in \{1,\dots,m\}\right] \\
&= \prod_{j=1}^{m} \mathbf{P}\left[N_{b_j} - N_{a_j} \;=\; r_j\right] \\
&= \prod_{j=1}^{m} \frac{(b_j - a_j)^{r_j} \exp(-(b_j - a_j))}{r_j!}\,,
\end{aligned}$$

that is, $\{\Gamma_i\}_{i\in\mathbb{N}}$ is a PPP on the positive half-line with Lebesgue intensity $\mathrm{d}x$. Since for

$$T \;:\; (0,\infty) \;\to\; \mathbb{R}\,, \qquad z \;\mapsto\; \log(1/z)\,,$$

and $a < b$ it holds

$$\begin{aligned}
&\int_{\mathbb{R}} \mathbb{1}\{x \in (a,b]\} \exp(-x)\,\mathrm{d}x \;=\; \exp(-a) \;-\; \exp(-b) \\
&\;=\; \int_{\mathbb{R}} \mathbb{1}\{x \in T^{-1}((a,b])\}\,\mathrm{d}x\,,
\end{aligned}$$

it follows from Theorem 2.1.5 that $\{\log(1/\Gamma_i)\}_{i\in\mathbb{N}}$ is a PPP on the real line with intensity measure $\exp(-x)\mathrm{d}x$; see Figure 2.1 for an illustration.

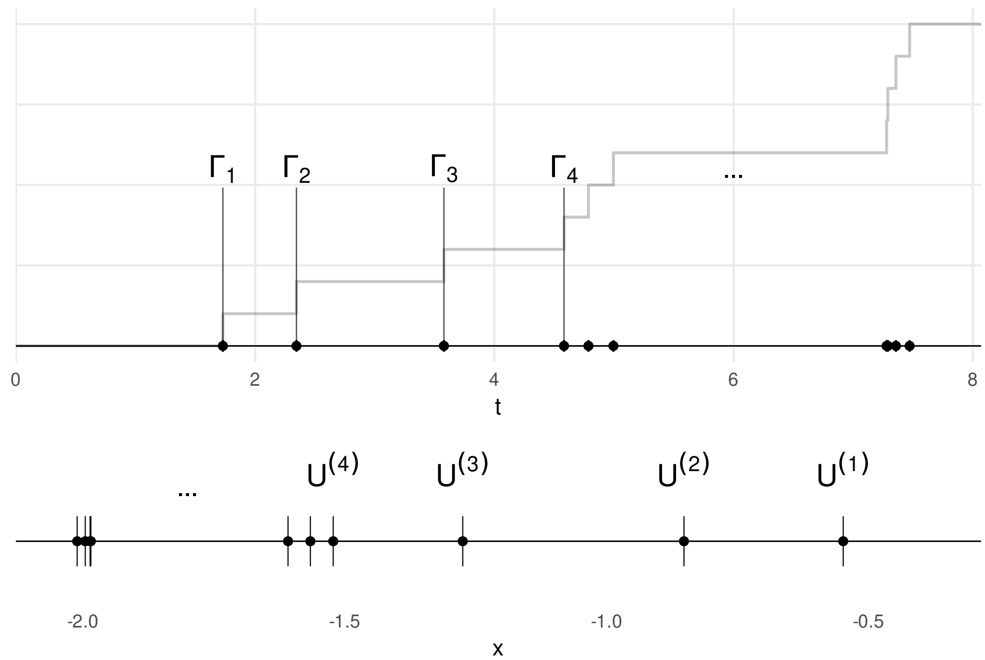

Figure 2.1: Jump times $\{\Gamma_i\}$ of a unit Poisson process and their image points $\{U^{(i)}\} = \{\log(1/\Gamma_i)\}$.

2.1.2.2 Markovian particle system

As mentioned in the last section, the points $\{U^{(i)}\}_{i\in\mathbb{N}}$ of a PPP will serve as the random initial configurations of more general particle systems. From each such configuration point an independent copy of a given stochastic process is started. The resulting collection of particle trajectories then defines the particle system. The motions of the particles that we are going to consider have the Markov property, which we are going to define next. A stochastic process $(X_t)_{t\geq 0}$ adapted to a filtration $(\mathcal{F}_t)_{t\geq 0}$ is called a time-homogeneous Markov process if there exists a family of transition kernels $(P_t)_{t\geq 0}$ such that

$$\mathbf{P}[X_t \in B \mid \mathcal{F}_s] = P_{t-s}(X_s, B) \qquad \text{almost surely}$$

for all $0 \leq s \leq t$ and Borel sets $B \subset \mathbb{R}$ [Kallenberg, 2021, Chapter 11]. In what follows, we consider such Markov processes with sample paths in the Skorokhod space $\mathbb{D}([0,\infty))$. For $x \in \mathbb{R}$, let $(X_t(x))_{t\geq 0}$ denote the process started at x, that is, $X_0(x) = x$.

Given a Markov process $(X_t)_{t\geq 0}$ with paths in the Skorokhod space $\mathbb{D}([0,\infty))$, and a collection of points $\{U^{(i)}\}_{i\in\mathbb{N}}$, we consider the collection of particles obtained by independently starting in each point $U^{(i)}$ an i.i.d. copy $(X_t^{(i)})_{t\geq 0}$ of the Markov process $(X_t)_{t\geq 0}$, that is,

$$\left\{ \left(X_t^{(i)}(U^{(i)}) \right)_{t\geq 0} \right\}_{i\in\mathbb{N}} .$$

We call this a Markovian particle system. If $\{U^{(i)}\}_{i\in\mathbb{N}}$ is a PPP on a σ-finite measure space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$, then we call this the particle system associated with the process–measure pair $((X_t)_{t\geq 0}, \mu)$, and denote it by $\mathfrak{P}((X_t)_{t\geq 0}, \mu)$.

2.1.2.3 Lévy processes

A real-valued stochastic process $(L_t)_{t\geq 0}$ is called a Lévy process if $L_0 = 0$ almost surely, it has independent and stationary increments, it is stochastically continuous, and has sample paths in $\mathbb{D}([0,\infty))$. Thus, for $0 \leq s < t$, the increment $L_t - L_s$ is independent of the past and has the same distribution as L_{t-s}. In this thesis, Lévy processes are used to describe the motion of particles in time. A particle started at position $x \in \mathbb{R}$ is moved according to

$$X_t(x) \;=\; x \;+\; L_t\,.$$

By the stationarity and independence of increments of $(L_t)_{t\geq 0}$, $(X_t)_{t\geq 0}$ is a time-homogeneous Markov process, with transition probabilities

$$\mathbf{P}[X_t \in A \mid X_0 = x] \;=\; \mathbf{P}[X_t(x) \in A] \;=\; \mathbf{P}[x + L_t \in A]\,.$$

Random time change

Definition 2.1.7 (Random clock)**.** Given a Lévy process $(L_t)_{t\geq 0}$ and a function $\alpha : \mathbb{R} \to (0,\infty)$ that is bounded and bounded away from 0, we define

$$A_x(t) \;=\; \int_0^t \alpha(L_r + x)\,\mathrm{d}r \qquad \text{and} \qquad T_x(t) \;=\; \inf\{s \geq 0 \mid A_x(s) > t\}$$

for $t \geq 0\,, x \in \mathbb{R}\,$. We call $(T_x(t))_{t\geq 0}$ the random clock, $(A_x(t))_{t\geq 0}$ the inverse random clock, and α the mass function.

In this thesis we work with the time-changed process $X_t(x) = x + L_{T_x(t)}$, which under additional assumptions is a Markov process.

Invariant measures

We say that a σ-finite measure on $\mathbb{R}$ is invariant for a Markov process $(X_t)_{t\geq 0}$ if

$$\int_{\mathbb{R}} \mathbf{P}[X_t(x) \in B]\,\mu(\mathrm{d}x) \;=\; \mu(B) \qquad \text{for all Borel sets } B \text{ and all } t \geq 0\,. \tag{2.2}$$

In this thesis we consider a Lévy process $(L_t)_{t\geq 0}$ satisfying $\mathbf{E}[\exp(L_1)] = 1$. Since $(L_t)_{t\geq 0}$ has independent and stationary increments, this immediately extends to $\mathbf{E}[\exp(L_t)] = 1$ for all $t \geq 0$. Then, for all Borel sets B and all $t \geq 0$,

$$\begin{aligned}
&\int_{\mathbb{R}} \mathbf{P}[L_t + x \in B]\exp(-x)\,\mathrm{d}x \\
&= \mathbf{E}\left[\int_{\mathbb{R}} \mathbb{1}\{L_t + x \in B\}\exp(-x)\,\mathrm{d}x\right] \\
&= \mathbf{E}\left[\int_{\mathbb{R}} \mathbb{1}\{x \in B\}\exp(-(x - L_t))\,\mathrm{d}x\right] \;=\; \mathbf{E}\left[\exp(L_t)\right]\cdot\int_B \exp(-x)\,\mathrm{d}x \\
&= \int_B \exp(-x)\,\mathrm{d}x\,,
\end{aligned} \tag{2.3}$$

that is $\mu(\mathrm{d}x) = \exp(-x)\mathrm{d}x$ is an invariant measure for $(L_t)_{t\geq 0}$. In this thesis we will study how the invariant measure of $(L_t)_{t\geq 0}$ changes under random time change.

2.1.2.4 Stationary Markovian particle systems

The notion of an invariant measure introduced above provides the link between the dynamics of an individual particle and stationarity of the corresponding particle system. We therefore turn to the notion of stationarity for Markovian particle systems and subsequently make this connection precise.

For a finite sequence of times $0 \leq t_1 < \cdots < t_\ell < \infty$, $\ell \in \mathbb{N}$, we define a finite-dimensional particle system by

$$\mathfrak{P}_{t_1,\ldots,t_\ell}((X_t)_{t\geq 0}, \mu) \;:=\; \left\{\left(X^{(i)}_{t_1}(U^{(i)}), \ldots, X^{(i)}_{t_\ell}(U^{(i)})\right)\right\}_{i\in\mathbb{N}} \tag{2.4}$$

with points in $\mathbb{R}^\ell$. We say that the full particle system $\mathfrak{P}((X_t)_{t\geq 0},\mu)$ is stationary if, for all such finite collections of time points and all $h \geq 0$, it holds that

$$\mathfrak{P}_{t_1,\ldots,t_\ell}((X_t)_{t\geq 0},\mu) \quad \text{has the same distribution as} \quad \mathfrak{P}_{t_1+h,\ldots,t_\ell+h}((X_t)_{t\geq 0},\mu)\,. \tag{2.5}$$

It will be useful to have more general transformations of the particle system. To this end, let $(E,\mathcal{E})$ be a measurable space and g be a (deterministic) measurable function from $(\mathbb{D}([0,\infty)),\mathcal{B}(\mathbb{D}([0,\infty))))$ to $(E,\mathcal{E})$. We define the g–transformed particle system by

$$\mathfrak{P}_g\left((X_t)_{t\geq 0},\mu\right) \;=\; \left\{ g\left(\left(X_t^{(i)}(U^{(i)}) \right)_{t\geq 0} \right) \right\}_{i\in\mathbb{N}} \qquad \text{with points in} \qquad E\,.$$

The following theorem is a direct consequence of Theorem 1 in Brown [1970], see also Section 3 therein.

Theorem 2.1.8. *Let $(X_t)_{t\geq 0}$ be a Markov process, $(\mathbb{R},\mathcal{B}(\mathbb{R}),\mu)$ a σ-finite measure space with $\mu(\mathbb{R})=\infty$, and g a measurable function from $(\mathbb{D}([0,\infty)),\mathcal{B}(\mathbb{D}([0,\infty))))$ to a measurable space $(E,\mathcal{E})$. Then the following holds:*

1. *The particle system $\mathfrak{P}_g((X_t)_{t\geq 0},\mu)$ defines a PPP on $(E,\mathcal{E})$ with intensity measure*

$$\mu_g(A) \;=\; \int_{\mathbb{R}} \mathbf{P}\left[g\left((X_t(x))_{t\geq 0}\right)\in A\right]\,\mu(\mathrm{d}x) \qquad \textit{for } A\in\mathcal{E}\,.$$

2. *The particle system $\mathfrak{P}((X_t)_{t\geq 0},\mu)$ is stationary if and only if the process–measure pair $((X_t)_{t\geq 0},\mu)$ is invariant.*

The last theorem shows that Markovian particle systems based on invariant process–measure pairs are stationary. In this thesis we obtain, by random time change, from an invariant baseline pair $((L_t)_{t\geq 0},\exp(-x)\mathrm{d}x)$ a perturbed invariant pair. We then study the resulting stationary Markovian particle system of the time-changed particles.

2.2 Extreme Value Theory

As outlined in the introduction, EVT is commonly studied through the PoT approach or the Block Maxima approach. We now introduce the methodological background for both approaches in more detail.

2.2.1 Peaks-over-Threshold

2.2.1.1 Regular variation

Univariate

Let F be a distribution function with upper endpoint $+\infty$, and let $\overline{F}=1-F$ denote its tail function. For $\alpha>0$, we say that $\overline{F}$ is regularly varying at $+\infty$

with index $-\alpha$, written $\overline{F} \in \mathrm{RV}_{-\alpha}$, if, for every $x > 0$,

$$\frac{\overline{F}(u \cdot x)}{\overline{F}(u)} \;\to\; x^{-\alpha} \qquad \text{as } u \to \infty\,. \tag{2.6}$$

A distribution function F, or random variable X with distribution function F, satisfying $\overline{F} \in \mathrm{RV}_{-\alpha}$ is said to have a regularly varying tail. In particular, such a distribution is heavy-tailed.

Example 2.2.1 (Pareto distribution)**.** A canonical example of a regularly varying tail function is given by the Pareto distribution with shape parameter $\alpha > 0$, denoted by Pareto(α), whose tail function is

$$\overline{F}_{\mathrm{Pareto}(\alpha)}(x) \;=\; \begin{cases} 1, & x < 1, \\ x^{-\alpha}, & x \geq 1. \end{cases}$$

In this case, the limiting relation in (2.6) holds as an identity for every $x > 0$ and sufficiently large u. Indeed, if $u \geq 1$ and $u \cdot x \geq 1$,

$$\frac{\overline{F}_{\mathrm{Pareto}(\alpha)}(u \cdot x)}{\overline{F}_{\mathrm{Pareto}(\alpha)}(u)} \;=\; \frac{(ux)^{-\alpha}}{u^{-\alpha}} \;=\; x^{-\alpha}\,.$$

Remark 2.2.2 (Regular variation beyond tail functions)**.** Although regular variation has been introduced above only for tail functions with negative index, it is a general asymptotic concept and is not intrinsically tied to heavy-tailed distributions or to tail functions [Bingham et al., 1987]. The convergence in (2.6) may be defined for any eventually positive function f instead of $\overline{F}$ and for any index $\rho \in \mathbb{R}$ instead of $-\alpha$. For example, the logarithm is regularly varying with $\rho = 0$, called slowly varying, since,

$$\frac{\log(u \cdot x)}{\log(u)} \;=\; 1 \;+\; \frac{\log(x)}{\log(u)} \;\to\; 1 \qquad \text{for all } x > 0 \text{ as } u \to \infty\,.$$

Example 2.2.3 (Probability density function of the t-distribution)**.** Let f_ν be the probability density function of the t-distribution with degrees of freedom $\nu > 0$, that is,

$$f_\nu(t) \;=\; \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 \;+\; \frac{t^2}{\nu}\right)^{-(\nu+1)/2}, \qquad t \in \mathbb{R}\,.$$

Then, for $x > 0$, as $t \to \infty$,

$$\frac{f_\nu(t \cdot x)}{f_\nu(t)} \;=\; \left(\frac{1 \;+\; \frac{t^2}{\nu}}{1 \;+\; \frac{(t \cdot x)^2}{\nu}}\right)^{(\nu+1)/2} \;=\; \left(\frac{\nu/t^2 \;+\; 1}{\nu/t^2 \;+\; x^2}\right)^{(\nu+1)/2} \;\to\; x^{-(\nu+1)}\,,$$

that is, $f_\nu \in \mathrm{RV}_{-(\nu+1)}$. By Karamata's theorem [De Haan and Ferreira, 2006, Theorem B.1.5], regular variation of the density also determines the regular variation of the corresponding tail function. In this case, the upper tail function $\overline{F}_\nu$ has index of regular variation $-\nu$. Notice that, as $\nu \to \infty$, the t-distribution converges to the standard normal distribution, which is no longer regularly varying; see Figure 1.2 for a comparison of stationary long memory linear time series with Gaussian and t-distributed innovations.

Stationary sequences

For $\alpha > 0$, a real-valued stationary sequence $(X_t)_{t\in\mathbb{Z}}$ is said to be regularly varying with tail index α if $|X_0|$ is regularly varying with index $-\alpha$, and if, for every integer $h \in \mathbb{N}_0$, there exists a non-null Borel measure μ_h on $\mathbb{R}^{1+h} \setminus \{\mathbf{0}\}$ that is finite on Borel sets bounded away from $\mathbf{0}$, such that

$$\frac{\mathbf{P}[[X_0, \ldots, X_h] \in u \cdot A]}{\mathbf{P}[|X_0| > u]} \;\to\; \mu_h[A] \qquad \text{as } u \to \infty \tag{2.7}$$

for all Borel sets $A \subset \mathbb{R}^{1+h} \setminus \{\mathbf{0}\}$ bounded away from $\mathbf{0}$ satisfying $\mu_h[\partial A] = 0$ [Mikosch and Wintenberger, 2024, Definition 4.1.1]. The measures μ_h, $h \in \mathbb{N}_0$, are called $(h+1)$-dimensional tail measures of (X_t). Choosing, for $T \in \mathbb{N}$, Borel sets $A, B \subset \mathbb{R}^T$ bounded away from zero, the quantity

$$\gamma_{A,B}(h) \;:=\; \mu_{T-1+h}[(A \times \mathbb{R}^h) \cap (\mathbb{R}^h \times B)]\,, \qquad h \in \mathbb{N}_0$$

is called the extremogram of the time series (X_t) [Davis and Mikosch, 2009].

Example 2.2.4 (Tail measures of heavy-tailed long memory linear time series)**.** Let (X_t) be a causal linear process as in Section 2.1.1, and let its innovation law ε be regularly varying, for example, with tail index $\nu \in (1, 2)$ as in Assumption 2(ii). Let $B \subset \mathbb{R}$ be a Borel set bounded away from 0, whose boundary has Lebesgue measure 0. Then, by regular variation and symmetry of ε,

$$\frac{\mathbf{P}[\varepsilon \in u \cdot B]}{\mathbf{P}[|\varepsilon| > u]} \;\to\; \mu_\varepsilon[B] \;:=\; \int_B \frac{\nu}{2}|z|^{-1-\nu}\,\mathrm{d}z \qquad \text{as } u \to \infty.$$

Let $A \subset \mathbb{R}^{1+h} \setminus \{\mathbf{0}\}$, $h \in \mathbb{N}_0$. By Proposition 5.2.8 in Mikosch and Wintenberger [2024], the $(h+1)$-dimensional tail measure of (X_t) evaluated at A is

$$\mu_h[A] \;=\; \frac{1}{\sum_{i=0}^{\infty} |a_i|^\nu} \sum_{j=-h}^{\infty} \mu_\varepsilon[\{x \in \mathbb{R} \mid x \cdot [a_j, \ldots, a_{j+h}] \in A\}]\,,$$

and the stationary sequence (X_t) is regularly varying with tail index ν.

2.2.1.2 Peaks-over-Threshold estimators

Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary regularly varying sequence with tail index $\alpha > 0$, and assume additionally that $\overline{F}_{X_0} \in \mathrm{RV}_{-\alpha}$. In the univariate case, regular variation of the tail function $\overline{F}_{X_0}$ of X_0 implies, for $x \geq 1$,

$$\mathbf{P}[X_0 > u \cdot x \mid X_0 > u] \;=\; \frac{\overline{F}_{X_0}(u \cdot x)}{\overline{F}_{X_0}(u)} \;\to\; x^{-\alpha} \qquad \text{as } u \to \infty.$$

Equivalently,

$$X_0/u \mid X_0 > u \;\to_d\; Z \sim \text{Pareto}(\alpha) \qquad \text{(see Example 2.2.1)} \tag{2.8}$$

as $u \to \infty$. This motivates studying observations relative to a high threshold, so-called Peaks-over-Threshold (PoT). In all the following examples, (u_n) is a sequence of thresholds growing to ∞ as $n \to \infty$.

Example 2.2.5 (Empirical tail function)**.** Since $\overline{F}_{X_0}$ is usually not known, a natural estimator of the tail function is

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{X_i > u_n\}\,.$$

This is often used as a normalization estimator for more complicated PoT estimators.

Example 2.2.6 (Hill estimator)**.** The convergence in (2.8) together with Potter's bounds [De Haan and Ferreira, 2006, (B.1.19)] implies

$$\mathbf{E}[\log(X_0/u) \mid X_0 > u] \;\to\; \int_1^\infty \log(x)\alpha x^{-1-\alpha}\,\mathrm{d}x \;=\; \frac{1}{\alpha}\,.$$

Therefore, a natural estimator for the reciprocal tail index $1/\alpha$ is

$$\frac{\frac{1}{n}\sum_{i=1}^{n}\log(X_i/u_n)\mathbb{1}\{X_i > u_n\}}{\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{X_i > u_n\}}\,.$$

This is the so-called Hill estimator.

Example 2.2.7 (Empirical tail measures)**.** For a regularly varying stationary sequence (X_t), the convergence to tail measures in (2.7) motivates the natural estimator

$$\frac{\frac{1}{n}\sum_{t=1}^{n-h}\mathbb{1}\{[X_t,\ldots,X_{t+h}] \in u_n\cdot A\}}{\frac{1}{n}\sum_{t=1}^{n}\mathbb{1}\{|X_t| > u_n\}}$$

of the $(h+1)$-dimensional tail measure of (X_t) evaluated at a set $A \subset \mathbb{R}^{1+h}\setminus\{\mathbf{0}\}$ that is bounded away from $\mathbf{0}$ and satisfies $\mu_h[\partial A] = 0$.

2.2.1.3 Random thresholds

In applications the threshold sequence (u_n) is usually not fixed in advance. Instead one chooses sufficiently high upper order statistics. Let $k = k_n$ be an intermediate sequence satisfying

$$k_n \;\to\; \infty\,, \qquad k_n/n \;\to\; 0$$

and let

$$X_{1:n} \;\le\; \cdots \;\le\; X_{n:n}$$

be the order statistics of the sample $X_1,\ldots,X_n$. The deterministic threshold u_n in the estimators is then replaced by the random threshold

$$\widehat{u}_n \;=\; X_{n-k:n}\,.$$

2.2.2 Block Maxima

2.2.2.1 Univariate theory

Let $Z_1, Z_2, \ldots$ be independent copies of a real-valued random variable Z, for example, block maxima. Consider the partial maxima

$$M_n \;=\; \max_{i=1,\ldots,n} Z_i\,, \qquad n \in \mathbb{N}\,.$$

We say that Z belongs to the max-domain of attraction of a non-degenerate random variable ζ, written $Z \in \mathrm{MDA}(\zeta)$, if there exist sequences of constants $(a_n)_{n\in\mathbb{N}} \subset (0,\infty)$ and $(b_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ such that

$$\frac{M_n - b_n}{a_n} \qquad \text{converges in distribution to} \qquad \zeta \qquad \text{as } n \to \infty\,.$$

Any such limit is max-stable in the sense that there exist sequences $(\alpha_n)_{n\in\mathbb{N}} \subset (0,\infty)$ and $(\beta_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ such that

$$\frac{\max_{i=1,\ldots,n} \zeta_i - \beta_n}{\alpha_n} \qquad \text{has the same distribution as} \qquad \zeta \qquad \text{for all } n \in \mathbb{N}\,,$$

where $\zeta_1, \zeta_2, \ldots$ are independent copies of ζ. By the extremal types theorem [De Haan and Ferreira, 2006, Theorem 1.1.3], ζ has a generalized extreme value distribution. In this thesis we work with the standard Gumbel law

$$\Lambda(x) \;=\; \exp(-\exp(-x))\,, \qquad x \in \mathbb{R}\,.$$

If $\zeta \sim \Lambda$, by independence,

$$\begin{aligned}
&\mathbf{P}\left[\max_{i=1,\ldots,n} \zeta_i - \log(n) \le x\right] \\
&\;=\; \mathbf{P}\left[\forall_{i=1,\ldots,n}\ \zeta_i \le x + \log(n)\right] \;=\; (\Lambda(x + \log(n)))^n \\
&\;=\; \exp(-n \cdot \exp(-(x + \log(n)))) \;=\; \exp(-\exp(-x)) \\
&\;=\; \Lambda(x)\,,
\end{aligned}$$

so in this case, the constants are $\alpha_n = 1$ and $\beta_n = \log(n)$, and max-stability means

$$\max_{i=1,\ldots,n} \zeta_i - \log(n) \qquad \text{has the same distribution as} \qquad \zeta\,.$$

Example 2.2.8 (Tail equivalent Gumbel)**.** Let $\alpha : \mathbb{R} \to (0,\infty)$ be a bounded function, bounded away from 0, and define

$$F_\alpha(x) \;=\; \exp\left(-\int_x^\infty \alpha(s)\exp(-s)\,\mathrm{d}s\right)\,.$$

If $\alpha \equiv 1$, then

$$F_1(x) \;:=\; \exp\left(-\int_x^\infty \exp(-s)\,\mathrm{d}s\right) \;=\; \exp(-\exp(-x)) \;=\; \Lambda(x)\,,$$

that is, F_1 is the Gumbel distribution function. For nonconstant α satisfying $\lim_{x\to\infty} \alpha(x) = 1$, we generally have $F_\alpha \neq \Lambda$, but

$$\begin{aligned}(F_\alpha(x + \log(n)))^n &= \exp\left(-n \int_{x+\log(n)}^{\infty} \alpha(s) \exp(-s)\, \mathrm{d}s\right)\\ &= \exp\left(-\int_{x}^{\infty} \alpha(s + \log(n)) \exp(-s)\, \mathrm{d}s\right)\\ &\to \exp\left(-\int_{x}^{\infty} \exp(-s)\, \mathrm{d}s\right) = \Lambda(x)\,,\end{aligned}$$

by the boundedness of α and the dominated convergence theorem. Therefore, on the level of distribution functions, $F_\alpha \in \mathrm{MDA}(\Lambda)$. This example is relevant for the stationary stochastic processes constructed in Scheffel [2026].

2.2.2.2 Max-stable processes

Let $(Z_t)_{t\geq 0}$ be a real-valued stationary stochastic process and consider, for independent copies $(Z_t^{(1)})_{t\geq 0}, (Z_t^{(2)})_{t\geq 0}, \ldots$ of $(Z_t)_{t\geq 0}$, the pointwise partial maxima

$$M_n(t) = \max_{i=1,\ldots,n} Z_t^{(i)}\,.$$

In analogy to the univariate case, we say that $(Z_t)_{t\geq 0}$ belongs to the MDA of a non-degenerate process $\zeta = (\zeta(t))_{t\geq 0}$ with Gumbel margins, written $(Z_t)_{t\geq 0} \in \mathrm{MDA}(\zeta)$, if there exist sequences of constants $(a_n)_{n\in\mathbb{N}} \subset (0,\infty)$ and $(b_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ such that, for any finite collection of time points $t_1, \ldots, t_k \geq 0$, $k \in \mathbb{N}$,

$$\left[(\max_{i=1,\ldots,n} Z_{t_1}^{(i)} - b_n)/a_n\,, \ldots\,, (\max_{i=1,\ldots,n} Z_{t_k}^{(i)} - b_n)/a_n\right]$$

converges in distribution to

$$[\zeta(t_1), \ldots, \zeta(t_k)] \qquad \text{as } n \to \infty\,.$$

Any such limit is max-stable in the sense that

$$\left[\max_{i=1,\ldots,n} \zeta^{(i)}(t_1) - \log(n)\,, \ldots\,, \max_{i=1,\ldots,n} \zeta^{(i)}(t_k) - \log(n)\right]$$

has the same distribution as

$$[\zeta(t_1), \ldots, \zeta(t_k)]\,,$$

where $\zeta^{(1)}, \zeta^{(2)}, \ldots$ are independent copies of ζ.

Spectral representation

While in the univariate case, there exists a complete characterization of max-stable laws in terms of generalized extreme value distributions, the characterization for max-stable processes is more difficult. It follows from the work of de Haan [1984] that all sample-continuous stationary max-stable processes with marginal distribution Λ admit a spectral representation of the form

$$\zeta \;=\; \max_{i\in\mathbb{N}} \left(U^{(i)} + X^{(i)}\right) ,$$

where $\{U^{(i)}\}_{i\in\mathbb{N}}$ form a Poisson point process (see Section 2.1.2.1) on the real line with intensity $\exp(-x)\mathrm{d}x$, $X^{(i)}$, $i \in \mathbb{N}$ are i.i.d. copies of a stochastic process $X = (X_t)_{t\geq 0}$ which takes values in $\mathbb{R}\cup\{-\infty\}$ and satisfies the condition $\mathbf{E}[\exp(X_t)] = 1$ for all $t \geq 0$, and where $\{U^{(i)}\}_{i\in\mathbb{N}}$ and $\{X^{(i)}\}_{i\in\mathbb{N}}$ are independent.

2.2.2.3 Maxima over stationary Markovian particle systems

For a single particle $(X_t(U))_{t\geq 0}$, stationarity holds if and only if there exists an invariant probability measure μ of the process $(X_t)_{t\geq 0}$ and $U \sim \mu$. Extreme value analysis naturally shifts the viewpoint from a single particle to a system of infinitely many particles, where the starting points $\{U^{(i)}\}_{i\in\mathbb{N}}$ form a Poisson point process (PPP). For a distribution function F with density F', it is, for example, possible to choose the PPP $\{U^{(i)}\}_{i\in\mathbb{N}}$ such that $F(x) = \mathbf{P}\left[\max_{i\in\mathbb{N}} U^{(i)} \leq x\right]$, for $x \in \mathbb{R}$. If the arising Markovian particle system is stationary, then the maximum process

$$Z_t \;=\; \max_{i\in\mathbb{N}} X_t^{(i)}(U^{(i)}) , \qquad t \geq 0 , \tag{2.9}$$

is stationary as well, with marginal distribution function F. Such a process (regardless of stationarity) is called max-infinitely divisible (max-id) in the sense that it is the componentwise maximum over a PPP with points in $\mathbb{D}([0,\infty))$. The notion of max-id is analogous to the notion of infinite divisibility for sums rather than maxima. It follows from Theorem 2.1.8 that the Markovian particle system (and the max-id process) is stationary if and only if $\mu(\mathrm{d}x) = F'(x)/F(x)\mathbb{1}\{F(x) > 0\}\mathrm{d}x$ is invariant for $(X_t)_{t\geq 0}$ in the sense of (2.2). In this setting the invariant measure μ is necessarily infinite, so that, unlike in the single-particle case discussed above, it cannot be normalized to a probability measure.

A possible choice for F is the Gumbel distribution function Λ (see Section 2.2.2.1 for the definition of Λ) with $\mu(\mathrm{d}x) \;=\; \exp(-x)\mathrm{d}x$. Any invariant process–measure pair $((X_t)_{t\geq 0}, \exp(-x)\mathrm{d}x)$ in the sense of (2.2) then yields a stationary max-id process with Gumbel margins, as in (2.9). Brown and Resnick [1977] were the first to construct such a process, based on the observation that this specific choice of μ is invariant for the drifted Brownian motion $(B_t - t/2)_{t\geq 0}$. The resulting process ζ_{BR}, called Brown–Resnick process, is stationary with Gumbel margins:

$$\zeta_{\mathrm{BR}}(t) \;:=\; \max_{i\in\mathbb{N}} \left(B_t^{(i)} - t/2 \;+\; U^{(i)}\right) , \qquad t \geq 0 . \tag{2.10}$$

Even though the single particles $B_t^{(i)} - t/2 + U^{(i)}$ are nonstationary (they converge to $-\infty$ for $t \to \infty$), the particle system $\{(B_t^{(i)} - t/2 + U^{(i)})_{t \geq 0}\}_{i \in \mathbb{N}}$, and the max-id process ζ_{BR} are stationary. Brown and Resnick also showed that ζ_{BR} is not only max-id, but even max-stable.

Max-stable processes obtained from systems of Lévy particles beyond the drifted Brownian motion have been studied in Stoev [2008], where they show stationarity and Gumbel margins of the process

$$\zeta_L(t) \;:=\; \max_{i \in \mathbb{N}} \left(L_t^{(i)} \;+\; U^{(i)} \right) , \qquad t \geq 0 \, ,$$

if the Lévy process $(L_t)_{t \geq 0}$ satisfies $\mathbf{E}[\exp(L_1)] = 1$. We refer to such processes as Lévy–Brown–Resnick processes. By (2.3), the Lévy process $(L_t)_{t \geq 0}$ has invariant measure $\exp(-x)\mathrm{d}x$, so that Theorem 2.1.8 yields stationarity.

Chapter 3

Main Results and Outlook

3.1 Main Results

In this section, we describe the main results of this thesis, which are contained in the following papers:

1. Ioan Scheffel, Marco Oesting, and Gilles Stupfler, *Central limit theory for Peaks-over-Threshold partial sums of long memory linear time series*, published in *Stochastic Processes and their Applications*, 2026.
   Reference: Scheffel et al. [2026a]
   Link: `https://doi.org/10.1016/j.spa.2026.104929`
   Link to arXiv version: `https://doi.org/10.48550/arXiv.2506.20789`
   This article (specifically Version 3 on arXiv, which corresponds to the accepted version of the published paper) is reproduced in Chapter 5.

2. Ioan Scheffel, Marco Oesting, and Gilles Stupfler, *Central limit theory for serial tail dependence estimators in heavy-tailed long memory linear time series*, submitted to *Probability Theory and Related Fields*, preprint available on arXiv, 2026.
   Reference: Scheffel et al. [2026b]
   Link to arXiv version: `https://doi.org/10.48550/arXiv.2608.10944`
   This article (specifically Version 1 on arXiv) is reproduced in Chapter 6.

3. Ioan Scheffel, *Maxima of stationary systems of randomly time-changed Lévy particles*, submitted to *Stochastic Processes and their Applications*, preprint available on arXiv, 2026.
   Reference: Scheffel [2026]
   Link to arXiv version: `https://doi.org/10.48550/arXiv.2604.11434`
   This article (specifically Version 2 on arXiv) is reproduced in Chapter 7.

Chapter 4 contains a statement on the contributions of the author of this thesis to the articles above. The following three subsections will present the main results of the respective articles.

3.1.1 Central limit theory for Peaks-over-Threshold partial sums of long memory linear time series

We consider long memory linear time series as introduced in Section 2.1.1. The main difficulty in applying existing limit theory for long memory linear time series to PoT statistics is that the transformation of the data itself depends on the sample size. Classical central limit theory for subordinated long memory linear processes studies partial sums $\sum_{t=1}^{n} G(X_t)$ of a fixed subordination function G; see Section 2.1.1.1 for an overview. In the PoT setting, however, one has a sequence of functions (G_n) depending on a threshold $u_n \to \infty$, for example,

$$G_n(x) \;=\; \mathbb{1}\{x > u_n\} \qquad \text{or} \qquad G_n(x) \;=\; \log(x/u_n)\mathbb{1}\{x > u_n\}\,; \tag{3.1}$$

see Section 2.2.1.2 for an overview of PoT statistics. This dependence on sample size changes the rate of convergence and the asymptotic scale and is the source of the main phenomenon discovered in this work.

As we discussed in Section 2.1.1, long memory linear processes are well defined under Assumptions 1 and 2. The next assumption on the sequence of subordination functions (G_n) is more general than the two examples in (3.1) since it allows also for polynomial growth in x.

Assumption 3 (Regularity of G_n)**.** It holds that

$$|G_n(x)| \;\lesssim\; (1+|x|)^{\gamma_G}\mathbb{1}\{x > u_n\}\,,$$

for some $\gamma_G \geq 0$ and a deterministic sequence (u_n) of thresholds with $u_n \to \infty$ as $n \to \infty$. Here, $\lesssim$ denotes inequality up to a multiplicative constant independent of n and x.

The next theorem is the main result.

Theorem 3.1.1 (Central limit theorem)**.** *Assume that the following holds:*

(i) Assumptions 1, 2, and 3 with $\gamma_G \leq \gamma_0$, where γ_0 depends on α and d (see Equation (4) in Scheffel et al. [2026a]).

(ii) $\mathbf{E}[G_n(X_0)] \neq 0$ and $G'_{\infty,n}(0) = \frac{\mathrm{d}}{\mathrm{d}y}\mathbf{E}[G_n(X_0+y)]\Big|_{y=0} \neq 0$, for sufficiently large n

(iii) for some $\delta > 0$ it holds that $n^{\kappa_0+\delta-d-1/\alpha}/G'_{\infty,n}(0) \to 0$ as $n \to \infty$, where κ_0 depends on α and d (see Equation (5) in Scheffel et al. [2026a]).

Then, as $n \to \infty$,

$$n^{1-(d+1/\alpha)} \cdot \frac{\mathbf{E}[G_n(X_0)]}{G'_{\infty,n}(0)} \cdot \frac{1}{n}\sum_{t=1}^{n}\left(\frac{G_n(X_t)}{\mathbf{E}[G_n(X_0)]} \;-\; 1\right) \;\xrightarrow{\mathrm{d}}\; Z_\alpha,$$

where Z_α is a symmetric α-stable random variable defined in Theorem 2.1.2.

Remark 3.1.2 (On the rate of convergence in Theorem 3.1.1)**.** Since we consider G_n rather than fixed G, the rate of convergence in Theorem 3.1.1 is also governed by the ratio of $\mathbf{E}[G_n(X_0)]$ and $G'_{\infty,n}(0)$. This can lead to markedly different rates, which is best illustrated by considering again $G_n(x) = \mathbb{1}\{x > u_n\}$, where

$$\frac{\mathbf{E}[G_n(X_0)]}{G'_{\infty,n}(0)} = \frac{\mathbf{P}[X_0 > u_n]}{f_{X_0}(u_n)}\,.$$

For Gaussian innovations, assuming $\mathrm{Var}[X_0] = 1$, this becomes

$$\frac{\mathbf{P}[X_0 > u_n]}{f_{X_0}(u_n)} \sim \frac{1}{u_n} \to 0\,,$$

which is in line with the intuition that the total rate should slow down when considering only extremes. For $S\alpha S$ innovations with $\alpha \in (1,2)$ we get

$$\frac{\mathbf{P}[X_0 > u_n]}{f_{X_0}(u_n)} \sim \frac{u_n}{\alpha} \to \infty\,,$$

which is surprising because the total rate speeds up when looking only at extremes.

We illustrate this phenomenon with an application to heavy tails. The light-tailed case, which features new phenomena by itself, can also be found in the paper.

Theorem 3.1.3 (Applications to heavy tails – deterministic thresholds)**.** *Let Assumptions 1 and 2 hold, and assume that* $f_{X_0} \in \mathrm{RV}_{-\nu-1}$ *with* $\nu > 1$*. If* $u_n \to \infty$ *is such that for some* $\delta > 0$, $u_n = \mathrm{o}(n^{(d+1/(\nu\wedge 2)-\kappa_0-\delta)/(\nu+1)})$*. Then, as* $n \to \infty$,

$$n^{1-(d+1/(\nu\wedge 2))} u_n \cdot \left(\frac{\sum_{t=1}^{n} \mathbb{1}\{X_t > u_n\}}{n\mathbf{P}[X_0 > u_n]} - 1 \right) \xrightarrow{\mathrm{d}} \nu Z_{\nu\wedge 2}$$

and

$$n^{1-(d+1/(\nu\wedge 2))} u_n \cdot \left(\frac{\sum_{t=1}^{n} \log(X_t/u_n)\mathbb{1}\{X_t > u_n\}}{n\mathbf{P}[X_0 > u_n]} - \mathbf{E}[\log(X_0/u_n) \mid X_0 > u_n] \right) \xrightarrow{\mathrm{d}} \frac{\nu}{\nu+1} Z_{\nu\wedge 2}.$$

As we discussed in Section 2.2.1.3, in statistical practice the choice of the threshold u_n is crucial, and it is common to have random instead of deterministic thresholds, for example, sample quantiles. The next result, which is the counterpart of Theorem 3.1.3 for random thresholds, is obtained by proving a separate derandomization device; see the explanation before Corollary 3.10 in Scheffel et al. [2026a].

Theorem 3.1.4 (Application to heavy tails – random thresholds)**.** *Let the Assumptions of Theorem 3.1.3 hold. Then, with* $k = \lfloor n\mathbf{P}[X_0 > u_n] \rfloor$*, it holds that, as* $n \to \infty$*,*

$$\begin{aligned} & n^{1-(d+1/(\nu\wedge 2))} q_{X_0}\left(1-\frac{k}{n}\right) \\ & \cdot \left(\frac{\sum_{i=1}^{k}\log(X_{n-i+1:n}/X_{n-k:n})}{n\mathbf{P}[X_0>u_n]} - \mathbf{E}[\log(X_0/u_n) \mid X_0 > u_n]\right) \\ & \xrightarrow{\mathrm{d}} \frac{1}{\nu+1} Z_{\nu\wedge 2}\,. \end{aligned}$$

Remark 3.1.5 (Comparison with known results in the literature)**.** As in the case of the empirical tail function in the first display of Theorem 3.1.3, the Hill estimators with deterministic thresholds (second display of Theorem 3.1.3) and with random thresholds (Theorem 3.1.4) show faster speeds compared with classical convergence rates in long memory settings. In the deterministic threshold case (Theorem 3.1.3), the Hill estimator shows the same increased speed as the empirical tail function, namely, with the extra factor u_n. In the random threshold case (Theorem 3.1.4), the factor of increase $q_{X_0}(1-k/n)$ may be seen as the deterministic counterpart of the random threshold $X_{n-k:n}$, since $X_{n-k:n}/q_{X_0}(1-k/n) \to_{\mathbf{P}} 1$. It stands in sharp contrast with what is observed in i.i.d. and short memory settings.

3.1.2 Central limit theory for serial tail dependence estimators in heavy-tailed long memory linear time series

We again consider long memory linear time series as introduced in Section 2.1.1. Fix $T \in \mathbb{N}$. For $s > 0$ and $A \subset \mathbb{R}^T \setminus \{\mathbf{0}\}$, denote the T-dimensional empirical tail measure (see Example 2.2.7) and its expectation by

$$\widehat{\mathbf{P}}_n(s, A) \;:=\; \frac{1}{n-T+1} \sum_{t=1}^{n-T+1} \frac{\mathbb{1}\{(X_t,\ldots,X_{t+T-1}) \in u_n \cdot s \cdot A\}}{\mathbf{P}[|X_0| > u_n]}$$

and

$$\mathbf{P}_n(s, A) \;:=\; \frac{\mathbf{P}\left[(X_t,\ldots,X_{t+T-1}) \in u_n \cdot s \cdot A\right]}{\mathbf{P}[|X_0| > u_n]}\,.$$

Theorem 3.1.6 (Empirical tail measure with deterministic thresholds)**.** *Let Assumptions 1, 2(ii) hold, let* $f_{X_0} \in \mathrm{RV}_{-1-\nu}$*, let* $A_0, \ldots, A_h$*,* $h \in \mathbb{N}_0$*, be bounded away from 0, and let* $u_n \to \infty$ *with* $u_n = o(n^{(d+1/(\nu\wedge 2)-\kappa_0-\delta)/(\nu+1)})$ *for some* $\delta > 0$*, where* κ_0 *is as in Theorem 3.1.1(iii). If* $A_0, \ldots, A_h$ *satisfy additional technical assumptions, then*

$$n^{1-d-1/\alpha} u_n \left[\widehat{\mathbf{P}}_n(1, A_i) \;-\; \mathbf{P}_n(1, A_i)\right]_{i=0,\ldots,h} \to_d Z_\alpha v_1 \qquad \text{as } n \to \infty\,,$$

where $v_1 \in \mathbb{R}^{h+1}$ *is a deterministic vector (see Theorem 4.2 in Scheffel et al. [2026b]).*

Remark 3.1.7 (Increased speed in the extreme value setting)**.** Compared to standard central limit theory for long memory linear time series, the increased speed $n^{1-d-1/\alpha}u_n$ has already been observed in the univariate setting of Scheffel et al. [2026a], see Remark 3.1.2.

Theorem 3.1.8 (Empirical tail measure with sample quantiles as thresholds)**.** *Let Assumptions 1, 2(ii) hold, let $f_{X_0} \in \mathrm{RV}_{-1-\nu}$ and $(x \mapsto x \cdot f_\varepsilon(x)) \in L^1(\mathbb{R})$, let $A_0, \ldots, A_h$, $h \in \mathbb{N}_0$, be bounded away from 0, and let $u_n \to \infty$ with $u_n = o(n^{(d+1/(\nu\wedge 2)-\kappa_0-\delta)/(\nu+1)})$ for some $\delta > 0$, where κ_0 is as in Theorem 3.1.1(iii). If $A_0, \ldots, A_h$ satisfy additional technical assumptions, then, for $k = \lfloor n \cdot \mathbf{P}[X_0 > u_n] \rfloor$,*

$$\begin{aligned} &n^{1-d-1/\alpha} q_{X_0}\left(1 - \frac{k}{n}\right)\left[\widehat{\mathbf{P}}_n(X_{n-k:n}/u_n, A_i) \;-\; \mathbf{P}_n(1, A_i)\right]_{i=0,\ldots,h} \\ &\to_d \; Z_\alpha v_2 \qquad \text{as } n \to \infty\,, \end{aligned}$$

where $v_2 \in \mathbb{R}^{h+1}$ is a deterministic vector, possibly different from v_1 (see Theorem 4.8 in Scheffel et al. [2026b]).

Remark 3.1.9 (Different limiting behavior for random and deterministic thresholds)**.** It may happen that $v_2 = \mathbf{0}$ while $v_1 \neq \mathbf{0}$. Then, in the setting of deterministic thresholds (Theorem 3.1.6), we get a non-degenerate limit, while sample quantiles as thresholds (Theorem 3.1.8) give a degenerate limit at the same rate.

3.1.3 New class of processes in the MDA

As we discussed in Section 2.2.2.3, all specific examples of stationary max-id processes in the literature, constructed from nonstationary single particles, are (up to marginal transformations) max-stable. The goal of this work is to generalize this construction by allowing for random time change of the underlying Lévy particles. Starting from a Lévy–Brown–Resnick process ζ_L, we construct infinite-dimensional classes $\{Z^\alpha\}$ of stationary max-id processes. These classes are indexed by admissible mass functions α that induce state-dependent time changes of the underlying Lévy particles; see Section 2.1.2.3 for background. In contrast to $\alpha \equiv 1$, which yields the Lévy–Brown–Resnick process $\zeta_L = Z^1$, nonconstant mass functions α allow for variability in the dependence structure of the max-id process that goes beyond the max-stable setting. We then use this variability in α to explore the extent of the so-called max-domain of attraction (MDA) of a given Lévy–Brown–Resnick process ζ_L; for background on MDA see Section 2.2.2.2.

The first part of the following assumption is needed to obtain existence of the time-changed process that we are going to consider, while the second part of the assumption is only needed for the convergence result in Theorem 3.1.11.

Assumption 4 (Assumption on mass function)**.** The mass function α is continuous and satisfies

$$0 \;<\; \underline{\alpha} \;:=\; \inf_{z\in\mathbb{R}} \alpha(z) \;\leq\; \sup_{z\in\mathbb{R}} \alpha(z) \;=:\; \overline{\alpha} \;<\; \infty \qquad \text{and} \qquad \lim_{z\to\infty} \alpha(z) \;=\; 1\,.$$

Theorem 3.1.10 (Existence and stationarity of max-id processes)**.** *Let α be a mass function satisfying Assumption 4, let $(L_t)_{t\geq 0}$ be a Lévy process satisfying $\mathbf{E}[\exp(L_1)] = 1$, let $(X_t)_{t\geq 0}$ be the corresponding time-changed process, and consider the process $(Z_t)_{t\geq 0}$ defined by*

$$Z_t \;=\; \max \mathfrak{P}_t((X_t)_{t\geq 0}, \alpha(x)\exp(-x)\mathrm{d}x)\,.$$

Then the process $(Z_t)_{t\geq 0}$ has paths in $\mathbb{D}([0,\infty))$ and is stationary. If, furthermore, $(L_t)_{t\geq 0}$ has continuous paths, then $(Z_t)_{t\geq 0}$ has paths in $\mathbb{C}([0,\infty))$.

Theorem 3.1.11 (max-domain of attraction of the constructed max-id process)**.** *Let α be a mass function satisfying Assumption 4, let $(L_t)_{t\geq 0}$ be a Lévy process satisfying $\mathbf{E}[\exp(L_1)] = 1$, and let $(X_t)_{t\geq 0}$ be the corresponding time-changed process. Consider the max-id process $(Z_t)_{t\geq 0}$ defined by*

$$Z_t \;=\; \max \mathfrak{P}_t\left((X_t)_{t\geq 0}, \alpha(x)\exp(-x)\mathrm{d}x\right)\,,$$

let

$$\zeta_n(t) \;:=\; \max_{k=1,\ldots,n} Z_t^{(k)} - \log(n)\,, \qquad n\in\mathbb{N}\,,$$

where $((Z_t^{(k)})_{t\geq 0})_{k=1,\ldots,n}$ are i.i.d. copies of $(Z_t)_{t\geq 0}$, and let

$$\zeta_L(t) \;:=\; \max \mathfrak{P}_t((L_t)_{t\geq 0}, \exp(-x)\mathrm{d}x)\,.$$

Then ζ_n converges to ζ_L in finite-dimensional distributions. If, furthermore, $(L_t)_{t\geq 0}$ has continuous paths, then convergence holds in $\mathbb{D}([0,\infty))$ equipped with the Skorokhod topology.

3.2 Summary and Outlook

In the papers contained in this thesis, we have studied extreme value theory for dependent stochastic processes, with a focus on PoT statistics and serial tail dependence estimators for long memory linear time series, as well as maxima of stationary particle systems.

The first paper develops central limit theory for PoT partial sums of long memory linear time series. Simulation studies indicate that the resulting central limit theorems cannot in general be used directly for statistical inference, at least in the univariate setting. An important next step is therefore to develop inference procedures that account for the nonstandard normalization and the potentially slow convergence to the limiting distribution. One possible direction is self-normalization. Another is a more precise analysis of finite-sample bias, in particular second-order bias corrections for tail estimators.

The second paper develops central limit theory for serial tail dependence estimators of heavy-tailed long memory linear time series. For specific rectangular sets that are relevant in practice, the asymptotic scale vanishes when passing from deterministic to random thresholds. Consequently, the convergence rate obtained for deterministic thresholds is no longer the correct one, and determining the appropriate rate and limiting distribution for random thresholds in this case remains an open problem. A further question arising from both the first and second papers is what happens outside the parameter regimes in which the respective reduction principles apply.

The third paper studies maxima of stationary systems of randomly time-changed Lévy particles. It develops a general construction principle that yields infinite-dimensional classes of stationary max-infinitely divisible processes. Since the constructed max-infinitely divisible processes are shown to belong to a prespecified max-domain of attraction, they provide flexible benchmark models for studying the finite-sample behavior of extreme-value methods that are designed to be valid throughout a max-domain of attraction. Their PPP representation also suggests that simulation may be approached using extremal-function techniques. Such methods were developed for max-stable processes by Dombry et al. [2016] and extended to max-id processes by Zhong et al. [2024].

Our construction is based on perturbing the invariant process–measure pair of an underlying Lévy–Brown–Resnick process by means of a state-dependent random time change, where the perturbation can be controlled through the resolvent of the time-changed process. An alternative direction would be to perturb the transition semigroup directly, using tools from the theory of Feller processes and stochastic differential equations; see Remark 3.8 in Scheffel [2026].

Bibliography

B. Basrak and J. Segers. Regularly varying multivariate time series. *Stochastic Processes and their Applications*, 119(4):1055–1080, 2009.

J. Beran. *Statistics for Long-Memory Processes*. Chapman and Hall, New York, 1994.

J. Beran, Y. Feng, S. Ghosh, and R. Kulik. *Long-Memory Processes: Probabilistic Properties and Statistical Methods*. Springer, Berlin, Heidelberg, 2013.

S. M. Berman. Limit theorems for the maximum term in stationary sequences. *The Annals of Mathematical Statistics*, 35(2):502–516, 1964.

N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation*. Cambridge University Press, Cambridge, 1987.

P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer, New York, 2nd edition, 1991.

B. M. Brown and S. I. Resnick. Extreme Values of Independent Stochastic Processes. *Journal of Applied Probability*, 14(4):732–739, 1977.

M. Brown. A Property of Poisson Processes and its Application to Macroscopic Equilibrium of Particle Systems. *The Annals of Mathematical Statistics*, 41(6): 1935–1941, 1970.

A. Bücher and C. Zhou. A horse race between the block maxima method and the peak-over-threshold approach. *Statistical Science*, 36(3):360–378, 2021.

E. Castillo, A. S. Hadi, N. Balakrishnan, and J. M. Sarabia. *Extreme Value and Related Models with Applications in Engineering and Science*. Wiley, Hoboken, NJ, 2005.

S. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer, London, 2001.

H. Cramér. On the intersections between the trajectories of a normal stationary stochastic process and a high level. *Arkiv för Matematik*, 6:337–349, 1966.

L. Daston and K. Park. *Wonders and the Order of Nature 1150–1750*. Zone Books, New York, 1998.

R. A. Davis and T. Hsing. Point process and partial sum convergence for weakly dependent random variables with infinite variance. *The Annals of Probability*, 23(2):879–917, 1995.

R. A. Davis and T. Mikosch. The extremogram: A correlogram for extreme events. *Bernoulli*, 15(4):977–1009, 2009.

L. de Haan. A Spectral Representation for Max-stable Processes. *The Annals of Probability*, 12(4):1194–1204, 1984.

L. De Haan and A. Ferreira. *Extreme Value Theory*. Springer, New York, NY, 2006.

C. Dombry, S. Engelke, and M. Oesting. Exact simulation of max-stable processes. *Biometrika*, 103(2):303–317, 2016.

R. Durrett. *Probability: Theory and Examples*. Cambridge University Press, Cambridge, 5th edition, 2019.

P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events*. Springer, Berlin, Heidelberg, 1997.

R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 24(2):180–190, 1928.

M. Fréchet. Sur la loi de probabilité de l'écart maximum. *Annales de la Société Polonaise de Mathématique*, 6:93–116, 1927.

L. Giraitis, H. L. Koul, and D. Surgailis. *Large Sample Inference for Long Memory Processes*. Imperial College Press, London, 2012.

B. V. Gnedenko. Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, 44(3):423–453, 1943.

O. Kallenberg. *Foundations of Modern Probability*. Springer International Publishing, Cham, 3rd edition, 2021.

A. Klenke. *Probability Theory: A Comprehensive Course*. Springer International Publishing, Cham, 2020.

R. Kulik and P. Soulier. The tail empirical process for long memory stochastic volatility sequences. *Stochastic Processes and their Applications*, 121(1):109–134, 2011.

G. Last and M. Penrose. *Lectures on the Poisson Process*. Cambridge University Press, Cambridge, 2017.

M. R. Leadbetter. On extreme values in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 28:289–303, 1974.

M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York, 1983.

R. M. Loynes. Extreme values in uniformly mixing stationary stochastic processes. *The Annals of Mathematical Statistics*, 36(3):993–999, 1965.

T. Mikosch and O. Wintenberger. *Extreme Value Theory for Time Series: Models with Power-Law Tails*. Springer Nature Switzerland, Cham, 2024.

R.-D. Reiss and M. Thomas. *Statistical Analysis of Extreme Values: With Applications to Insurance, Finance, Hydrology and Other Fields*. Birkhäuser, Basel, 3rd edition, 2007.

S. I. Resnick. *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York, 2007.

I. Scheffel. Maxima of stationary systems of randomly time-changed Lévy particles, 2026. arXiv preprint arXiv:2604.11434.

I. Scheffel, M. Oesting, and G. Stupfler. Central limit theory for peaks-over-threshold partial sums of long memory linear time series. *Stochastic Processes and their Applications*, 197:104929, 2026a.

I. Scheffel, M. Oesting, and G. Stupfler. Central limit theory for serial tail dependence estimators in heavy-tailed long memory linear time series, 2026b. arXiv preprint arXiv:2608.10944.

S. A. Stoev. On the ergodicity and mixing of max-stable processes. *Stochastic Processes and their Applications*, 118(9):1679–1705, 2008.

G. S. Watson. Extreme values in samples from m-dependent stationary stochastic processes. *The Annals of Mathematical Statistics*, 25(4):798–800, 1954.

P. Zhong, R. Huser, and T. Opitz. Exact Simulation of Max-Infinitely Divisible Processes. *Econometrics and Statistics*, 30:96–109, 2024.

Part II

Cumulative Part

Chapter 4

Declaration to the Cumulative Part

The cumulative part of this thesis contains three research papers (Chapters 5–7). Their titles and publication statuses are as follows:

1. Ioan Scheffel, Marco Oesting, and Gilles Stupfler, *Central limit theory for Peaks-over-Threshold partial sums of long memory linear time series*, published in *Stochastic Processes and their Applications*, 2026.
   Reference: Scheffel et al. [2026a]
   Link: `https://doi.org/10.1016/j.spa.2026.104929`
   Link to arXiv version: `https://doi.org/10.48550/arXiv.2506.20789`
   This article (specifically Version 3 on arXiv, which corresponds to the accepted version of the published paper) is reproduced in Chapter 5.

2. Ioan Scheffel, Marco Oesting, and Gilles Stupfler, *Central limit theory for serial tail dependence estimators in heavy-tailed long memory linear time series*, submitted to *Probability Theory and Related Fields*, preprint available on arXiv, 2026.
   Reference: Scheffel et al. [2026b]
   Link to arXiv version: `https://doi.org/10.48550/arXiv.2608.10944`
   This article (specifically Version 1 on arXiv) is reproduced in Chapter 6.

3. Ioan Scheffel, *Maxima of stationary systems of randomly time-changed Lévy particles*, submitted to *Stochastic Processes and their Applications*, preprint available on arXiv, 2026.
   Reference: Scheffel [2026]
   Link to arXiv version: `https://doi.org/10.48550/arXiv.2604.11434`
   This article (specifically Version 2 on arXiv) is reproduced in Chapter 7.

I, *Ioan Scheffel*, hereby declare that the co-author lists are complete and that I have not reproduced, without acknowledgment, the work of another. I declare that I made major contributions to these articles, including in particular the phase of problem selection, literature research, and the derivation of theoretical and numerical results as well as the writing process. This applies to all three articles listed above.

In particular, in Articles 1 and 2, except where stated otherwise, I developed the main results and proofs and wrote the initial drafts of both manuscripts. Marco Oesting and Gilles Stupfler supervised the research, contributed to improving and completing proofs, strengthening some of the results, and identifying examples and counterexamples that helped shape the statements. Gilles Stupfler also provided the basic idea for the proof of the derandomization device [Scheffel et al., 2026a, Lemma C.1], and provided the complete proof of Theorem 5.5 in Scheffel et al. [2026b]. All authors reviewed and approved the final manuscripts.

Chapters 5–7 reproduce the original articles.

Place and Date: Stuttgart, August 14, 2026

Signature: ______________________________

Hiermit bestätige ich, *Ioan Scheffel*, dass Artikel 1 der auf Seite 39 gelisteten Artikel

i in einem englischsprachigen, international verbreiteten Publikationsorgan mit Peer Review zur Veröffentlichung akzeptiert und online erschienen ist,

ii ich den Hauptbeitrag zu den neuen inhaltlichen Erkenntnissen dieser Publikation geleistet habe,

iii ich federführend bin, das heißt, den Hauptbeitrag zur Abfassung, (das heißt Umsetzung der inhaltlichen Erkenntnisse der Publikation) geleistet habe.

Für Artikel 2 und 3 auf Seite 39, welche in englischsprachigen, international verbreiteten Publikationsorganen mit Peer Review eingereicht, jedoch bisher noch nicht zur Veröffentlichung akzeptiert sind, treffen die Punkte ii und iii ebenfalls zu.

Place and Date: Stuttgart, August 14, 2026

Signature: ______________________________

Chapter 5

Central limit theory for Peaks-over-Threshold partial sums of long memory linear time series

Central limit theory for Peaks-over-Threshold partial sums of long memory linear time series

Ioan Scheffel[a], Marco Oesting[a,b], Gilles Stupfler[c]

[a] Institute for Stochastics and Applications, University of Stuttgart, D-70563 Stuttgart, Germany

[b] Stuttgart Center for Simulation Science (SC SimTech), University of Stuttgart, D-70569 Stuttgart, Germany

[c] Univ Angers, CNRS, LAREMA, SFR MATHSTIC, F-49000 Angers, France

Abstract

Over the last 30 years, extensive work has been devoted to developing central limit theory for partial sums of subordinated long memory linear time series. A much less studied problem, motivated by questions that are ubiquitous in extreme value theory, is the asymptotic behavior of such partial sums when the subordination mechanism has a threshold depending on sample size, so as to focus on the right tail of the time series. This article substantially extends longstanding asymptotic techniques by allowing the subordination mechanism to depend on the sample size in this way and to grow at a polynomial rate, while permitting the innovation process to have infinite variance. The cornerstone of our theoretical approach is a tailored reduction principle, which enables the use of classical results on partial sums of long memory linear processes. In this way we obtain asymptotic theory for certain Peaks-over-Threshold estimators with deterministic or random thresholds. Applications cover both heavy- and light-tailed regimes, yielding unexpected results which, to the best of our knowledge, are new to the literature. A simulation study illustrates the relevance of our findings in finite samples.

MSC 2020 subject classifications: 60F05, 60G70
Keywords: Central limit theory, Heavy tails, Peaks-over-Threshold model, Stable distribution, Subordinated linear time series

1 Introduction

1.1 Background and literature review

A causal linear (or MA(∞)) time series $X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$, $t \in \mathbb{Z}$, built upon i.i.d. innovations (ε_t) with at least a fractional moment and real-valued coefficients (a_j), is (roughly speaking) said to have long memory if the decay of the coefficients to 0 is slow enough to yield a divergent series, yet fast enough to accommodate tail heaviness of the innovations. Long memory linear time series models include the ARFIMA model introduced by [18], which is arguably the simplest example of a stochastic process having long memory. The ARFIMA model has been successfully applied to various fields, such as finance and economics [6] and medicine [41], and is the subject of active mathematical research, for example on fundamental aspects linked to model selection [23].

Central limit theorems for partial sums of long memory linear time series are well-known; see among others [5, Chapter 4]. Instead of the linear time series (X_t), however, in many applications one is interested in a subordinated time series, that is, a transformation $G(X_t)$. Typical examples are $G(X_t) = |X_t|^p$, yielding the empirical pth moment, or $G(X_t) = \mathbb{1}\{X_t \leq x\}$, providing an estimator of the distribution function of X_0 at the point x. This problem is discussed by [21] (innovations with finite eighth moment), [20] (innovations with finite fourth moment and $G(X_t) = \mathbb{1}\{X_t \leq x\}$), [39] (innovations with finite second moment but no fourth moment) and [26] (innovations with finite first moment but no second moment); see also [13, Section III] and [29, Section 16.1]. The cornerstone of a standard proof is to show a (weak) reduction principle of the form

$$n^{-d-1/\min(\nu,2)} \sum_{t=1}^{n} (G(X_t) - \mathbf{E}[G(X_0)]) = n^{-d-1/\min(\nu,2)} G'_\infty(0) \sum_{t=1}^{n} X_t + o_{\mathbf{P}}(1), \qquad (1)$$

where $\nu = \sup\{p > 1 \mid \|\varepsilon\|_{L^p(\mathbf{P})} < \infty\}$ is the (algebraic) number of finite moments of the innovations, assumed to be larger than 1, d is a long memory parameter, reminiscent of the fractional differencing parameter in ARFIMA models, that satisfies $d \in (0, 1-1/\min(\nu,2))$ and is such that $(n^{1-d}a_n)$ converges to a finite, nonzero limit, and $G'_\infty(0) = \mathrm{d}(\mathbf{E}[G(X_0 + x)])/\mathrm{d}x\big|_{x=0}$ is assumed not to vanish. By [5, Theorem 4.6 ($\nu \geq 2$), Theorem 4.17 ($1 < \nu < 2$)], the weak limit of the left-hand side of (1) is then a stable distribution, which is non-Gaussian for $\nu \in (1,2)$, and whose scale depends on the asymptotic behavior of the coefficients (a_j), tail heaviness of the innovations, and the derivative $G'_\infty(0)$.

Fitting extreme value analysis into this framework requires shifting from fixed transformations G to a sequence of sample-size-dependent transformations (G_n). Typical examples arise from counting threshold exceedances – the simplest example among a broader class of estimators called Peaks-over-Threshold (PoT) estimators –, that is,

$$\frac{1}{n}\sum_{t=1}^{n} \mathbb{1}\{X_t > u_n\}, \qquad \text{with } G_n(x) = \mathbb{1}\{x > u_n\}$$

for a threshold sequence (u_n) with $u_n \to \infty$ as $n \to \infty$, or

$$\frac{1}{n}\sum_{t=1}^{n} \log(X_t/u_n)\mathbb{1}\{X_t > u_n\}, \qquad \text{with } G_n(x) = (\log(x) - \log(u_n))\mathbb{1}\{x > u_n\}$$

which is reminiscent of the Hill estimator [19] of the extreme value index for heavy-tailed data. In practice the threshold sequence is often chosen as an order statistic from the sample, that is, $u_n = X_{n-k:n}$ for an intermediate sequence $k = k(n)$ with $k \to \infty$ and $k/n \to 0$ for $n \to \infty$, where $X_{1:n} \leq \ldots \leq X_{n:n}$ is the ordered sample. This corresponds to the realistic setting where the chosen threshold estimates an unknown quantile threshold of the form $q_X(1 - k/n)$. The asymptotic behavior of these estimators, whether with deterministic or random thresholds, is well-studied under various notions of short-range dependence; see for instance [22] in a strong mixing setting, as well as later work by [15] for extensions to a wider class of estimators in the $\beta-$mixing framework under an anti-clustering condition (see also [13, Section V]). Recent contributions, such as [32] (based on the notion of dependence introduced in [30]), or [3] and [9], have focused on weakening the mixing requirement, which may be difficult to check in a given model.

The question of handling long-range dependence in PoT extreme value analysis has remained much less explored. Works by [27], [7] and [29, Section 16.3] study the tail empirical process with random thresholds, that is, $(\frac{1}{k}\sum_{t=1}^{n} \mathbb{1}\{X_t > X_{n-k:n}s\})_{s\geq 1}$, and show that it does not suffer from long memory in certain stochastic volatility models. [28] discusses the existence of multivariate versions of the extremogram in the same class of processes. This is, to the best of our knowledge, the only class of processes that has been studied in this context; in particular, long memory linear time series have been left untouched.

1.2 Contribution of the paper

The main contribution of our article is to derive central limit theorems for subordinated long memory linear time series, in an extreme value framework where the subordination mechanism depends on sample size. This is challenging because the mixing properties that are typically assumed in the extreme value literature (in particular, strong mixing, $\beta-$mixing or anti-clustering conditions) do not hold in this context. To be more specific, under mild smoothness assumptions on the innovation sequence, and provided each member of the sequence of transformations (G_n) grows at most like a (suitable) power function and is supported on an interval of the form (u_n, ∞) with $u_n \to \infty$, we show that the centered and scaled subordinated linear time series

$$n^{-d-1/\min(\nu,2)} \sum_{t=1}^{n} (G_n(X_t) - \mathbf{E}[G_n(X_0)])$$

has the same asymptotic behavior as

$$\frac{\mathrm{d}}{\mathrm{d}x} \mathbf{E}[G_n(X_0 + x)] \bigg|_{x=0} n^{-d-1/\min(\nu,2)} \sum_{t=1}^{n} X_t \,.$$

This new, tailored reduction principle is achieved by explicitly deriving an upper bound for an appropriate moment of the difference between the two random sums. This technical result, specific to the long memory framework, leads to our main central limit theorem for partial sums of subordinated long memory linear time series, yielding the same limiting distribution as the partial sums of the original time series, but with a different rate of convergence due to focusing only on the extremes of (X_t).

Our results stand in contrast to what one would expect from the available i.i.d. or short-range-dependent theory of extreme values. For example, with heavy-tailed innovations, we observe that the speed of convergence is *faster* than the classical long memory speed arising from a setting with fixed transformation. This is markedly different from independent or short-range dependent settings, where it is well-known that the speed of convergence of PoT extreme value estimators based on the k largest observations only is $\sqrt{k}$, which is *slower* than the standard speed $\sqrt{n}$. A further contribution of our work is to allow the replacement of the deterministic threshold (u_n) with its random counterpart; in doing so, we obtain a second set of asymptotic results for the extreme value estimators under consideration, showing that, unlike in the i.i.d. and short memory settings [37], Hill estimators with deterministic and random thresholds have different asymptotic distributions. A similar phenomenon was observed by [27] in stochastic volatility models with long memory.

We organize the presentation of these theoretical findings as follows. In Section 2, we describe the model and assumptions in detail. Section 3 contains the announced central limit results. In addition, we assess the practical relevance of these theoretical results in a simulation study, presented in Section 4. While confirming the expected rate of convergence, the simulations show that in finite samples convergence to the actual form of the asymptotic distribution is slow, even in relatively simple settings. This is partly due to the extreme value setting and also due to the intrinsic difficulty of working with long memory time series, which is pronounced already at central levels. Section 5 finally discusses our findings and perspectives for future work. The proofs of all the results are deferred to an online Supplementary Material document.

2 Model and main assumptions

Throughout this article, $(\Omega, \mathcal{A}, \mathbf{P})$ denotes a probability space rich enough to support all random variables under consideration. For $p \geq 1$, we write $L^p(\mathbf{P})$ for the space of random variables with finite p-th moment. For a generic, continuous random variable Z, we let F_Z (resp. f_Z, q_Z) denote its cumulative distribution function (resp. probability density function, quantile function). We write $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$ for $a, b \in \mathbb{R}$, and use the symbol $\sim$ to denote asymptotic equivalence of sequences and functions. We write $\lesssim$ to mean inequality up to a generic multiplicative constant $C > 0$ that can change from place to place. If the dependence of this constant on the surrounding variables is relevant to the argument, we will make this explicit, by saying, for example, that $C = C_n$ depends on n, or that C is independent of n. The symbol $C^p(\mathbb{R})$ denotes the vector space of real-valued functions on $\mathbb{R}$ with continuous derivatives up to order p.

Let $(\varepsilon_t)_{t\in\mathbb{Z}}$ be a sequence of independent and identically distributed (i.i.d.) copies of a random variable ε. Set

$$\nu = \nu_\varepsilon := \sup\{p > 1 \mid \|\varepsilon\|_{L^p(\mathbf{P})} < \infty\}.$$

We have $\nu = \infty$, for example, if ε is standard Gaussian, and if the complementary distribution function $x \mapsto \mathbf{P}[|\varepsilon| > x]$ is regularly varying with index $-1/\xi < 0$, then $\nu = 1/\xi$. In this work, we focus on linear time series $X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$, where (a_j) is a sequence of real-valued coefficients, featuring long memory in the following sense.

Assumption 1 (Long memory)**.** It holds that $\nu > 1$, and that the sequence $(n^{1-d} a_n)$ converges to a finite, nonzero limit c_a for some $d \in (0, 1 - 1/\alpha)$, where $\alpha := 2 \wedge \nu$.

Remark 2.1 (On Assumption 1)**.** Assumption 1 guarantees that the process (X_t) is well-defined and stationary; see [29, Section 15.3]. It imposes that the innovations have at least a finite moment of order larger than 1, and that the coefficient sequence (a_j), while forming a divergent series and thus granting the long memory property to (X_t), is able to accommodate potentially heavy tails of the innovations.

Our main goal is to derive a central limit theorem for partial sums of subordinated long memory linear time series, in the sense of Section 4.2.5 in [5], taking the form

$$v_n \sum_{t=1}^{n} (G_n(X_t) - \mathbf{E}[G_n(X_0)]) \xrightarrow{\mathrm{d}} Z\,,$$

where (G_n) is a given sequence of real-valued functions, and (v_n), Z are respectively a nonrandom sequence and a nondegenerate stable limiting random variable to be determined. This is tailored to extreme value theory within the Peaks-over-Threshold (PoT) framework, where G_n may for instance be $G_n(x) = \mathbb{1}\{x > u_n\}$ (for counting the number of exceedances above a high threshold $u_n \to \infty$) or $G_n(x) = (\log(x) - \log(u_n))\, \mathbb{1}\{x > u_n\}$, resulting in the Hill estimator [19]. The first of the regularity assumptions we require regards the sequence of functions (G_n).

Assumption 2 (Regularity of G_n)**.** It holds that

$$|G_n(x)| \;\lesssim\; (1 + |x|)^{\gamma_G} \mathbb{1}\{x > u_n\}\,,$$

for some $\gamma_G \geq 0$ and a deterministic sequence (u_n) of thresholds with $u_n \to \infty$ as $n \to \infty$. Here, $\lesssim$ denotes inequality up to a multiplicative constant $C > 0$ independent of n and x.

Remark 2.2 (On Assumption 2)**.** The functions G_n appearing in the simple threshold exceedance estimator and the Hill estimator [19] satisfy Assumption 2, with $\gamma_G = 0$ and any $\gamma_G > 0$, respectively. In Section 3, we introduce further technical restrictions on γ_G that prevent G_n from growing too fast, depending on the tail heaviness of the innovations.

Our second main regularity assumption focuses on the innovation sequence (ε_t) and is tailored to the two main regimes of interest in our work. On the one hand, we will consider the (easier) case when ε has a finite second moment with sufficiently fast decay of the characteristic function. On the other hand, we shall be interested in the (harder) case of heavy-tailed time series whose innovations have regularly varying densities and a finite first moment, but no finite second moment.

Assumption 3 (Regularity of the innovations)**.** Suppose that $F_\varepsilon \in C^2(\mathbb{R})$, and that ε has a symmetric distribution. Furthermore, assume that one of the following alternatives holds:

(i) (Case $\alpha = 2$) One has $\mathbf{E}[\varepsilon^2] < \infty$, and, for some $\delta > 0$, the characteristic function of ε is such that $(1 + |s|)^\delta \mathbf{E}[\exp(\mathrm{i}s\varepsilon)]$ is uniformly bounded in $s \in \mathbb{R}$.

(ii) (Case $\alpha \in (1, 2)$) Assume that there is a finite positive constant A such that $\lim_{x\to\infty} x^\alpha \mathbf{P}[\varepsilon > x] = \lim_{x\to\infty} x^\alpha \mathbf{P}[\varepsilon < -x] = A/2$, and suppose that the probability density function $f_\varepsilon \in C^1(\mathbb{R})$ satisfies:

$$\begin{aligned} |f_\varepsilon'(x)| &\lesssim (1 + |x|)^{-\alpha} \qquad \text{for } x \in \mathbb{R}\,, \\ |f_\varepsilon'(x) - f_\varepsilon'(y)| &\lesssim |x - y| \cdot (1 + |x|)^{-\alpha} \qquad \text{for } x, y \in \mathbb{R}\,, |x - y| < 1\,, \end{aligned}$$

where $\lesssim$ denotes inequality up to a multiplicative constant $C_\varepsilon > 0$ depending only on the distribution of ε.

Our main interest in this work is Case (ii) in Assumption 3, that is, the setting $\alpha \in (1, 2)$, which should be viewed as the harder case of the two. Let us point out that our assumptions in Case (i) might be weakened further using dedicated techniques from [20]. The rationale for the precise form of our Assumption 3(i) is to allow a comparison of asymptotic behavior between the infinite and finite variance contexts in a unified fashion.

Remark 2.3 (On Assumption 3(i))**.** Verifying Assumption 3(i) typically boils down to checking integrability properties of the density f_ε and its derivative. Take, for example, any distribution of ε having finite second moment with density vanishing asymptotically at $\pm\infty$ and integrable first derivative. Then on the one hand, for $|s| \leq 1$, it holds that

$$|\mathbf{E}[\exp(\mathrm{i}s\varepsilon)]| \leq 1 \leq \frac{1}{1 + |s|}\,,$$

and on the other hand, for $|s| > 1$, integration by parts yields

$$|\mathbf{E}[\exp(\mathrm{i}s\varepsilon)]| = \left| \frac{1}{\mathrm{i}s} \int_{\mathbb{R}} \exp(\mathrm{i}sx) f_\varepsilon'(x)\, \mathrm{d}x \right| \leq \frac{1}{|s|} \left\| f_\varepsilon' \right\|_{L^1(\mathbb{R})} \lesssim \frac{1}{1 + |s|}$$

due to the inequality $1/|s| \leq 2/(1 + |s|)$ for $|s| > 1$. Then Assumption 3(i) is satisfied with $\delta = 1$. Therefore, it covers a large class of distributions, in particular centered Gaussian distributions, symmetric Beta distributions with parameter larger than 2, and Student distributions with $\nu > 2$ degrees of freedom. Note that this nonetheless ensures that X_0 has an infinitely differentiable distribution function; see [16, Lemma 1].

Remark 2.4 (On Assumption 3(ii)). Assumption 3(ii) is satisfied by symmetric α-stable ($S\alpha S$) distributed innovations; see [26, after (2.2)]. More generally, the two inequalities there hold if f_ε' and f_ε'' are continuous and regularly varying (with index $\leq -\alpha$). In particular, the Student $t-$distribution satisfies this stronger criterion. We note that the symmetry property is assumed mainly to simplify the application of the central limit theorem for partial sums when $\alpha \in (1,2)$; it would suffice for the innovations to be centered. Without symmetry, the limiting distribution becomes asymmetric, but the convergence rates remain the same, see [29, Theorem 8.3.5].

3 Main results

This section is split into four subsections. In Section 3.1, we develop our reduction principle, which is the key technical tool behind our main results. The reader primarily interested in central limit theory can skip this subsection and go directly to Section 3.2, where we provide and discuss a general central limit theorem arising as a corollary of this reduction principle. This result is then applied to the cases when the stationary distribution of (X_t) is heavy-tailed (resp. light-tailed) in Section 3.3 (resp. Section 3.4).

3.1 Technical tool: Reduction principle

At a conceptual level, the reduction principle works as follows. For $k \in \mathbb{Z}$, let $\mathcal{F}_k = \sigma(\varepsilon_k, \varepsilon_{k-1}, \ldots)$ denote the past σ-algebra generated by the sequence (ε_t) up to index k. For $k \geq 0$, let $X_{t,k} := \sum_{j=0}^{k} a_j \varepsilon_{t-j}$ be the k-truncated time series, and write

$$G_{k,n}(y) \ := \ \mathbf{E}[G_n(X_{0,k} + y)] \qquad \text{and} \qquad G_{k,n}'(y) \ = \ \frac{\mathrm{d}}{\mathrm{d}y} G_{k,n}(y)\,,$$

where we set $G_{\infty,n}(y) := \mathbf{E}[G_n(X_{0,\infty} + y)] = \mathbf{E}[G_n(X_0 + y)]$ for $k = \infty$ in the natural way. The reduction principle consists in showing that

$$\begin{aligned}
&\sum_{t=1}^{n} (G_n(X_t) \ - \ \mathbf{E}[G_n(X_0)]) \\
&= \ \sum_{t=1}^{n} \sum_{k=0}^{\infty} \left(\mathbf{E}\left[G_n(X_t) \mid \mathcal{F}_{t-k}\right] \ - \ \mathbf{E}\left[G_n(X_t) \mid \mathcal{F}_{t-(k+1)}\right]\right) \\
&= \ \sum_{t=1}^{n} \sum_{k=0}^{\infty} (G_{k-1,n}(X_t - X_{t,k-1}) \ - \ G_{k,n}(X_t - X_{t,k})) \\
&\approx \ \sum_{t=1}^{n} \sum_{k=0}^{\infty} (G_{k,n}(X_t - X_{t,k-1}) \ - \ G_{k,n}(X_t - X_{t,k})) \\
&\approx \ \sum_{t=1}^{n} \sum_{k=0}^{\infty} a_k \varepsilon_{t-k} G_{k,n}'(X_t - X_{t,k}) \\
&= \ G_{\infty,n}'(0) \sum_{t=1}^{n} X_t \ + \ \sum_{t=1}^{n} \sum_{k=0}^{\infty} a_k \varepsilon_{t-k} \left(G_{k,n}'(X_t - X_{t,k}) \ - \ G_{\infty,n}'(0)\right)\,.
\end{aligned} \tag{2}$$

Here, the asymptotic profile turns out to be determined only by the linear term, for which central limit theory is readily available. Before we make the argument in (2) rigorous, we provide central limit theory for partial sums of long memory linear time series. For a proof of the latter we refer to [5, Theorem 4.6 ($\alpha = 2$) and Theorem 4.17 ($\alpha \in (1,2)$)].

Theorem 3.1 (Central limit theorems for partial sums)**.** *Let Assumptions 1 and 3 hold. Then*

$$n^{-d-1/\alpha}\sum_{t=1}^{n} X_t \xrightarrow{\mathrm{d}} Z_\alpha \qquad \text{as } n\to\infty\,,$$

where

$$Z_\alpha \quad \text{has distribution} \quad \begin{cases}\mathcal{N}(0,\sigma^2) & \text{if } \alpha=2 \text{ with } \mathbf{E}[\varepsilon^2]<\infty,\\ S\alpha S(\eta) & \text{if } \alpha\in(1,2)\,,\end{cases}$$

with variance

$$\sigma^2 := c_a^2\mathbf{E}[\varepsilon^2]\frac{B(1-2d,d)}{d(2d+1)} \quad \text{when } \mathbf{E}[\varepsilon^2]<\infty,$$

or scale

$$\eta := \frac{c_a}{d}\cdot\left(A\cdot\frac{\Gamma(2-\alpha)\cos(\pi\alpha/2)}{1-\alpha}\right)^{1/\alpha}\left(\int_{-\infty}^{1}\left((1-v)_+^d-(-v)_+^d\right)^\alpha \mathrm{d}v\right)^{1/\alpha}.$$

Remark 3.2 (On the limiting distribution in Theorem 3.1)**.** We define the distribution $S\alpha S(\eta)$ in the usual sense, having characteristic function $x\mapsto \exp(-\eta^\alpha|x|^\alpha)$. Also, it may not be immediately obvious that the integral in the definition of η is finite in our setting. To clarify, using Jensen's inequality and the condition $1<(1-d)\alpha<2$, we have

$$\begin{aligned}\int_1^\infty\left((1+v)^d \;-\; v^d\right)^\alpha \mathrm{d}v &= \int_1^\infty\left(d\int_0^1 (v+s)^{-(1-d)}\,\mathrm{d}s\right)^\alpha \mathrm{d}v\\ &\le \int_1^\infty\left(\int_0^1 (v+s)^{-(1-d)\alpha}\,\mathrm{d}s\right)\mathrm{d}v\\ &= \int_0^1\left(\int_1^\infty (v+s)^{-(1-d)\alpha}\,\mathrm{d}v\right)\mathrm{d}s\\ &\le \int_1^\infty v^{-(1-d)\alpha}\,\mathrm{d}v<\infty\,.\end{aligned}$$

Therefore,

$$\begin{aligned}&\int_{-\infty}^1\left((1-v)_+^d-(-v)_+^d\right)^\alpha \mathrm{d}v\\ &= \int_{-1}^1\left((1-v)_+^d-(-v)_+^d\right)^\alpha \mathrm{d}v+\int_1^\infty\left((1+v)^d-v^d\right)^\alpha \mathrm{d}v<\infty\,.\end{aligned}$$

The next result makes the argument in (2) rigorous. In the proof we adapt techniques of [26] to our extreme value setting.

Theorem 3.3 (L^r-moment bound)**.** *Let Assumptions 1, 2 and 3 hold. Then*

$$G'_{\infty,n}(0) \;:=\; \frac{\mathrm{d}}{\mathrm{d}x}\mathbf{E}[G_n(X_0+x)]\bigg|_{x=0}$$

is well-defined. Moreover, if γ_G is such that there are $\gamma,r\in\mathbb{R}$ satisfying

$$\gamma_G \;<\; \gamma \;<\; \min\left\{\frac{d}{1-d}\,,1-\frac{1}{r(1-d)}\,,\frac{\alpha}{r}-1\right\} \qquad \text{and} \qquad \frac{1}{1-d} \;<\; r \;<\; \alpha\,, \tag{3}$$

we have

$$\begin{aligned}
&\left\| \sum_{t=1}^{n} (G_n(X_t) \ - \ \mathbf{E}[G_n(X_0)]) \ - \ G'_{\infty,n}(0) \sum_{t=1}^{n} X_t \right\|_{L^r(\mathbf{P})} \\
&\lesssim \ u_n^{(\gamma_G-\gamma)} n^{1+1/r-(1-d)(1+\gamma)} \ + \ n^{1/r} \left(\|G_n(X_0)\|_{L^r(\mathbf{P})} \vee |G'_{\infty,n}(0)| \right) \\
&\lesssim \ n^{1+1/r-(1-d)(1+\gamma)} \,,
\end{aligned}$$

where $\lesssim$ *denotes inequality up to a multiplicative constant* $C > 0$ *independent of* n*.*

It follows from Theorem 3.3 that if $G'_{\infty,n}(0) \neq 0$ for sufficiently large n, then

$$\left\| \frac{n^{-d-1/\alpha}}{G'_{\infty,n}(0)} \sum_{t=1}^{n} (G_n(X_t) \ - \ \mathbf{E}[G_n(X_0)]) \ - \ n^{-d-1/\alpha} \sum_{t=1}^{n} X_t \right\|_{L^r(\mathbf{P})} \ \lesssim \ \frac{n^{\kappa(\gamma,r)-d-1/\alpha}}{|G'_{\infty,n}(0)|} \,,$$

where $\kappa(\gamma, r) := 1 + 1/r - (1-d)(1+\gamma)$. Therefore, if $n^{\kappa(\gamma,r)-d-1/\alpha}/G'_{\infty,n}(0) \to 0$ as $n \to \infty$, then the following $L^r(\mathbf{P})$−reduction principle applies:

$$\frac{n^{-d-1/\alpha}}{G'_{\infty,n}(0)} \sum_{t=1}^{n} (G_n(X_t) \ - \ \mathbf{E}[G_n(X_0)]) \ - \ n^{-d-1/\alpha} \sum_{t=1}^{n} X_t \overset{L^r(\mathbf{P})}{\longrightarrow} 0 \qquad \text{as } n \to \infty.$$

Remark 3.4 (On the existence of γ and r in Theorem 3.3)**.** Note that the upper bound for γ in (3) gets maximal for

$$r = r^\star := \frac{1}{2}\left(\alpha + \frac{1}{1-d}\right),$$

that is, for r being equal to the midpoint of its domain. In this case, the upper bound is given by

$$\gamma^\star = \gamma(r^\star) = \min\left\{\frac{d}{1-d}, \frac{\alpha(1-d)-1}{\alpha(1-d)+1}\right\} = \begin{cases} \dfrac{d}{1-d} & \text{if } \dfrac{1}{(1-d)(1-2d)} < \alpha\,, \\[2ex] \dfrac{\alpha(1-d)-1}{\alpha(1-d)+1} & \text{otherwise.} \end{cases}$$

Thus, $\gamma_G < \gamma^\star$ is necessary for the existence of a pair (γ, r) satisfying (3). Conversely, if $\gamma_G < \gamma^\star$, the pair $(\gamma^\star - \delta, r^\star)$ satisfies Condition (3) for sufficiently small $\delta > 0$. Consequently, Condition (3) is nonempty if and only if $\gamma_G < \gamma^\star$. Simple calculations show that

$$\max_{d\in(0,1-1/\alpha)} \min\left\{\frac{d}{1-d}, \frac{\alpha(1-d)-1}{\alpha(1-d)+1}\right\} = \frac{8(1-1/\alpha)}{(1+\sqrt{9-8(1-1/\alpha)})(3+\sqrt{9-8(1-1/\alpha)})}$$

with the maximum attained at $d(\alpha) = (3 - \sqrt{9-8(1-1/\alpha)})/4$. This constitutes a hard upper bound for γ_G in our framework. Note that $d(1) = 0$ and $d(2) = (3-\sqrt{5})/4 \approx 0.191$.

As illustrated by the example $G_n(x) = \mathbb{1}\{x > u_n\}$, where $G'_{\infty,n}(0) = f_{X_0}(u_n)$, the condition $n^{\kappa(\gamma,r)-d-1/\alpha}/G'_{\infty,n}(0) \to 0$ restricts the growth of (u_n) to infinity. Consequently, the choice of (γ, r) has a significant impact on the admissible growth rates of the threshold sequence (u_n). Therefore, we aim to relax this condition as much as possible, which amounts to minimizing the function $(\gamma, r) \mapsto \kappa(\gamma, r)$. It turns out (see Proposition B.6 in the supplement) that κ is minimized over the admissible values of γ and r at

$$(\gamma_0, r_0) := \begin{cases} \left(\dfrac{d}{1-d},\ \alpha(1-d)\right) & \text{if } \dfrac{1}{(1-d)(1-2d)} < \alpha\,, \\ \left(\dfrac{\alpha(1-d)-1}{\alpha(1-d)+1},\ \dfrac{1}{2}\left(\alpha + \dfrac{1}{1-d}\right)\right) & \text{otherwise,} \end{cases} \tag{4}$$

yielding the optimal value $\kappa_0 := \kappa(\gamma_0, r_0)$ as

$$\kappa_0 = \begin{cases} \dfrac{1}{\alpha(1-d)} & \text{if } \dfrac{1}{(1-d)(1-2d)} < \alpha\,, \\ \dfrac{2(1-d) + (1-\alpha(1-d)(1-2d))}{\alpha(1-d)+1} = d + (1-d)\dfrac{3-\alpha(1-d)}{\alpha(1-d)+1} & \text{otherwise.} \end{cases} \tag{5}$$

With the notation of Remark 3.4, $\gamma_0 = \gamma^\star$, and $r_0 = r^\star$ when $\alpha(1-d)(1-2d) \geq 1$. Noting that $\kappa_0 - d - 1/\alpha < 0$, we obtain the following optimized moment bound.

Theorem 3.5 (Optimal L^r−moment bound)**.** *Let Assumptions 1, 2 and 3 hold. If $\gamma_G < \gamma_0$, then, for any $\delta > 0$,*

$$\left\| \sum_{t=1}^n (G_n(X_t) \ - \ \mathbf{E}[G_n(X_0)]) \ - \ G'_{\infty,n}(0) \sum_{t=1}^n X_t \right\|_{L^{r_0}(\mathbf{P})} \lesssim n^{\kappa_0+\delta}\,,$$

where $\lesssim$ denotes inequality up to a multiplicative constant $C > 0$ independent of n. Consequently, if one has $G'_{\infty,n}(0) \neq 0$ for sufficiently large n, and if for some $\delta > 0$ one has $n^{\kappa_0+\delta-d-1/\alpha}/G'_{\infty,n}(0) \to 0$ as $n \to \infty$, then

$$\frac{n^{-d-1/\alpha}}{G'_{\infty,n}(0)} \sum_{t=1}^n (G_n(X_t) \ - \ \mathbf{E}[G_n(X_0)]) \ - \ n^{-d-1/\alpha} \sum_{t=1}^n X_t \xrightarrow{L^r(\mathbf{P})} 0.$$

In the next section, we apply these moment bounds in order to obtain general central limit theorems.

3.2 General central limit theory

Recall, from Section 3.1, that if $n^{\kappa(\gamma,r)-d-1/\alpha}/G'_{\infty,n}(0) \to 0$ as $n \to \infty$, then Theorem 3.3 yields

$$\frac{n^{-d-1/\alpha}}{G'_{\infty,n}(0)} \sum_{t=1}^n (G_n(X_t) \ - \ \mathbf{E}[G_n(X_0)]) \ - \ n^{-d-1/\alpha} \sum_{t=1}^n X_t \xrightarrow{\mathbf{P}} 0 \qquad \text{as } n \to \infty.$$

In particular, by Slutsky's theorem, the asymptotic behavior of the subordinated long memory linear time series reduces to that of $n^{-d-1/\alpha} \sum_{t=1}^n X_t$. The same conclusion holds if, with the notation of (4), $\gamma_G < \gamma_0$ and $n^{\kappa_0+\delta-d-1/\alpha}/G'_{\infty,n}(0) \to 0$ as $n \to \infty$, by Theorem 3.5. This leads to the following central limit theorem.

Corollary 3.6 (Central limit theorem)**.** *Assume that at least one of the following sets of conditions is met:*

1. *(i) the assumptions of Theorem 3.3 hold*

   (ii) $\mathbf{E}[G_n(X_0)] \neq 0$ and $G'_{\infty,n}(0) \neq 0$, for sufficiently large n

(iii) there exist γ, r satisfying Condition (3) *such that* $n^{\kappa(\gamma,r)-d-1/\alpha}/G'_{\infty,n}(0) \to 0$ *as* $n \to \infty$, *where* $\kappa(\gamma, r) = 1 + 1/r - (1-d)(1+\gamma)$.

2. *(i) the assumptions of Theorem 3.5 hold*

   (ii) $\mathbf{E}[G_n(X_0)] \neq 0$ *and* $G'_{\infty,n}(0) \neq 0$, *for sufficiently large* n

   (iii) for some $\delta > 0$ *it holds that* $n^{\kappa_0+\delta-d-1/\alpha}/G'_{\infty,n}(0) \to 0$ *as* $n \to \infty$, *where* κ_0 *is defined in* (5).

Then, as $n \to \infty$,

$$n^{1-(d+1/\alpha)} \cdot \frac{\mathbf{E}[G_n(X_0)]}{G'_{\infty,n}(0)} \cdot \frac{1}{n} \sum_{t=1}^{n} \left(\frac{G_n(X_t)}{\mathbf{E}[G_n(X_0)]} - 1 \right) \xrightarrow{\mathrm{d}} Z_\alpha,$$

where Z_α *is a symmetric* α*-stable random variable defined in Theorem 3.1.*

Remark 3.7 (On the rate of convergence in Corollary 3.6)**.** Since we consider G_n rather than fixed G, the rate of convergence in Corollary 3.6 is also governed by the ratio of $\mathbf{E}[G_n(X_0)]$ and $G'_{\infty,n}(0)$. This can lead to markedly different rates, which is best illustrated by considering again $G_n(x) = \mathbb{1}\{x > u_n\}$, where

$$\frac{\mathbf{E}[G_n(X_0)]}{G'_{\infty,n}(0)} = \frac{\mathbf{P}[X_0 > u_n]}{f_{X_0}(u_n)}.$$

We compare the case of standard Gaussian innovations with that of $S\alpha S$ innovations with $\alpha \in (1, 2)$, covered by Assumption 3(i) and (ii), respectively, so that the stationary distribution of (X_t) is Gaussian and $S\alpha S$ with $\alpha \in (1, 2)$, respectively. For simplicity, suppose that $\sum_{j=0}^{\infty} a_j^2 = 1$. Then by Mills' ratio it follows in the first case

$$\frac{\mathbf{P}[X_0 > u_n]}{f_{X_0}(u_n)} \sim \frac{1}{u_n}.$$

This is in line with the intuition that the total rate should slow down when considering only extremes. In the second case, however, by Karamata's theorem (see [12, Theorem B.1.5]) it holds that

$$\frac{\mathbf{P}[X_0 > u_n]}{f_{X_0}(u_n)} \sim \frac{u_n}{\nu}.$$

This is surprising, because compared with the classical long memory rate $1/n^{1-(d+1/\alpha)}$ the total rate speeds up when looking only at extremes.

The next section discusses this phenomenon and other examples in detail.

3.3 Central limit theory for Peaks-over-Threshold estimators (heavy tails)

We first discuss the case when the stationary distribution of (X_t) is heavy-tailed with index $\nu > 1$. By [29, Equation (15.3.5)], this will be the case if the ε_t are heavy-tailed with index $\nu > 1$, and an application of [8, Theorem 1.7.2] yields that, if f_{X_0} is ultimately decreasing, then $f_{X_0} \in \mathrm{RV}_{-\nu-1}$. In this setting, we have the following consequence of Corollary 3.6; here and throughout $\log(x)_+ = \log(x) \vee 0 = \log(x)\mathbb{1}\{x > 1\}$.

Corollary 3.8 (Applications to heavy tails – deterministic thresholds)**.** *Let Assumptions 1 and 3 hold, and assume that* $f_{X_0} \in \mathrm{RV}_{-\nu-1}$ *with* $\nu > 1$*. If* $u_n \to \infty$ *is such that for some* $\delta > 0$, $u_n = \mathrm{o}(n^{(d+1/(\nu\wedge 2)-\kappa_0-\delta)/(\nu+1)})$, *where* κ_0 *is defined in* (5) *then, as* $n \to \infty$,

$$n^{1-(d+1/(\nu\wedge 2))} u_n \cdot \left(\frac{\sum_{t=1}^n \mathbb{1}\{X_t > u_n\}}{n\mathbf{P}[X_0 > u_n]} - 1 \right) \xrightarrow{\mathrm{d}} \nu Z_{\nu\wedge 2}$$

and

$$n^{1-(d+1/(\nu\wedge 2))} u_n \cdot \left(\frac{\sum_{t=1}^n \log(X_t/u_n)_+}{n\mathbf{P}[X_0 > u_n]} - \mathbf{E}[\log(X_0/u_n) \mid X_0 > u_n] \right) \xrightarrow{\mathrm{d}} \frac{\nu}{\nu+1} Z_{\nu\wedge 2}.$$

In a remark after [5, Theorem 4.17] the authors point out that (for a fixed subordination function G) a stable limit may be obtained from summation of bounded random variables. This observation extends to our work, where the subordination function depends on the sample size.

Remark 3.9 (Equivalent form of Corollary 3.8)**.** An equivalent form of the second convergence in Corollary 3.8, closer to Corollary 3.6, is

$$n^{1-(d+1/(\nu\wedge 2))} u_n \cdot \left(\frac{\sum_{t=1}^n \log(X_t/u_n)_+}{n\mathbf{E}[\log(X_0/u_n)_+]} - 1 \right) \xrightarrow{\mathrm{d}} \frac{\nu^2}{\nu+1} Z_{\nu\wedge 2}.$$

This is due to the well-known convergence $\mathbf{E}[\log(X_0/u) \mid X_0 > u] \to 1/\nu$ as $u \to \infty$.

In statistical practice, the choice of the threshold u_n is crucial. Typical solutions [36] are data-driven, and it is therefore more common to have random instead of deterministic thresholds. In theory, one way of shifting from deterministic to random thresholds is to prove results for deterministic thresholds first and, often by means of empirical process theory, extend them to random thresholds. The Hill estimator (among others) has enough structure to avoid reliance on empirical process theory, which can be technically delicate. Our approach is based on a derandomization device (Lemma C.1 in the supplement), specifically avoiding empirical process theory, and proving directly the convergence of the random threshold version based on the joint convergence of the deterministic threshold version and the order statistic $X_{n-k:n}$. In this way, the next result, which is the counterpart of Corollary 3.8 for random thresholds, is obtained by combining Lemma C.1 and Lemma D.1 in the supplement.

Corollary 3.10 (Application to heavy tails – random thresholds)**.** *Let the Assumptions of Corollary 3.8 hold. Then, with* $k = \lfloor n\mathbf{P}[X_0 > u_n] \rfloor$, *it holds that, as* $n \to \infty$,

$$n^{1-(d+1/(\nu\wedge 2))} q_{X_0}\left(1 - \frac{k}{n}\right)$$
$$\cdot \left(\frac{\sum_{i=1}^k \log(X_{n-i+1:n}/X_{n-k:n})}{n\mathbf{P}[X_0 > u_n]} - \mathbf{E}[\log(X_0/u_n) \mid X_0 > u_n] \right) \xrightarrow{\mathrm{d}} \frac{1}{\nu+1} Z_{\nu\wedge 2}.$$

Remark 3.11 (Comparison with known results in the literature I)**.** As in the case of the empirical tail function in the first display of Corollary 3.8, the Hill estimators with deterministic thresholds (second display of Corollary 3.8) and with random thresholds (Corollary 3.10) show faster speeds compared with classical convergence rates in long memory settings. In the deterministic threshold case (Corollary 3.8), the Hill estimator shows the same increased speed as the empirical tail function, namely, with the extra factor u_n. In the random threshold case (Corollary 3.10), the factor of increase $q_{X_0}(1 -$

$k/n)$ may be seen as the deterministic counterpart of the random threshold $X_{n-k:n}$, since $X_{n-k:n}/q_{X_0}(1-k/n) \to_{\mathbf{P}} 1$. This follows from the convergence of the empirical tail function in the first display of Corollary 3.8 along with inverting this function using Lemma C.2 in the supplement. It stands in sharp contrast with what is observed in i.i.d. and short memory settings.

Although the limiting structure of the Hill estimator is the same for deterministic and random thresholds, the asymptotic scale is lower in the random case. Specifically, it shifts from $\nu/(\nu+1)$ to $1/(\nu+1)$, where, as $\nu \to \infty$, the former converges to 1, while the latter tends to 0. This suggests a phase transition (in the sense of [35]) in the case of random thresholds, occurring as we pass from heavy to light tails, that is, in the limit $\nu \to \infty$. This motivates a second set of examples featuring light tails.

3.4 Central limit theory for Peaks-over-Threshold estimators (light tails)

In the light-tailed case, the asymptotic behavior of PoT estimators should be expected to be different. For example, one has then $\mathbf{E}[\log(X_0/u) \mid X_0 > u] \to 0$ as $u \to \infty$. This is crucial when it comes to investigating the convergence of the moment estimator of the extreme value index [14], which is partly built upon the Hill estimator; see [12, Section 3.5] for a review.

Corollary 3.12 (Applications to light tails – deterministic thresholds)**.** *Let Assumptions 1 and 3 hold, and suppose that $f_{X_0}(x) = \omega(x)\exp(-|x|^\beta/\beta)$ for some $\beta > 0$, where $\omega \in C^1(\mathbb{R})$ is a symmetric function satisfying $|x|^{\beta-1}\omega(x) \to 0$ as $x \to 0$, $\omega(x) > 0$ for x large enough, and with derivative ω' either regularly varying at infinity or constant equal to 0 for $|x|$ large enough. If $u_n \to \infty$ is such that for some $\delta > 0$, $u_n = (-\beta\log(n)(\kappa_0 + \delta - 1/2 - d) + o(1))^{1/\beta}$, where κ_0 is defined as in* (5) *with $\alpha = 2$, then, as $n \to \infty$,*

$$n^{1/2-d}u_n^{1-\beta} \cdot \left(\frac{\sum_{t=1}^n \mathbb{1}\{X_t > u_n\}}{n\mathbf{P}[X_0 > u_n]} - 1\right) \xrightarrow{\mathrm{d}} Z_2$$

and

$$n^{1/2-d}u_n^{1-\beta} \cdot \left(\frac{\sum_{t=1}^n \log(X_t/u_n)_+}{n\mathbf{P}[X_0 > u_n]} - \mathbf{E}[\log(X_0/u_n) \mid X_0 > u_n]\right) \xrightarrow{\mathrm{d}} Z_2.$$

The factor $u_n^{1-\beta}$ allows for interpolation between slowly decaying (e.g., polynomial) and rapidly decaying (e.g., exponential) tails, depending on β. For instance, $\beta = 1$ corresponds to Laplace-type tails, while $\beta = 2$ gives Gaussian decay. The assumptions of Corollary 3.12 can be verified easily for normally distributed innovations, that is, where $\beta = 2$. We highlight this special case in the following corollary.

Corollary 3.13 (Gaussian innovations – deterministic thresholds)**.** *Let the ε_t be i.i.d. Gaussian centered, and suppose that Assumption 1 holds. If $u_n \to \infty$ is such that for some $\delta > 0$, $u_n = (-2\log(n)(\kappa_0 + \delta - 1/2 - d) + o(1))^{1/2}$, where κ_0 is defined as in* (5) *with $\alpha = 2$, then, as $n \to \infty$,*

$$\frac{n^{1/2-d}}{u_n} \cdot \left(\frac{\sum_{t=1}^n \mathbb{1}\{X_t > u_n\}}{n\mathbf{P}[X_0 > u_n]} - 1\right) \xrightarrow{\mathrm{d}} Z_2$$

and

$$\frac{n^{1/2-d}}{u_n} \cdot \left(\frac{\sum_{t=1}^n \log(X_t/u_n)_+}{n\mathbf{P}[X_0 > u_n]} - \mathbf{E}[\log(X_0/u_n) \mid X_0 > u_n]\right) \xrightarrow{\mathrm{d}} Z_2.$$

Remark 3.14 (Comparison with known results in the literature II)**.** Another comparison with the i.i.d. setting is worthwhile here. On the one hand, when the X_i are i.i.d. with density satisfying $f_{X_0}(x) = \omega(x)\exp(-|x|^\beta/\beta)$ for some $\beta > 0$ and a function ω converging to a positive constant ω_0 at infinity, then a repeated use of Lemma 1 in [38] and a straightforward application of the Lindeberg central limit theorem yield

$$\sqrt{n\mathbf{P}[X_0 > u_n]}u_n^\beta \left(\frac{\sum_{t=1}^n \log(X_t/u_n)_+}{n\mathbf{P}[X_0 > u_n]} \;-\; \mathbf{E}[\log(X_0/u_n) \mid X_0 > u_n] \right) \xrightarrow{\text{d}} \mathcal{N}(0, \sigma^2)$$

when $n\mathbf{P}[X_0 > u_n] \to \infty$, where $\sigma^2 = \sigma^2(\beta, \omega_0)$ can be explicitly calculated. On the other hand, when the X_i are i.i.d. heavy-tailed with $f_{X_0} \in \mathrm{RV}_{-\nu-1}$, then another straightforward application of the Lindeberg central limit theorem yields

$$\sqrt{n\mathbf{P}[X_0 > u_n]} \left(\frac{\sum_{t=1}^n \log(X_t/u_n)_+}{n\mathbf{P}[X_0 > u_n]} \;-\; \mathbf{E}[\log(X_0/u_n) \mid X_0 > u_n] \right) \xrightarrow{\text{d}} \mathcal{N}(0, 2/\nu^2)$$

when $n\mathbf{P}[X_0 > u_n] \to \infty$. In other words, when the X_i are i.i.d., the speed of convergence goes from

$$\sqrt{n\mathbf{P}[X_0 > u_n]} \sim \sqrt{n u_n^{-1/\nu}}$$

in the heavy-tailed setting to

$$\sqrt{n\mathbf{P}[X_0 > u_n]}u_n^\beta \sim \sqrt{n u_n^{1+\beta} \exp(-u_n^\beta/\beta)}$$

in the light-tailed setting. The exponentially decaying term $\exp(-u_n^\beta/\beta)$ dramatically slows down the speed as one goes from heavy to light tails. By contrast, in the long memory case, the speed of convergence goes from $n^{1/2-d}u_n$ in the heavy-tailed setting (when $\nu > 2$) to $n^{1/2-d}u_n^{1-\beta}$ in the light-tailed case. The slowdown is thus much more pronounced in the i.i.d. setting than in the long memory case.

As in Section 3.3, we finally consider an analogue of Corollary 3.10 in the light-tailed context.

Corollary 3.15 (Application to light tails – random thresholds)**.** *Under the Assumptions of Corollary 3.12 and with the notation of Corollary 3.10 it holds, as $n \to \infty$,*

$$n^{1/2-d} \left(q_{X_0}\left(1 - \frac{k}{n}\right)\right)^{1-\beta} \cdot \left(\frac{\sum_{i=1}^k \log(X_{n-i+1:n}/X_{n-k:n})}{n\mathbf{P}[X_0 > u_n]} \;-\; \mathbf{E}[\log(X_0/u_n) \mid X_0 > u_n] \right) \xrightarrow{\mathbf{P}} 0\,.$$

The analogous result for heavy-tails (Corollary 3.10) suggests that there is a phase transition in the limit between heavy- and light tails. Observe that the speed coming from deterministic thresholds and light tails (Corollary 3.12) is too slow to identify a non-trivial limiting structure for light tails and random thresholds (Corollary 3.15). This hints at a weakening or even absence of a long memory effect in this case. Absence of long memory has indeed been observed in a similar setting by [27], where the authors recovered i.i.d. behavior with the standard speed $\sqrt{k}$.

4 Numerical illustrations

4.1 Main objective and setting

The purpose of this section is to illustrate the central limit theorem in Corollary 3.6 when the innovations are heavy-tailed. Our motivation is to get a grasp of the extent to which the asymptotic distributions of the PoT estimators obtained in Section 3.3 show up in finite samples. To the best of our knowledge, there is no exact method for simulating long memory linear time series with heavy tails. We therefore use a (long enough) truncated coefficient sequence,

$$a_0 = 0 \qquad \text{and} \qquad a_j \ = \ j^{-(1-d)}\,, \qquad j = 1, \dots, p = 10^8\,,$$

with $d = 0.1$, and symmetric α-stable innovations ε_t with $\alpha = 1.9$ and scale parameter $\eta = 1$. We simulate $N = 10{,}000$ i.i.d. copies of $(X_t, t = 1, \dots, n = p/10 = 10^7)$, using GPU acceleration; the sample size $n = p/10$ is such that the number of coefficients p is large enough compared with the length of the time series, in order to simulate long memory behavior of the series $X_t = \sum_{j=1}^{\infty} j^{-(1-d)} \varepsilon_{t-j}$.

4.2 Slow convergence of pre-asymptotic scaling factors

The main ingredient in our central limit theory is Theorem 3.1, which provides the weak convergence behavior of partial sums of the X_t, namely,

$$n^{-d-1/\alpha} \sum_{t=1}^{n} X_t \ \xrightarrow{\mathrm{d}} \ Z_\alpha \qquad \text{as } n \to \infty,$$

where Z_α is symmetric α-stable with scale

$$\eta \ := \ \frac{c_a}{d} \cdot \left(A \cdot \frac{\Gamma(2-\alpha)\cos(\pi\alpha/2)}{1-\alpha} \right)^{1/\alpha} \left(\int_{-\infty}^{1} \left((1-v)_+^d - (-v)_+^d \right)^\alpha \, \mathrm{d}v \right)^{1/\alpha} .$$

Observe that in our case $c_a = 1$ so that, by symmetry of the α-stable innovations,

$$\eta = \frac{1}{d} \left(\int_{-\infty}^{1} \left((1-v)_+^d - (-v)_+^d \right)^\alpha \, \mathrm{d}v \right)^{1/\alpha} .$$

For $d = 0.1$, $\alpha = 1.9$, by numerical integration we get

$$\eta \ = \ \eta_{d=0.1,\alpha=1.9} \ = \ 10 \left(\int_{-\infty}^{1} \left((1-v)_+^{0.1} - (-v)_+^{0.1} \right)^{1.9} \, \mathrm{d}v \right)^{1/1.9} \ \approx \ 9.344\,.$$

To assess how fast the asymptotic scale η is approached in finite samples, one may write the rescaled partial sum as

$$\begin{aligned} n^{-d-1/\alpha} \frac{1}{\eta} \sum_{t=1}^{n} X_t \ &= \ n^{-d-1/\alpha} \frac{1}{\eta} \sum_{t=1}^{n} \sum_{j=1}^{\infty} j^{-(1-d)} \varepsilon_{t-j} \\ &= \ n^{-d-1/\alpha} \frac{1}{\eta} \left(\sum_{t=1}^{n-1} \varepsilon_{n-t} \sum_{j=1}^{t} j^{-(1-d)} \ + \ \sum_{t=0}^{\infty} \varepsilon_{-t} \sum_{j=1}^{n} (j+t)^{-(1-d)} \right) . \end{aligned}$$

Since the ε_t are i.i.d. symmetric α-stable, so is this rescaled partial sum, with scale

$$s = s_{n,d,\alpha} = n^{-d-1/\alpha}\frac{1}{\eta}\left(\sum_{t=1}^{n-1}\left(\sum_{j=1}^{t} j^{-(1-d)}\right)^{\alpha} + \sum_{t=0}^{\infty}\left(\sum_{j=1}^{n}(j+t)^{-(1-d)}\right)^{\alpha}\right)^{1/\alpha}. \quad (6)$$

If η were to be a correct approximation of the scale of $n^{-d-1/\alpha}\sum_{t=1}^{n} X_t$, then $s_{n,d,\alpha}$ would be close to 1. Numerical approximations, however, show that the ratio $s_{n,d,\alpha}$ converges to 1 slowly, suggesting that η does not correctly approximate the scale of $n^{-d-1/\alpha}\sum_{t=1}^{n} X_t$. We illustrate this in the top panel of Figure 1, where even for the large value $n = 10^6$ one finds $s \approx 0.72$.

Another scaling quantity, namely $\mathbf{E}[G_n(X_0)]/G'_{\infty,n}(0)$, appears in Corollary 3.6. For exceedance counts, using the indicator function $G_n(x) = \mathbb{1}\{x > u_n\}$, this ratio is equal to $\mathbf{P}[X_0 > u_n]/f_{X_0}(u_n)$, which is known given the distribution of X_0. But even for relatively simple extensions such as $G_n(x) = (\log(x) - \log(u_n))\mathbb{1}\{x > u_n\}$, the exact calculation of $\mathbf{E}[G_n(X_0)]$ and $G'_{\infty,n}(0)$ becomes an issue. This is because the resulting integrals typically do not have a simple closed form. A solution is an approximation by Monte Carlo simulation, using for example

$$\mathbf{E}[G_n(X_0)] \;=\; \mathbf{E}[\log(X_0/u_n)_+] \;\approx\; \widehat{\mathbf{E}}[\log(X_0/u_n)_+] \;=\; \frac{1}{m}\sum_{i=1}^{m}\log(X_0^{(i)}/u_n)\mathbb{1}\{X_0^{(i)} > u_n\}$$

and, by partial integration (see the proof of Corollary 3.8),

$$\begin{aligned} G'_{\infty,n}(0) \;&=\; \int_{u_n}^{\infty}\frac{f_{X_0}(x)}{x}\,\mathrm{d}x \;=\; \mathbf{E}[X_0^{-1}\mathbb{1}\{X_0 > u_n\}] \;\approx\; \widehat{\mathbf{E}}[X_0^{-1}\mathbb{1}\{X_0 > u_n\}] \\ &=\; \frac{1}{m}\sum_{i=1}^{m}\frac{1}{X_0^{(i)}}\mathbb{1}\{X_0^{(i)} > u_n\}\,, \end{aligned}$$

where $X_0', \ldots, X_0^{(m)}$, $m \in \mathbb{N}$ are i.i.d. samples of X_0; here we take $m = 10^7$.

It should be highlighted that in statistical applications, the distribution of X_0 is not assumed to be known. A workaround is to combine Karamata's and Slutsky's theorem, as is done in the proof of Corollary 3.8, to replace $\mathbf{E}[G_n(X_0)]/G'_{\infty,n}(0)$ by an asymptotic equivalent. For example, for $G_n(x) = \mathbb{1}\{x > u_n\}$,

$$\frac{\mathbf{E}[G_n(X_0)]}{G'_{\infty,n}(0)} = \frac{\mathbf{P}[X_0 > u_n]}{f_{X_0}(u_n)} \quad\text{and}\quad \frac{u_n}{\nu}$$

are asymptotically equivalent. As illustrated in the bottom panel of Figure 1, when X_0 has a stable distribution these two quantities may, however, differ substantially, even at fairly high thresholds u_n of order, say, $q_{X_0}(0.999)$. This is largely due to the speed of convergence of $\mathbf{P}[X_0 > u_n]$ to its power law equivalent, where the remainder term is of order $O(u_n^{-\alpha})$ and vanishes slowly as $n \to \infty$; see [42, Corollary 2 p.94] for details. It is thus not advisable to use the asymptotic equivalent u_n/ν in place of $\mathbf{P}[X_0 > u_n]/f_{X_0}(u_n)$ when $G_n(x) = \mathbb{1}\{x > u_n\}$. More accurate asymptotic equivalents could be obtained by estimating the remainder term in Karamata's theorem, for instance using second-order extended regular variation [12, Sections 2.3 and B.3]. The construction of such refined asymptotic equivalents is outside of the scope of this work.

This discussion motivates, in our particular case of stable distributed X_0, comparing the finite-sample distributions of

$$n^{1-(d+1/\alpha)}\frac{1}{\eta}\frac{\mathbf{P}[X_0 > u_n]}{f_{X_0}(u_n)} \cdot \left(\frac{\sum_{t=1}^{n}\mathbb{1}\{X_t > u_n\}}{n\mathbf{P}[X_0 > u_n]} \;-\; 1\right) \quad (7)$$

and

$$n^{1-(d+1/\alpha)}\frac{1}{\eta}\frac{\widehat{\mathbf{E}}[\log(X_0/u_n)_+]}{\widehat{\mathbf{E}}[X_0^{-1}\mathbb{1}\{X_0>u_n\}]}\cdot\left(\frac{\sum_{t=1}^{n}\log(X_t/u_n)_+}{n\mathbf{E}[\log(X_0/u_n)_+]}\;-\;1\right) \tag{8}$$

with the symmetric unit stable distribution with index α.

4.3 Histograms of exceedance counts and Hill estimators

The top rows of Figures 2 and 3 report histograms of $N = 10{,}000$ copies from (7) and (8), respectively, when the length n of the time series (X_t) belongs to $\{10^3, 10^5, 10^7\}$. It can be seen there that the structure of the histograms roughly matches that of a symmetric unit stable distribution, only with incorrect scale. From the discussion in Section 4.2, it follows that a possible reason for that is incorrect scaling of the partial sums; the middle rows of Figures 2 and 3 show the histograms of these same realizations divided by the scale parameter $\widehat{s}$ of a stable distribution fitted to the realizations of the rescaled partial sums $n^{-d-1/\alpha}\eta^{-1}\sum_{t=1}^{n}X_t$. This fit and subsequent stable fits are obtained using the `stableFit` function from the `fBasics` package in `R`, with argument `type="q"`, based on the method proposed by [31]. For exceedance counts (Figure 2), this scaling results in a much better fit of the symmetric unit stable distribution, with only minor deviations for $n = 10^3$. For the Hill estimator (Figure 3), this still does not result in a good fit for all values of n, and lack of symmetry due to skewness can be observed. To separate scaling and skewness issues, we report in the bottom row of Figures 2 and 3 the histograms of the realizations of (7) and (8), rescaled not by $\widehat{s}$, but by a scale parameter $\widetilde{s}$ obtained by fitting a stable distribution directly to the realizations. In the case of the Hill estimator we confirm the presence of skewness, and while its degree seems to decrease from $n = 10^3$ (bottom row of Figure 3, left panel) to $n = 10^5$ (bottom row of Figure 3, middle panel), this is not the case from $n = 10^5$ to $n = 10^7$ (bottom row of Figure 3, right panel). As a conclusion, it seems that, in the simplest case of exceedance counts, the finite-sample distribution is dictated by the behavior of partial sums of X_t, while in more difficult cases such as the Hill estimator, a degree of skewness appears, which could be due to the particular structure of the subordinating function G_n.

5 Discussion

Our work yields a variety of unexpected results, already highlighted in the previous sections. We now discuss these findings, offer interpretations, and place them in a broader context. We highlighted that the ratio $\mathbf{E}[G_n(X_0)]/G'_{\infty,n}(0)$, which appears in the rate of the central limit theorem Corollary 3.6, is widely different in the heavy- and light-tailed cases when $G_n(x) = \mathbb{1}\{x > u_n\}$. Subsequently, we showed that this applies also to the Hill estimator, where $G_n(x) = \log(x/u_n)\mathbb{1}\{x > u_n\}$. This raises the question of how the heavy- and light-tailed setting can differ so markedly in this regard, and how the increased rate in the heavy-tailed case arises. We offer an interpretation based on the observation that heavy tails exhibit strong asymptotic dependence, leading to extremal clustering, while light tails, such as those of the normal distribution, exhibit asymptotic independence. To be more precise, we define asymptotic (in)dependence by

$$\mathbf{P}[X_t > u \mid X_0 > u] \;\overset{u\to\infty}{\to}\; \chi(t)\begin{cases} = \;0 & \text{for all } t>0 \text{ when asymptotic independent,}\\ > \;0 & \text{for some } t>0 \text{ when asymptotic dependent.}\end{cases}$$

The quantity $\chi(t)$ is called tail-dependence coefficient. It is linked to the tail process (Y_t) of (X_t) (see [4] for a reference to tail processes) via the identity $\chi(t) = \mathbf{P}[Y_t > 1]$. Let us first discuss the asymptotically dependent, that is, heavy-tailed setting. To control extremal clustering, all PoT results in the literature, to the best of our knowledge, assume some form of anti-clustering condition. For example, one assumes that for an intermediate rate (r_n), a threshold sequence (u_n), and all $s,t>0$ it holds

$$\lim_{m\to\infty}\limsup_{n\to\infty}\frac{1}{\mathbf{P}[|X_0|>u_n]}\sum_{j=m}^{r_n}\mathbf{P}[|X_0|>u_n s, |X_j|>u_n t] \;=\; 0\,.$$

This condition is called $\mathcal{S}(r_n,u_n)$ in [29, Section 9.2 p. 243]. According to (15.3.33) therein, in the case of heavy-tailed linear time series there exists a sufficient condition for $S(r_n,u_n)$ to hold, that is,

$$\sum_{j=1}^{\infty} j\cdot|a_j|^{\delta} \;<\; \infty \qquad \text{for some } \delta\in(0,\alpha)\cap(0,2]\,. \tag{9}$$

It, however, follows immediately that under Assumption 1 this simpler condition is never satisfied. We even have that a sufficient condition for $\mathcal{S}(r_n,u_n)$ to fail is that $\sum_{t=1}^{\infty}\mathbf{P}[|Y_t|>1]=\infty$, where (Y_t) is the tail process of the time series. This follows from applying Fatou's lemma, that is,

$$\sum_{t=1}^{\infty}\mathbf{P}[|Y_t|>1] \;\le\; \liminf_{n\to\infty}\frac{1}{\mathbf{P}[|X_0|>u_n]}\sum_{j=1}^{r_n}\mathbf{P}[|X_0|>u_n, |X_j|>u_n]\,.$$

With [29, (15.3.34)], if moreover the $|a_j|$ are decreasing, then $\sum_{t=1}^{\infty}\mathbf{P}[|Y_t|>1]<\infty$ is equivalent to

$$\sum_{j=1}^{\infty} j\cdot|a_j|^{\nu} \;<\; \infty \quad \text{which under Assumption 1 becomes} \quad \sum_{j=1}^{\infty} j^{1-(1-d)\nu} \;<\; \infty\,.$$

Since this holds if and only if $\nu\ge 2/(1-d)$, we know that the anti-clustering condition fails for such combinations of (ν,d). In particular, it fails for all $\nu\in(1,2)$. It therefore seems plausible that long clusters of extremes may accelerate the speed of convergence obtained in Corollary 3.6.

Having discussed the asymptotically dependent setting and its connection to anti-clustering conditions, we now turn to the asymptotically independent case featured in our work (Corollary 3.13). In the case of Gaussian innovations, and thus Gaussian time series, we observe behavior more consistent with existing literature: for deterministic thresholds, focusing on extremes, that is, reducing the effective number of summands, slows down the speed of convergence. Since the time series is asymptotically independent, anti-clustering conditions hold, and we return to the standard setting.

Another point we want to emphasize is the sensitivity of PoT estimators to threshold choice. The assumption $u_n=\mathrm{o}(n^{(d+1/(\nu\wedge 2)-\kappa_0-\delta)/(\nu+1)})$ in Corollary 3.8, where κ_0 is defined in (5), reflects that the threshold u_n should be chosen carefully. In simulations not reported here, we tried using higher quantile levels such as $q_{X_0}(1-n^{-1/3.5})$, and we observed that this could make the finite-sample distribution even less close to the asymptotic limit. As a consequence, one should not blindly base inference procedures upon the asymptotic distribution obtained in our results. Similar problems have been reported already in i.i.d. extreme value settings irrespective of whether the second moment is finite or not [10, 11, 33]. A solution may be to develop dedicated self-normalization or subsampling procedures, as in [2, 24, 34], or to rely on a smaller, more tractable class of models such as ARFIMA models [40]. On a different note, but still about threshold choice, observe that, unlike in the i.i.d. or weakly dependent case, results for deterministic and random thresholds differ in our long memory setting. While [27] show that, in an asymptotically independent setting, long memory effects for the Hill estimator in stochastic volatility models arise only with deterministic thresholds, our work allows for a comparison of both asymptotically dependent and independent settings (Corollaries 3.10 and 3.15). We note that stochastic volatility models with asymptotic dependence exist [25]; see also [29, p.487] for a discussion of the literature. Future work could take these models into consideration.

Acknowledgments

Financial support from the French CNRS within the project "Extreme value analysis of time series through de-randomization techniques", funded through the IEA program, is gratefully acknowledged. The authors also acknowledge computing support by the state of Baden-Württemberg through bwHPC. G. Stupfler acknowledges further financial support from the French *Agence Nationale de la Recherche* under the grants ANR-23-CE40-0009 (EXSTA project) and ANR-11-LABX-0020-01 (Centre Henri Lebesgue), as well as from the TSE-HEC ACPR Chair "Regulation and systemic risks" and the Chair Stress Test, RISK Management and Financial Steering of the Foundation Ecole Polytechnique. I. Scheffel thanks Julian Weitz who pointed him towards the use of GPU acceleration for simulating linear time series.

References

[1] von Bahr, B., Esseen, C.G., 1965. Inequalities for the rth absolute moment of a sum of random variables, $1 \leq r \leq 2$. The Annals of Mathematical Statistics 36, 299–303.

[2] Bai, S., Taqqu, M.S., 2017. On the validity of resampling methods under long memory. The Annals of Statistics 45, 2365–2399.

[3] Bai, S., Zhang, T., 2024. Tail adversarial stability for regularly varying linear processes and their extensions. Extremes 27, 33–65.

[4] Basrak, B., Segers, J., 2009. Regularly varying multivariate time series. Stochastic Processes and their Applications 119, 1055–1080.

[5] Beran, J., Feng, Y., Ghosh, S., Kulik, R., 2013. Long-Memory Processes: Probabilistic Properties and Statistical Methods. Springer, Berlin, Heidelberg.

[6] Bhardwaj, G., Swanson, N.R., 2006. An empirical investigation of the usefulness of arfima models for predicting macroeconomic and financial time series. Journal of Econometrics 131, 539–578.

[7] Bilayi-Biakana, C., Ivanoff, G., Kulik, R., 2019. The tail empirical process for long memory stochastic volatility models with leverage. Electronic Journal of Statistics 13, 3453–3484.

[8] Bingham, N.H., Goldie, C.M., Teugels, J.L., 1987. Regular Variation. Encyclopedia of Mathematics and Its Applications, Cambridge University Press, Cambridge.

[9] Cao, H., Gao, J., Shao, Y., Sriram, T.N., Wang, W., Wen, F., Zhang, T., 2025+. Tail index estimation for tail adversarial stable time series with an application to high-dimensional tail clustering. Journal of Time Series Analysis.

[10] Daouia, A., Stupfler, G., Usseglio-Carleve, A., 2024. Bias-reduced and variance-corrected asymptotic Gaussian inference about extreme expectiles. Statistics and Computing 34, 130.

[11] Daouia, A., Stupfler, G., Usseglio-Carleve, A., 2025. Corrected inference about the extreme Expected Shortfall in the general Max-Domain of Attraction. *Information and Inference*, to appear (HAL preprint hal-04648262).

[12] De Haan, L., Ferreira, A., 2006. Extreme Value Theory. Springer Series in Operations Research and Financial Engineering, Springer, New York, NY.

[13] Dehling, H., Mikosch, T., Sørensen, M. (Eds.), 2002. Empirical Process Techniques for Dependent Data. Birkhäuser, Boston, MA.

[14] Dekkers, A.L.M., Einmahl, J.H.J., de Haan, L., 1989. A moment estimator for the index of an extreme-value distribution. The Annals of Statistics 17, 1833–1855.

[15] Drees, H., 2003. Extreme quantile estimation for dependent data, with applications to finance. Bernoulli 9, 617–657.

[16] Giraitis, L., Koul, H.L., Surgailis, D., 1996. Asymptotic normality of regression estimators with long memory errors. Statistics & Probability Letters 29, 317–335.

[17] Giraitis, L., Surgailis, D., 1999. Central limit theorem for the empirical process of a linear sequence with long memory. Journal of Statistical Planning and Inference 80, 81–93.

[18] Granger, C.W.J., Joyeux, R., 1980. An introduction to long-memory time series and fractional differencing. Journal of Time Series Analysis 1, 15–29.

[19] Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. The Annals of Statistics 3, 1163–1174.

[20] Ho, H.C., Hsing, T., 1996. On the asymptotic expansion of the empirical process of long-memory moving averages. The Annals of Statistics 24, 992–1024.

[21] Ho, H.C., Hsing, T., 1997. Limit theorems for functionals of moving averages. The Annals of Probability 25, 1636–1669.

[22] Hsing, T., 1991. On tail index estimation using dependent data. The Annals of Statistics 19, 1547–1569.

[23] Huang, H.H., Chan, N.H., Chen, K., Ing, C.K., 2022. Consistent order selection for ARFIMA processes. The Annals of Statistics 50, 1297–1319.

[24] Jach, A.J., McElroy, T., Politis, D.N., 2012. Subsampling inference for the mean of heavy-tailed long-memory time series. Journal of Time Series Analysis 33, 96–111.

[25] Janssen, A., Drees, H., 2016. A stochastic volatility model with flexible extremal dependence structure. Bernoulli 22, 1448–1490.

[26] Koul, H.L., Surgailis, D., 2001. Asymptotics of empirical processes of long memory moving averages with infinite variance. Stochastic Processes and their Applications 91, 309–336.

[27] Kulik, R., Soulier, P., 2011. The tail empirical process for long memory stochastic volatility sequences. Stochastic Processes and their Applications 121, 109–134.

[28] Kulik, R., Soulier, P., 2013. Estimation of limiting conditional distributions for the heavy tailed long memory stochastic volatility process. Extremes 16, 203–239.

[29] Kulik, R., Soulier, P., 2020. Heavy-Tailed Time Series. Springer Series in Operations Research and Financial Engineering, Springer, New York, NY.

[30] Kulik, R., Spodarev, E., 2021. Long range dependence of heavy-tailed random functions. Journal of Applied Probability 58, 569–593.

[31] McCulloch, J.H., 1986. Simple consistent estimators of stable distribution parameters. Communications in Statistics – Simulation and Computation 15, 1109–1136.

[32] Oesting, M., Rapp, A., 2023. Long memory of max-stable time series as phase transition: Asymptotic behaviour of tail dependence estimators. Electronic Journal of Statistics 17, 3316–3336.

[33] Padoan, S., Stupfler, G., 2022. Joint inference on extreme expectiles for multivariate heavy-tailed distributions. Bernoulli 28, 1021–1048.

[34] Romano, J., Wolf, M., 1999. Subsampling inference for the mean in the heavy-tailed case. Metrika 50, 55–69.

[35] Samorodnitsky, G., 2016. Stochastic Processes and Long Range Dependence. Springer Series in Operations Research and Financial Engineering, Springer International Publishing, Cham.

[36] Scarrott, C., MacDonald, A., 2012. A review of extreme value threshold estimation and uncertainty quantification. REVSTAT – Statistical Journal 10, 33–60.

[37] Stupfler, G., 2019. On a relationship between randomly and non-randomly thresholded empirical average excesses for heavy tails. Extremes 22, 749–769.

[38] Stupfler, G., Usseglio-Carleve, A., 2025. Simple sufficient criteria for second-order extended regular variation in the Gumbel domain of attraction: The case of Weibull-tailed distributions. Extremes 28, 393–435.

[39] Surgailis, D., 2004. Stable limits of sums of bounded functions of long-memory moving averages with finite variance. Bernoulli 10, 327–355.

[40] Verma, V., Stoev, S., Chen, Y., 2025+. On the optimal prediction of extreme events in heavy-tailed time series with applications to solar flare forecasting. Journal of Time Series Analysis.

[41] Wang, Y., Liang, Z., Qing, S., Liu, X., Xu, C., 2024. Application of an ARFIMA model to estimate hepatitis C epidemics in Henan, China. The American Journal of Tropical Medicine and Hygiene 110, 404–411.

[42] Zolotarev, V.M., 1986. One-dimensional Stable Distributions. volume 65. American Mathematical Society.

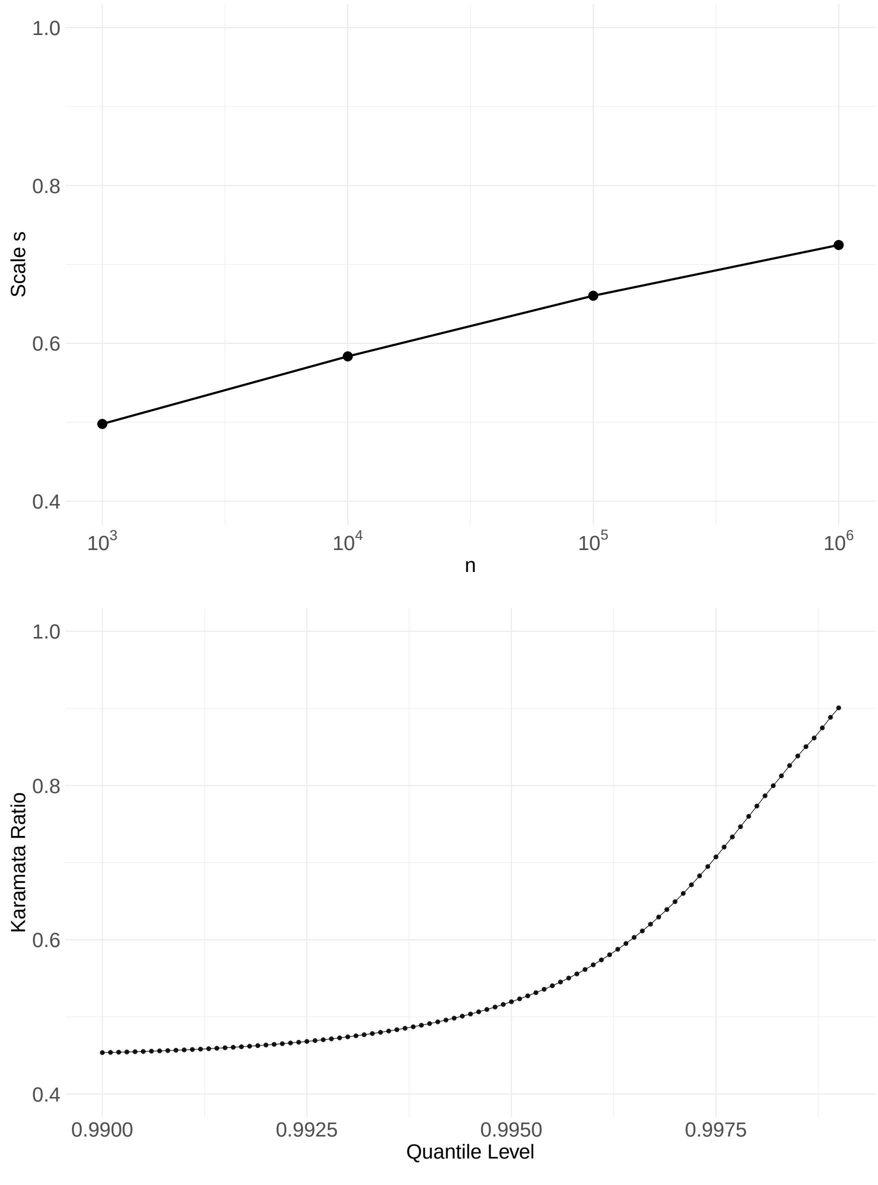

Figure 1: Illustration of the convergence of pre-asymptotic scaling factors. Top: Values of $s_{n,d,\alpha}$ in (6) for $n \in \{10^3, 10^4, 10^5, 10^6\}$, $d = 0.1$ and $\alpha = 1.9$. Bottom: Values of $\alpha\mathbf{P}[X > u]/(uf_X(u))$ for the unit α-stable distribution along the levels $u = q_X(\tau)$ for $\tau \in [0.99, 0.999]$, with $\alpha = 1.9$.

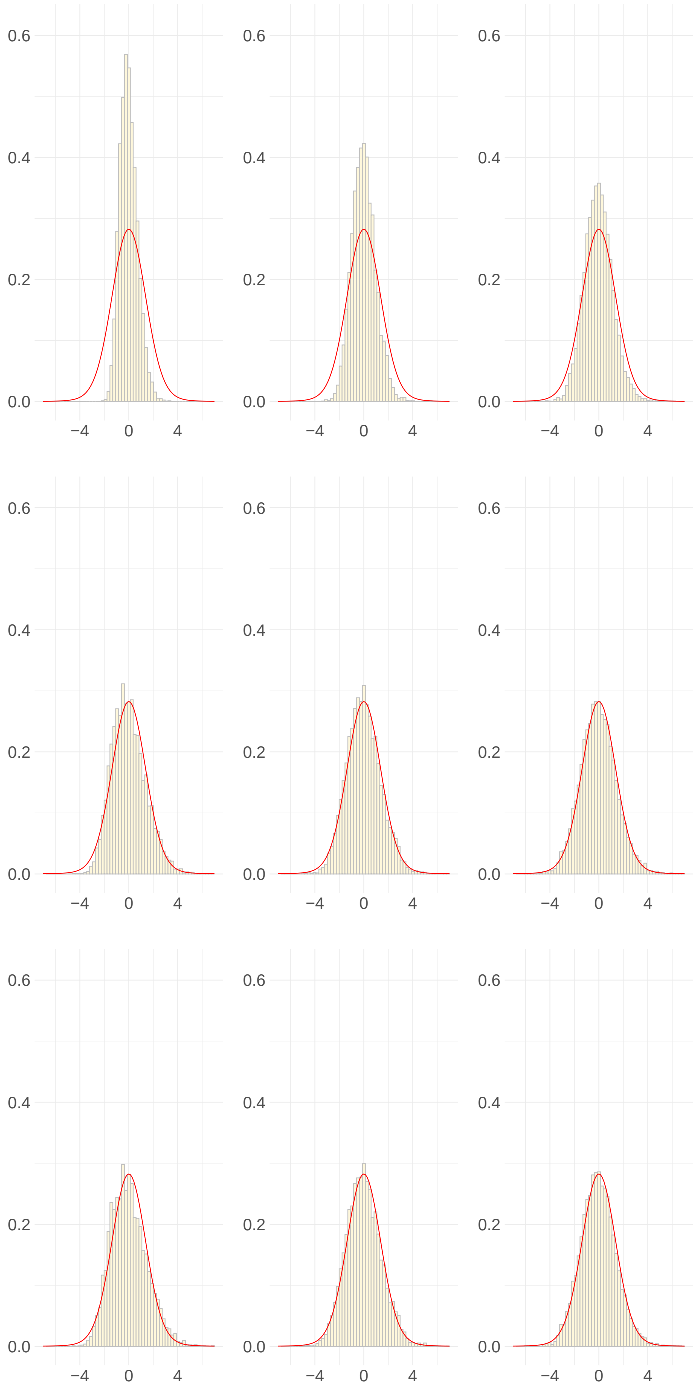

Figure 2: Histograms of $N = 10{,}000$ realizations of the centered and rescaled exceedance counts in (7). Top panels: raw realizations, middle panels: realizations rescaled by the parameter $\widehat{s}$ (of a stable distribution fitted to the partial sums), bottom panels: realizations rescaled by the parameter $\widetilde{s}$ (of a stable distribution fitted directly to these raw realizations). The length of the time series (X_t), generated with $a_j = j^{-(1-d)}$ and symmetric unit α-stable innovations (for $d = 0.1$ and $\alpha = 1.9$), is $n = 10^k$ with $k = 3$ (left panels), 5 (middle panels), 7 (right panels), with $u_n = 1 - n^{-1/5} = 1.38, 2.67, 3.74$.

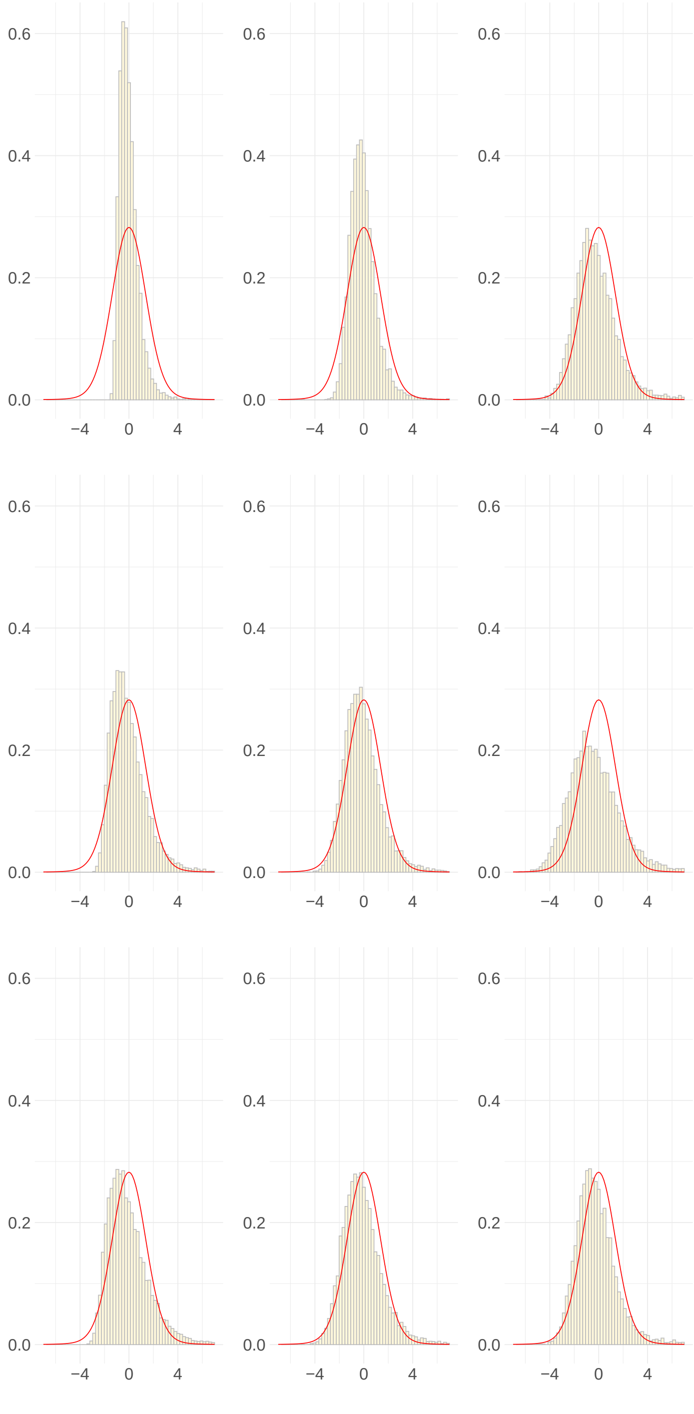

Figure 3: Histograms of $N = 10{,}000$ realizations of the centered and rescaled Hill estimator in (8). Top panels: raw realizations, middle panels: realizations rescaled by the parameter $\widehat{s}$ (of a stable distribution fitted to the partial sums), bottom panels: realizations rescaled by the parameter $\widetilde{s}$ (of a stable distribution fitted directly to these raw realizations). The length of the time series (X_t), generated with $a_j = j^{-(1-d)}$ and symmetric unit α-stable innovations (for $d = 0.1$ and $\alpha = 1.9$), is $n = 10^k$ with $k = 3$ (left panels), 5 (middle panels), 7 (right panels), with $u_n = 1 - n^{-1/5} = 1.38, 2.67, 3.74$.

Supplementary Material to the article Central limit theory for Peaks-over-Threshold partial sums of long memory linear time series

Ioan Scheffel, Marco Oesting, Gilles Stupfler

A Auxiliary results

In this section we collect a host of results that are subsequently used in the proofs, but may also be of independent interest.

A.1 Martingale difference inequality

We start this section by recalling—and refining—a fundamental inequality for martingale difference arrays (cf. [1]) which is satisfied in particular by independent, centered random variables.

Lemma A.1 (Martingale Difference Inequality). *Let $p \in [1,2]$, and let $(Y_{m,i}\,, i = 1,\ldots,m\,, m \in \mathbb{N}) \subset L^p(\mathbf{P})$ be an array of random variables satisfying*

$$\mathbf{E}\left[Y_{m,\ell+1} \,\middle|\, \sum_{i=1}^{\ell} Y_{m,i} \right] \;=\; 0 \qquad \text{for all } 1 \le \ell \le m-1 \text{ and all } m \in \mathbb{N}\,. \tag{10}$$

1. *It holds that*

$$\mathbf{E}\left|\sum_{i=1}^{m} Y_{m,i}\right|^p \;\le\; 2\sum_{i=1}^{m} \mathbf{E}\,|Y_{m,i}|^p \qquad \text{for all } m \in \mathbb{N}\,. \tag{11}$$

2. *If there exists a sequence $(Y_i)_{i\in\mathbb{N}} \subset L^p(\mathbf{P})$ such that*

$$\lim_{m\to\infty} \sum_{i=1}^{m} Y_{m,i} = \sum_{i=1}^{\infty} Y_i \quad \mathbf{P}\text{-almost surely, and} \qquad \liminf_{m\to\infty} \sum_{i=1}^{m} \mathbf{E}|Y_{m,i}|^p \le \sum_{i=1}^{\infty} \mathbf{E}|Y_i|^p\,, \tag{12}$$

then

$$\mathbf{E}\left|\sum_{i=1}^{\infty} Y_i\right|^p \;\le\; 2\sum_{i=1}^{\infty} \mathbf{E}\,|Y_i|^p\ . \tag{13}$$

Proof. Part (i) is exactly [1, Theorem 2]. To prove Part (ii), we apply Fatou's lemma together with (11) and (12) to obtain

$$\begin{aligned}
\mathbf{E}\left|\sum_{i=1}^{\infty} Y_i\right|^p \;=\; \mathbf{E}\left(\liminf_{m\to\infty}\left|\sum_{i=1}^{m} Y_{m,i}\right|^p\right) \;\le\; \liminf_{m\to\infty} \mathbf{E}\left|\sum_{i=1}^{m} Y_{m,i}\right|^p \;&\le\; \liminf_{m\to\infty} 2\sum_{i=1}^{m} \mathbf{E}\,|Y_{m,i}|^p \\
&\le 2\sum_{i=1}^{\infty} \mathbf{E}\,|Y_i|^p\ ,
\end{aligned}$$

as announced. □

A.2 Regularity of distribution functions and their derivatives

For $\gamma \in (0,1)$, we define

$$g_\gamma \;:\; \mathbb{R} \to (0,1], \qquad x \;\mapsto\; g_\gamma(x) \;=\; \frac{1}{(1+|x|)^{\gamma+1}}.$$

For $k \in \mathbb{N}$, we define the truncated time series as

$$X_{t,k} \;:=\; \sum_{j=0}^{k} a_j \varepsilon_{t-j} \qquad \text{and} \qquad \widetilde{X}_{t,k} \;:=\; X_t \;-\; X_{t,k} \;=\; \sum_{j=k+1}^{\infty} a_j \varepsilon_{t-j}\,.$$

By convention, we write $X_{t,\infty} := X_t$. In the following, let f_k denote the density of $X_{0,k}$, and f_∞ denote the density of $X_{0,\infty} = X_0$. Note that $X_{0,k} =_d X_{t,k}$ for all $t \in \mathbb{Z}$, due to stationarity of the sequence (ε_t). The next lemma is needed when dealing with heavy tails ($\alpha \in (1,2)$). It is a substitute for the mean value theorem that may be applied in settings with finite fourth moments (compare [26, 20]). Also note that, as explained in the introduction, we assimilate the case $\alpha = 2$ to fit in the framework of $\alpha \in (1,2)$, which is essentially done here.

Lemma A.2. *Let Assumptions 1 and 3 hold. Then there exists $k_0 \in \mathbb{N}$ such that for all $k > k_0$ the distribution functions of X_0 and $X_{0,k}$ belong to $C^2(\mathbb{R})$. Furthermore, for all $\gamma \in (0, \alpha-1)$ and $r \in (1,2)$ with $1+\gamma < r < \alpha$ and $1/(1-d) < r$, it holds for $x, y \in \mathbb{R}$ with $|x-y| \leq 1$ and all $k > k_0$*

$$|f'_\infty(x)| \;+\; |f'_k(x)| \;\lesssim\; g_\gamma(x)\,, \tag{14}$$
$$|f'_\infty(x) - f'_\infty(y)| \;+\; |f'_k(x) - f'_k(y)| \;\lesssim\; |x-y| \cdot g_\gamma(x)\,, \tag{15}$$
$$|f'_\infty(x) - f'_k(x)| \;\lesssim\; k^{-(1-d)+1/r} \cdot g_\gamma(x)\,, \tag{16}$$

where $\lesssim$ is $\leq$ up to a constant $C_{\gamma,r} > 0$ only depending on γ and r.

Proof. Under Assumption 3(ii), the statement is proved in [26, Lemma 4.2]. Under Assumption 3(i), a proof of a slightly different statement is found in [17, Lemma 2]. Indeed, the only difference is in (14)-(16), that is, [17, Lemma 2] prove that

$$|f'_\infty(x)| \;+\; |f'_k(x)| \;\lesssim\; g_1(x)\,,$$
$$|f'_\infty(x) - f'_\infty(y)| \;+\; |f'_k(x) - f'_k(y)| \;\lesssim\; |x-y| \cdot g_1(x)\,,$$
$$|f'_\infty(x) - f'_k(x)| \;\lesssim\; k^{2d-1} \cdot g_1(x)\,.$$

Since $g_1 \leq g_\gamma$ for $\gamma \in (0,1)$, we immediately recover (14) and (15). The only remaining difference is in the third inequality. For this, note that $k^{2d-1} \;\leq\; k^{-(1-d)+1/r}$ due to

$$(2d-1) \;+\; (1-d) - 1/r \;=\; d \;-\; 1/r \;\leq\; 1 \;-\; \frac{2}{\alpha} \;\leq\; 0$$

since $1/\alpha < 1/r$, $d < 1 - 1/\alpha$, and $\alpha \leq 2$. Inequality (16) follows. $\square$

The next lemma is an analytic tool to deal with g_γ, which may arise after applying Lemma A.2.

Lemma A.3. *Let $\gamma \in (0,1)$. Then the following hold:*

(i) For all $y, z \in \mathbb{R}$,

$$g_\gamma(z+y) \;\lesssim\; g_\gamma(z) \cdot (1 \vee |y|)^{1+\gamma}\,. \tag{17}$$

(ii) Let $a, b \in \mathbb{R}$ with $a < b$. Then, for all $z \in \mathbb{R}$,

$$\int_a^b g_\gamma(z-w)\,\mathrm{d}w \;\lesssim\; g_\gamma(z) \cdot ((b-a) \vee 1)^{1+\gamma}\,. \tag{18}$$

Here $\lesssim$ denotes inequality up to a constant $C_\gamma > 0$ depending only on γ.

Proof. The Inequality (17) follows from [26, Lemma 5.1]. We now prove (18). Let $a, b \in \mathbb{R}$ with $a < b$, and define $u := b - a + 1 \;\geq\; 1$ and $\tilde{z} := z - a + 1$. Then

$$\int_a^b g_\gamma(z-w)\,\mathrm{d}w \;=\; \int_1^u g_\gamma(\tilde{z}-w)\,\mathrm{d}w\,.$$

Since $g_\gamma(\tilde{z}) \lesssim g_\gamma(z)$, it suffices to consider the case $a = 1 < b$. Note that for all $b > 1$, we have $b \;\leq\; 2((b-1) \vee 1)$, and thus it suffices to show for all $z \in \mathbb{R}$

$$\frac{1}{b^{1+\gamma}} \int_1^b g_\gamma(z-w)\,\mathrm{d}w \;\lesssim\; g_\gamma(z)\,. \tag{19}$$

We now consider two cases.
Case $|z| \leq 1$: It holds

$$\frac{|z|+1}{2} \;\leq\; 1 \;<\; b \qquad \text{and hence} \qquad \frac{1}{b^{1+\gamma}} \;\lesssim\; \frac{1}{(1+|z|)^{1+\gamma}} \;=\; g_\gamma(z)\,.$$

Since

$$\int_1^b g_\gamma(z-w)\,\mathrm{d}w \;\leq\; \int_{\mathbb{R}} g_\gamma(w)\,\mathrm{d}w \;<\; \infty \qquad \text{for all } z \in \mathbb{R}\,, \tag{20}$$

we get (19).
Case $|z| > 1$: We will use, without further mention, the inequalities $1/|z| \leq 2/(1+|z|) \leq 2/|z|$, and consider three subcases.
Subcase $b \geq |z|/2$: It holds

$$\frac{1}{b^{1+\gamma}} \;\lesssim\; \frac{1}{(1+|z|)^{1+\gamma}}\,,$$

and by (20), this implies (19).
Subcase $z/2 > b \geq 1$: In this case it holds $z - 1 > z - b > 0$, and

$$\begin{aligned}\int_1^b g_\gamma(z-w)\,\mathrm{d}w \;\leq\; \int_{z-b}^{z-1} \frac{1}{w^{1+\gamma}}\,\mathrm{d}w \;&=\; \frac{1}{\gamma}\left(\frac{1}{(z-b)^\gamma} \;-\; \frac{1}{(z-1)^\gamma}\right) \\ &=\; \frac{1}{\gamma(z-b)^\gamma}\left(1 \;-\; \left(\frac{z-b}{z-1}\right)^\gamma\right)\,.\end{aligned}$$

By the fundamental theorem of calculus, we obtain

$$\begin{aligned}1 \;-\; \left(\frac{z-b}{z-1}\right)^\gamma \;=\; -\int_1^b \frac{\mathrm{d}}{\mathrm{d}v}\left(\left(\frac{z-v}{z-1}\right)^\gamma\right)\mathrm{d}v \;&=\; \frac{\gamma}{(z-1)^\gamma} \cdot \int_1^b (z-v)^{\gamma-1}\,\mathrm{d}v \\ &\leq\; \frac{\gamma}{(z-1)^\gamma} \cdot b(z-b)^{\gamma-1}\end{aligned}$$

since $\gamma \in (0,1)$. Thus

$$\int_1^b g_\gamma(z-w)\,\mathrm{d}w \;\le\; \frac{b}{(z-1)^\gamma(z-b)} \;\le\; \frac{b}{(z-z/2)^\gamma(z-z/2)} \;\lesssim\; \frac{b}{z^{\gamma+1}} \quad \text{by } z/2 > b \ge 1\,.$$

It follows (19), that is,

$$\frac{1}{b^{1+\gamma}}\int_1^b g_\gamma(z-w)\,\mathrm{d}w \;\lesssim\; \frac{1}{b^\gamma}\frac{1}{(1+|z|)^{1+\gamma}} \;\lesssim\; \frac{1}{(1+|z|)^{1+\gamma}} \;=\; g_\gamma(z)\,. \tag{21}$$

Subcase $-z/2 > b \ge 1$: Using the symmetry of g_γ, we write

$$\begin{aligned}\int_1^b g_\gamma(z-w)\,\mathrm{d}w \;&=\; \int_1^b g_\gamma(w-z)\,\mathrm{d}w \;\le\; \int_{1-z}^{b-z} \frac{1}{w^{1+\gamma}}\,\mathrm{d}w\\ &=\; \frac{1}{\gamma}\left(\frac{1}{(1-z)^\gamma} \;-\; \frac{1}{(b-z)^\gamma}\right)\\ &=\; \frac{1}{\gamma(b-z)^\gamma}\left(\left(\frac{b-z}{1-z}\right)^\gamma \;-\; 1\right).\end{aligned}$$

Applying the fundamental theorem of calculus again and using $\gamma \in (0,1)$, we obtain

$$\begin{aligned}\left(\frac{b-z}{1-z}\right)^\gamma \;-\; 1 \;=\; \int_1^b \frac{\mathrm{d}}{\mathrm{d}v}\left(\left(\frac{v-z}{1-z}\right)^\gamma\right)\mathrm{d}v \;&=\; \frac{\gamma}{(1-z)^\gamma}\cdot\int_1^b (v-z)^{\gamma-1}\,\mathrm{d}v\\ &\le\; \frac{\gamma}{1-z}\cdot b\,.\end{aligned}$$

Thus,

$$\begin{aligned}&\int_1^b g_\gamma(z-w)\,\mathrm{d}w\\ &\le\; \frac{b}{(b-z)^\gamma(1-z)} \;\le\; \frac{b}{(1-z)^\gamma(1-z)} \;\lesssim\; \frac{b}{|z|^{\gamma+1}} \quad \text{by } -z/2 > b \ge 1\,.\end{aligned}$$

As in (21), this yields (19). This completes the proof. □

A.3 Swap integration and differentiation

The next lemma asserts that all the notions constructed from G_n and used in this work are well-defined. It justifies interchanging differentiation and integration, which simplifies the terms we are going to work with.

Lemma A.4 (Swap Integration and Differentiation)**.** *Let Assumptions 1, 2 and 3 hold with $\gamma_G < d/(1-d)$, and define $G_{k,n}(y) := \mathbf{E}[G_n(X_{0,k}+y)] = \int_{\mathbb{R}} G_n(x) f_k(x-y)\mathrm{d}x$. Then the following statements hold:*

(i) $\mathbf{E}|G_n(X_{0,k}+y)|^r < \infty$ *for all $k \in \mathbb{N}\cup\{\infty\}$, $n \in \mathbb{N}$, $y \in \mathbb{R}$, and for all $r \in (0,\alpha)$.*

(ii) With the notation of Lemma A.2, for all $k > k_0$ or $k = \infty$, it holds that the function $G_{k,n}$ is continuously differentiable with

$$G'_{k,n}(y) \;:=\; \frac{\mathrm{d}}{\mathrm{d}y}G_{k,n}(y) \;=\; -\int_{\mathbb{R}} G_n(x) f'_k(x-y)\,\mathrm{d}x \;<\; \infty \qquad \text{for all } y \in \mathbb{R}\,.$$

Proof. **Proof of Part (i):** By Assumption 1 there exists $\delta > 0$ such that $(\alpha-\delta)(1-d) > 1$. Let γ_G as in Assumption 2. Since $r < \alpha$ and

$$\gamma_G < \frac{d}{1-d} \ = \ \frac{1}{1-d} \ - \ 1 \ < \ \alpha - 1 \leq 1 \tag{22}$$

by assumption, we can choose $\delta > 0$ small enough such that $\alpha - \delta \geq 1$ and $r\gamma_G < \alpha - \delta$. For $y \in \mathbb{R}$ it follows

$$\begin{aligned}
\mathbf{E}|G_n(X_{0,k} + y)|^r &\lesssim \mathbf{E}[(1 + |y + X_{0,k}|)^{r\gamma_G}] \\
&\lesssim \mathbf{E}[(1 + |y + X_{0,k}|)^{\alpha-\delta}] \\
&\lesssim 1 \ + \ |y|^{\alpha-\delta} \ + \ \mathbf{E}|X_{0,k}|^{\alpha-\delta} \\
&\lesssim 1 \ + \ |y|^{\alpha-\delta} \ + \ \mathbf{E}|\varepsilon|^{\alpha-\delta} \sum_{j=0}^{\infty} |a_j|^{\alpha-\delta} \\
&\lesssim 1 \ + \ |y|^{\alpha-\delta} \ + \ \sum_{j=1}^{\infty} j^{-(\alpha-\delta)(1-d)} \ < \ \infty \,.
\end{aligned}$$

The third inequality is due to the convexity of $x \mapsto |x|^{\alpha-\delta}$, and the fourth inequality is due to the last part of Lemma A.1(i). This concludes the proof of Part (i).
Proof of Part (ii): Fix $y \in \mathbb{R}$ and assume $y \neq y'$. By Part (i), it follows $\mathbf{E}|G_n(X_{0,k} + y)| < \infty$ for all $y \in \mathbb{R}$, so that we can use linearity of the expectation to get

$$\frac{\mathbf{E}[G_n(X_{0,k} + y)] - \mathbf{E}[G_n(X_{0,k} + y')]}{y - y'} \ = \ \int_{\mathbb{R}} G_n(x) \frac{f_k(x-y) - f_k(x-y')}{y - y'} \, \mathrm{d}x \,. \tag{23}$$

If we may interchange the limit $y' \to y$ with the integral, then

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}y}\mathbf{E}[G_n(X_{0,k} + y)] \ = \ \lim_{y' \to y} \frac{\mathbf{E}[G_n(X_{0,k} + y)] - \mathbf{E}[G_n(X_{0,k} + y')]}{y - y'} \\
&= \ \int_{\mathbb{R}} G_n(x) \lim_{y' \to y} \frac{f_k(x-y) - f_k(x-y')}{y - y'} \, \mathrm{d}x \ = \ - \int_{\mathbb{R}} G_n(x) f_k'(x-y) \, \mathrm{d}x \,,
\end{aligned}$$

as required. We seek to use dominated convergence to interchange these limits, and to this end we show that there is a bivariate function g such that $g(\cdot, y) \in L^1(\mathbb{R})$ and

$$\left| G_n(x) \frac{f_k(x-y) - f_k(x-y')}{y - y'} \right| \ \lesssim \ g(x,y) \qquad \text{uniformly in } y' \text{ with } |y' - y| \leq 1. \tag{24}$$

First we apply the fundamental theorem of calculus and the triangle inequality to get

$$\left| G_n(x) \frac{f_k(x-y) - f_k(x-y')}{y - y'} \right| \ \leq \ \frac{G_n(x)}{|y - y'|} \int_{(y \wedge y')}^{(y \vee y')} \left| f_k'(x-s) \right| \, \mathrm{d}s \,.$$

Since $\gamma_G < \alpha - 1$ by (22), there exists γ with $\gamma_G < \gamma < \alpha - 1 \leq 1$ such that from Lemma A.2 with $1 + \gamma < (\alpha \wedge 2)$, and the first part of Lemma A.3 we get

$$\begin{aligned}
\int_{(y \wedge y')}^{(y \vee y')} \left| f_k'(x-s) \right| \, \mathrm{d}s &\lesssim g_\gamma(x) \int_{(y \wedge y')}^{(y \vee y')} (1 \vee |s|)^{1+\gamma} \, \mathrm{d}s \\
&\leq g_\gamma(x) |y - y'| (1 \vee |y| \vee |y'|)^{1+\gamma} \\
&\lesssim g_\gamma(x) |y - y'| (1 + |y|)^2
\end{aligned}$$

since $|y-y'| \le 1$. Then, with Assumption 2

$$\frac{G_n(x)}{|y-y'|}\int_{(y\wedge y')}^{(y\vee y')}\left|f_k'(x-s)\right|\,\mathrm{d}s$$
$$\lesssim\; G_n(x)g_\gamma(x)(1+|y|)^2 \;\lesssim\; (1+|x|)^{\gamma_G}g_\gamma(x)(1+|y|)^2 \;\lesssim\; g_{\gamma-\gamma_G}(x)(1+|y|)^2\,.$$

With $\gamma-\gamma_G>0$ and therefore

$$\int_{\mathbb{R}} g_{\gamma-\gamma_G}(x)\,\mathrm{d}x \;<\; \infty\,, \tag{25}$$

it follows (24) with $g(x,y)=g_{\gamma-\gamma_G}(x)(1+|y|)^2$. To complete the proof, we show that $G'_{k,n}$ is continuous. To this end, let $y\in\mathbb{R}$ and $(y_m)_{m\in\mathbb{N}}$ a bounded sequence with $y_m\to y$ for $m\to\infty$. We want to show

$$\begin{aligned}\lim_{m\to\infty}G'_{k,n}(y_m) \;&=\; \lim_{m\to\infty}-\int_{\mathbb{R}}G_n(x)f_k'(x-y_m)\,\mathrm{d}x\\ &=\; -\int_{\mathbb{R}}G_n(x)\lim_{m\to\infty}f_k'(x-y_m)\,\mathrm{d}x\\ &=\; -\int_{\mathbb{R}}G_n(x)f_k'(x-y)\,\mathrm{d}x\\ &=\; G'_{k,n}(y)\,.\end{aligned}$$

The third equality is due to the continuity of f_k' (see Lemma A.2). To show the second equality we want to apply dominated convergence. To this end, note that by (14) in Lemma A.2 and then Lemma A.3(i), with the choice of γ as above,

$$\left|G_n(x)f_k'(x-y_m)\right| \;\lesssim\; (1+|x|)^{\gamma_G}g_\gamma(x)(1\vee|y_m|)^{1+\gamma} \;\lesssim\; g_{\gamma-\gamma_G}(x)\,.$$

Together with (25) we may apply dominated convergence to finish the proof. □

Lemma A.4 tells us that

$$G_{k,n}(y) \;:=\; \mathbf{E}[G_n(X_{0,k}+y)] \qquad \text{and} \qquad G'_{k,n}(y) \;:=\; \frac{\mathrm{d}}{\mathrm{d}y}\mathbf{E}[G_n(X_{0,k}+y)]$$

are well-defined for $k>k_0$. The following lemma establishes a connection between consecutive values of $G_{k,n}$ and $G'_{k,n}$.

Lemma A.5. *Let Assumptions 1, 2 and 3 hold true. It holds for $j\in\{0,1\}$ and $k>k_0$*

$$G^{(j)}_{k+1,n}(y) \;=\; \mathbf{E}[G^{(j)}_{k,n}(a_{k+1}\varepsilon_{-k-1}+y)] \qquad \textit{and} \qquad G^{(j)}_{\infty,n}(y) \;=\; \mathbf{E}[G^{(j)}_{k,n}(\widetilde{X}_{0,k}+y)]\,.$$

Proof. Let $(f*g)(y)$ denote the convolution defined by $\int_{\mathbb{R}}f(y-x)g(x)\,\mathrm{d}x$. Since ε is symmetric, the density f_ε is even, and its derivative f_ε' is odd. Consequently, f_k is even and f_k' is odd for all $k\in\mathbb{N}\cup\{\infty\}$. Thus, for $k>k_0$,

$$G_{k+1,n}(y) \;=\; \int_{\mathbb{R}}G_n(x)f_{k+1}(x-y)\,\mathrm{d}x \;=\; (G_n*f_{k+1})(y)\,, \tag{26}$$

and as shown in Lemma A.4

$$G'_{k+1,n}(y) \;=\; -\int_{\mathbb{R}}G_n(x)f'_{k+1}(x-y)\,\mathrm{d}x \;=\; (G_n*f'_{k+1})(y)\,. \tag{27}$$

Observe that $f_{k+1} = f_k * f_{a_{k+1}\varepsilon_{-k-1}}$, by the independence of the innovations, and then, by dominated convergence,

$$f'_{k+1} = (f_k * f_{a_{k+1}\varepsilon_{-k-1}})' \; = \; (f'_k * f_{a_{k+1}\varepsilon_{-k-1}}) \tag{28}$$

since $f_{a_{k+1}\varepsilon_{-k-1}} \in L^1(\mathbb{R})$, and $f_k \in C^1(\mathbb{R})$ with bounded derivative for all $k > k_0$ by (14) in Lemma A.2. Thus for $j \in \{0, 1\}$,

$$\begin{aligned} G^{(j)}_{k+1,n}(y) &= (G_n * f^{(j)}_{k+1})(y) \\ &= ((G_n * f^{(j)}_k) * f_{a_{k+1}\varepsilon_{-k-1}})(y) \\ &= (G^{(j)}_{k,n} * f_{a_{k+1}\varepsilon_{-k-1}})(y) \\ &= \int_{\mathbb{R}} G^{(j)}_{k,n}(x) f_{a_{k+1}\varepsilon_{-k-1}}(x-y)\,\mathrm{d}x \\ &= \mathbf{E}[G^{(j)}_{k,n}(a_k\varepsilon_{-k-1} + y)]\,, \end{aligned}$$

where the first and third equalities are due to (26) and (27), and the second equality is due to (28) (and associativity of the convolution operator). Observe, finally, that $f_\infty = f_k * f_{\widetilde{X}_{0,k}}$, and that a similar argument can be written for $f_{a_{k+1}\varepsilon_{-k-1}}$ replaced by $f_{\widetilde{X}_{0,k}}$. Therefore the statement holds also for $k = \infty$. This completes the proof. □

B Proof of Theorem 3.3

The proof of Theorem 3.3 is long. It draws on ideas from [26], where the authors conducted their analysis without the benefit of finite second moments—a difficulty we must also address. Throughout this section, let $n \in \mathbb{N}$ unless stated otherwise. To track the deviation of the PoT statistic from its approximation, we define a centered process that isolates the remainder term after linear approximation. This term will be central in applying martingale arguments. In the sequel, for $n \in \mathbb{N}$, define

$$\mathcal{U}_n(X_t) \;:=\; G_n(X_t) \;-\; \mathbf{E}[G_n(X_t)] \;-\; G'_{\infty,n}(0)X_t \qquad \text{for } t \in \mathbb{Z}\,.$$

Note that under Assumptions 1, 2 and 3, the term $G'_{\infty,n}(0)$ is indeed well-defined by Lemma A.4. Hence, $\mathcal{U}_n(X_t)$ itself will be well-defined and integrable once we verify that $G_n(X_0)$ and X_0 are integrable. For $k \in \mathbb{Z}$, let $\mathcal{F}_k = \sigma(\varepsilon_k, \varepsilon_{k-1}, \ldots)$ denote the past σ-algebra generated by the sequence (ε_t) up to index k. For $Y \in L^1(\mathbf{P})$, we define the projection

$$P_k Y \;:=\; \mathbf{E}\,[Y \,|\, \mathcal{F}_k] \;-\; \mathbf{E}\,[Y \,|\, \mathcal{F}_{k-1}] \qquad \text{for all } k \in \mathbb{Z}\,.$$

We proceed with a sequence of lemmas that prepare the proof of the final result. The next lemma establishes integrability properties of the centered process $\mathcal{U}_n(X_t)$—first in its unprojected form, then after projection by P_k. This integrability is needed to apply the lemma on martingale difference arrays (Lemma A.1) later in the proof. Note that Part (ii) also provides an upper bound that contributes directly to the final result.

Lemma B.1. *Under Assumptions 1, 2 and 3, for $r \in [1, \alpha)$, the following statements hold:*

(i) $\mathcal{U}_n(X_0) \in L^r(\mathbf{P})$.

(ii) $\|P_k\mathcal{U}_n(X_0)\|_{L^r(\mathbf{P})} \lesssim \|G_n(X_0)\|_{L^r(\mathbf{P})} \vee |G'_{\infty,n}(0)| < \infty$ *for all $k \in \mathbb{Z}$.*

Proof. **Proof of Part (i):** Let $r \in [1, \alpha)$. By Assumption 2, the inequalities

$$\gamma_G \;<\; \frac{d}{1-d} \;<\; 1 \;<\; \frac{1}{1-d}\,,$$

which hold because of $d < 1/2$ (see Assumption 1), and the convexity of $x \mapsto |x|^r$, we obtain

$$\begin{aligned} \mathbf{E}|G_n(X_0)|^r \;&\lesssim\; \mathbf{E}(1+|X_0|)^{r\gamma_G} \;\leq\; \mathbf{E}(1+|X_0|)^r \;\leq\; 2^{r-1}\,(1+\mathbf{E}|X_0|^r) \\ &\leq\; 2\,(1+\mathbf{E}|X_0|^r)\;. \end{aligned} \tag{29}$$

To further bound $\mathbf{E}|X_0|^r$, we choose $\widetilde{r} \in (r \vee (1/(1-d)), \alpha)$, such that, by Hölder's inequality, $\mathbf{E}|X_0|^r \;\leq\; (\mathbf{E}|X_0|^{\widetilde{r}})^{r/\widetilde{r}}$, and continue to bound $\mathbf{E}|X_0|^{\widetilde{r}}$. To this end, we apply Lemma A.1 to the array $(Y_{m,i}) = (a_i\varepsilon_{-i} : i = 0, \ldots, m\,, m \in \mathbb{N})$ and the sequence $(Y_i) = (a_i\varepsilon_{-i})_{i\geq 0}$. By Assumption 3, both $(Y_{m,i})$ and $(Y_i) \subset L^{\widetilde{r}}(\mathbf{P})$. Since the innovations are independent and centered, Condition (10) is satisfied. Moreover, Condition (12) is clear, so by (13) in Lemma A.1 and Assumption 1, we get

$$\mathbf{E}|X_0|^{\widetilde{r}} \;=\; \mathbf{E}\left|\sum_{i=0}^{\infty} Y_i\right|^{\widetilde{r}} \;\lesssim\; \sum_{i=0}^{\infty}\mathbf{E}\,|Y_i|^{\widetilde{r}} \;\lesssim\; \sum_{i=1}^{\infty} i^{-(1-d)\widetilde{r}} \;<\; \infty\,, \tag{30}$$

since we chose $\widetilde{r} > 1/(1-d)$. Using (29), this shows that $G_n(X_0) \in L^r(\mathbf{P})$. Since $X_0 \in L^r(\mathbf{P})$ by (30), the proof of Part (i) is complete.

Proof of Part (ii): From Part (i), we know that $P_k\,\mathcal{U}_n(X_0)$ is well-defined. For fixed $j, k \in \mathbb{Z}$, observe that

$$P_{-k}\varepsilon_{-j} \;=\; \begin{cases} 0 & \text{if } j \neq k\,, \\ \varepsilon_{-k} & \text{if } j = k\,, \end{cases}$$

so, by dominated convergence, it follows that

$$P_k\,X_0 \;=\; \sum_{i=0}^{\infty} a_i P_k\,\varepsilon_{-i} \;=\; \begin{cases} 0 & \text{if } k > 0\,, \\ a_{-k}\varepsilon_k & \text{if } k \leq 0\,. \end{cases}$$

Thus,

$$\begin{aligned} P_k\,\mathcal{U}_n(X_0) \;&=\; P_k\,(G_n(X_0) \;-\; \mathbf{E}[G_n(X_0)] \;-\; G'_{\infty,n}(0)X_0) \\ &=\; \mathbf{E}[G_n(X_0)\,|\,\mathcal{F}_k] \;-\; \mathbf{E}[G_n(X_0)\,|\,\mathcal{F}_{k-1}] \;-\; G'_{\infty,n}(0)\cdot a_{-k}\cdot\varepsilon_k\cdot\mathbb{1}\{k\leq 0\}\,. \end{aligned} \tag{31}$$

Applying Jensen's inequality with $r \geq 1$, and using the properties of conditional expectation, we obtain

$$\mathbf{E}|\mathbf{E}[G_n(X_0)\,|\,\mathcal{F}_k]|^r \;\leq\; \mathbf{E}|G_n(X_0)|^r \;<\; \infty \qquad \text{as shown in Part (i)}\,.$$

Combining Equation (31) with the fact that $\varepsilon \in L^r(\mathbf{P})$ for $r < \alpha$ completes the proof of Lemma B.1. □

The next lemma provides an intermediate upper bound on the error term

$$\sum_{t=1}^{n}\big(G_n(X_t) \;-\; \mathbf{E}[G_n(X_0)] \;-\; G'_{\infty,n}(0)X_t\big) = \sum_{t=1}^{n}\mathcal{U}_n(X_t)$$

from Theorem 3.3(i). The bound is obtained by first decomposing the error term via projections, and then applying the martingale difference inequality (Lemma A.1).

Lemma B.2. *Under the assumptions of Lemma B.1, it holds that*

$$\mathbf{E}\left|\sum_{t=1}^{n}\mathcal{U}_n(X_t)\right|^r \lesssim \sum_{k=-\infty}^{n}\left(\sum_{t=k\vee 1}^{n}\|P_{k-t}\mathcal{U}_n(X_0)\|_{L^r(\mathbf{P})}\right)^r .$$

At this stage it is not obvious that the right-hand side is finite, and thus that this upper bound is informative; the next step of the proof, in Lemma B.3, turns this into an informative upper bound.

Proof. Since the process (X_t) is (strictly) stationary, we may define

$$T_n(G_n) := \sum_{t=1}^{n}\mathcal{U}_n(X_t) = \sum_{t=1}^{n}G_n(X_t) - \mathbf{E}[G_n(X_0)] - G'_{\infty,n}(0)X_t .$$

Because X_t is $\mathcal{F}_t$-measurable and the sequence $(\mathcal{F}_k)$ is increasing, $T_n(G_n)$ is $\mathcal{F}_n$-measurable. Then

$$P_k\,T_n(G_n) = 0 \quad \text{for } k>n \quad \text{and} \quad P_k\,\mathcal{U}_n(X_t) = 0 \quad \text{for } k>t .$$

We can therefore write, for any $\ell<n$,

$$T_n(G_n) - \mathbf{E}[T_n(G_n)\mid\mathcal{F}_{\ell-1}] = \sum_{k=\ell}^{n}P_k\,T_n(G_n).$$

Since $\cap_{k\in\mathbb{Z}}\mathcal{F}_k$ is the trivial $\sigma-$algebra, $\mathbf{E}[T_n(G_n)\mid\mathcal{F}_{\ell-1}]\to\mathbf{E}[T_n(G_n)]=0$ as $\ell\to-\infty$. Hence, we obtain

$$T_n(G_n) = \sum_{k=-\infty}^{n}P_k\,T_n(G_n)\,, \tag{32}$$

and then

$$P_k\,T_n(G_n) = \sum_{t=1}^{n}P_k\,\mathcal{U}_n(X_t) = \sum_{t=k\vee 1}^{n}P_k\,\mathcal{U}_n(X_t). \tag{33}$$

We aim to apply Lemma A.1 to Equation (32). To this end, define the array

$$(Y_{m,i}) = (P_{i+n-m}\,T_n(G_n):\ i=1,\dots,m\,,m\in\mathbb{N})\,,$$

so that

$$\sum_{i=1}^{m}Y_{m,i} = \sum_{k=n-m+1}^{n}P_k\,T_n(G_n)\,,$$

and define the associated sequence $(Y_i)_{i\geq 1}$ by $Y_i=P_{n-i+1}\,T_n(G_n)$. By Lemma B.1(ii) and Identity (33) it holds that $(Y_{m,i}),(Y_i)\subset L^r(\mathbf{P})$. Note that

$$\sum_{i=1}^{\infty}Y_i = \sum_{k=-\infty}^{n}P_k\,T_n(G_n) = \lim_{m\to\infty}\sum_{k=n-m+1}^{n}P_k\,T_n(G_n) = \lim_{m\to\infty}\sum_{i=1}^{m}Y_{m,i}\,,$$

and similarly $\sum_{i=1}^{\infty} \mathbf{E}|Y_i|^r = \lim_{m\to\infty} \sum_{i=1}^{m} \mathbf{E}|Y_{m,i}|^r$. To apply Lemma A.1, it remains to verify Condition (10). To this end, observe that Equation (32) represents an $\mathcal{F}_\ell$-martingale difference decomposition in the sense that, for $\ell \in \mathbb{Z}$,

$$\mathbf{E}[P_{\ell+1}\, T_n(G_n) \mid \mathcal{F}_\ell] \;=\; \mathbf{E}[\mathbf{E}[T_n(G_n) \mid \mathcal{F}_{\ell+1}] \mid \mathcal{F}_\ell] \;-\; \mathbf{E}[T_n(G_n) \mid \mathcal{F}_\ell] \;=\; 0\,.$$

Consequently

$$\mathbf{E}[Y_{m,\ell+1} \mid \mathcal{F}_{n-m+\ell}] \;=\; 0 \qquad \text{for all } 1 \le \ell \le m-1\,, m \in \mathbb{N}\,. \tag{34}$$

Let $\mathcal{V}_\ell^m$ denote the σ-algebra generated by

$$\sum_{i=1}^{\ell} Y_{m,i} \;=\; \sum_{i=1}^{\ell} P_{i+n-m}\, T_n(G_n) \;=\; \sum_{k=n-m+1}^{n-m+\ell} P_k\, T_n(G_n)\,.$$

Since the filtration $(\mathcal{F}_k)$ is increasing, all summands are $\mathcal{F}_{n-m+\ell}$-measurable. Therefore, $\mathcal{V}_\ell^m \subset \mathcal{F}_{n-m+\ell}$. Combined with (34) this verifies Condition (10). We can now apply Inequality (13) to obtain

$$\mathbf{E}\left|\sum_{k=-\infty}^{n} P_k T_n(G_n)\right|^r \;=\; \mathbf{E}\left|\sum_{i=1}^{\infty} Y_i\right|^r \;\lesssim\; \sum_{i=1}^{\infty} \mathbf{E}\,|Y_i|^r \;=\; \sum_{k=-\infty}^{n} \mathbf{E}\,|P_k T_n(G_n)|^r\,.$$

With Equation (32) and Equation (33), it follows that

$$\begin{aligned}
\mathbf{E}\left[|T_n(G_n)|^r\right] \;&\lesssim\; \sum_{k\le n} \mathbf{E}\left|\sum_{t=k\vee 1}^{n} P_k\, \mathcal{U}_n(X_t)\right|^r \\
&\le\; \sum_{k\le n} \left(\sum_{t=k\vee 1}^{n} \|P_{k-t}\, \mathcal{U}_n(X_0)\|_{L^r(\mathbf{P})}\right)^r.
\end{aligned} \tag{35}$$

The final inequality follows from the triangle inequality for the $L^r(\mathbf{P})$ norm, using that $\|P_k\, \mathcal{U}_n(X_t)\|_{L^r(\mathbf{P})} = \|P_{k-t}\, \mathcal{U}_n(X_0)\|_{L^r(\mathbf{P})}$ due to stationarity. This completes the proof. $\square$

The most delicate step now lies ahead: bounding each projection term individually. The first part of the next lemma sets the stage by identifying a suitable decomposition of the error. The core of the argument is in the second part, where we introduce auxiliary parameters γ and r to mitigate the effect of infinite second moments.

Lemma B.3 (Error Decomposition and Control)**.** *Let Assumptions 1, 2 and 3 hold, let k_0 be as defined by Lemma A.2, and let $k > k_0$.*

(i) We can write $P_{-k}\, \mathcal{U}_n(X_0) = R_1 + R_2 + R_3$, where

$$\begin{aligned}
R_1 \;&:=\; G_{k-1,n}(X_0 - X_{0,k-1}) \;-\; \int G_{k-1,n}(X_0 - X_{0,k} + a_k z)\, \mathrm{d}\mathbf{P}_\varepsilon(z) \\
&\qquad -\; a_k \cdot \varepsilon_{-k} \cdot G'_{k-1,n}(X_0 - X_{0,k})\,, \\
R_2 \;&:=\; a_k \cdot \varepsilon_{-k} \left(G'_{\infty,n}(X_0 - X_{0,k}) \;-\; G'_{\infty,n}(0)\right), \\
R_3 \;&:=\; a_k \cdot \varepsilon_{-k} \left(G'_{k-1,n}(X_0 - X_{0,k}) \;-\; G'_{\infty,n}(X_0 - X_{0,k})\right).
\end{aligned}$$

(ii) Let $\gamma \in (0,1)$ and $r \in [1,2]$ such that

$$\gamma_G \;<\; \gamma \;<\; \min\left\{\frac{d}{1-d}\,,1-\frac{1}{r(1-d)}\,,\frac{\alpha}{r}-1\right\} \qquad \textit{and} \qquad \frac{1}{1-d} \;<\; r \;<\; \alpha\,.$$

For each $j \in \{1,2,3\}$, we have $\|R_j\|_{L^r(\mathbf{P})} \;\lesssim\; k^{-(1-d)(1+\gamma)} \cdot u_n^{-\gamma+\gamma_G}$. Therefore

$$\|P_{-k}\,\mathcal{U}_n(X_0)\|_{L^r(\mathbf{P})} \;\lesssim\; k^{-(1-d)(1+\gamma)} \cdot u_n^{-\gamma+\gamma_G}\,. \tag{36}$$

Below and throughout, the symbol $\asymp$ means asymptotic equivalence of sequences and functions up to a constant $C \neq 0$ that may differ from 1.

B.1 Proof of Lemma B.3(i)

By definition

$$\begin{aligned} &R_1 \;+\; R_2 \;+\; R_3 \\ &\quad=\; G_{k-1,n}(X_0 - X_{0,k-1}) \;-\; \int G_{k-1,n}(X_0 - X_{0,k} + a_k z)\,\mathrm{d}\mathbf{P}_\varepsilon(z) \;-\; a_k \cdot \varepsilon_{-k} \cdot G'_{\infty,n}(0)\,. \end{aligned}$$

On the other hand, using (31), we have

$$P_{-k}\,\mathcal{U}_n(X_0) \;=\; \mathbf{E}[G_n(X_0)\,|\,\mathcal{F}_{-k}] \;-\; \mathbf{E}[G_n(X_0)\,|\,\mathcal{F}_{-(k+1)}] \;-\; a_k \cdot \varepsilon_{-k} \cdot G'_{\infty,n}(0)\,.$$

Note that $X_{0,k-1} = \sum_{j=0}^{k-1} a_j \varepsilon_{-j}$ is independent of $\mathcal{F}_{-k}$, and $X_0 - X_{0,k-1} = \sum_{j=k}^{\infty} a_j \varepsilon_{-j}$ is $\mathcal{F}_{-k}$-measurable. Therefore,

$$\begin{aligned} \mathbf{E}[G_n(X_0)\,|\,\mathcal{F}_{-k}] \;&=\; \mathbf{E}[G_n(X_{0,k-1} + (X_0 - X_{0,k-1}))\,|\,\mathcal{F}_{-k}] \\ &=\; \int G_n(z + (X_0 - X_{0,k-1}))\,\mathrm{d}\mathbf{P}_{X_{0,k-1}}(z) \\ &=\; G_{k-1,n}(X_0 - X_{0,k-1})\,. \end{aligned}$$

Similarly, by Lemma A.5,

$$\mathbf{E}[G_n(X_0)\,|\,\mathcal{F}_{-(k+1)}] \;=\; G_{k,n}(X_0 - X_{0,k}) \;=\; \int G_{k-1,n}(X_0 - X_{0,k} + a_k z)\,\mathrm{d}\mathbf{P}_\varepsilon(z)\,.$$

Consequently, combining the expressions gives

$$\begin{aligned} &P_{-k}\,\mathcal{U}_n(X_0) \\ &\quad=\; G_{k-1,n}(X_0 - X_{0,k-1}) \;-\; \int G_{k-1,n}(X_0 - X_{0,k} + a_k z)\,\mathrm{d}\mathbf{P}_\varepsilon(z) \;-\; a_k \cdot \varepsilon_{-k} \cdot G'_{\infty,n}(0) \\ &\quad=\; R_1 \;+\; R_2 \;+\; R_3\,, \end{aligned}$$

as claimed.

B.2 Proof of Lemma B.3(ii)

Given the bounds on the R_j, the Bound (36) follows directly from Part (i). We now turn to the proof of the announced bounds on the R_j. Fix $k > k_0$ for the remainder of the proof. Note that by Assumption 2, and for any γ with $\gamma_G < \gamma < d/(1-d)$, we have

$$\int_{\mathbb{R}} G_n(x) g_\gamma(x)\,\mathrm{d}x \;\lesssim\; \int_{u_n}^{\infty} (1+|x|)^{-(1+\gamma-\gamma_G)}\,\mathrm{d}x \;\asymp\; u_n^{-\gamma+\gamma_G}\,. \tag{37}$$

Moreover, the next inequalities ensure that the conditions of Lemma A.2 are met:

$$1 \;+\; \gamma \;<\; 1 \;+\; \frac{d}{1-d} \;=\; \frac{1}{1-d} \;<\; r \;<\; \alpha\,.$$

B.2.1 Analysis of R_1

Write R_1 as

$$\begin{aligned} R_1 \;&=\; G_{k-1,n}(X_0 - X_{0,k-1}) \;-\; \int_{\mathbb{R}} G_{k-1,n}(X_0 - X_{0,k} + a_k z)\,\mathrm{d}\mathbf{P}_\varepsilon(z) \\ &\quad -\; a_k \cdot \varepsilon_{-k} \cdot G'_{k-1,n}(X_0 - X_{0,k}) \\ &=\; \int_{\mathbb{R}} [G_{k-1,n}(X_0 - X_{0,k} + a_k\varepsilon_{-k}) \;-\; G_{k-1,n}(X_0 - X_{0,k} + a_k z) \\ &\quad -\; a_k \cdot \varepsilon_{-k} \cdot G'_{k-1,n}(X_0 - X_{0,k})]\,\mathrm{d}\mathbf{P}_\varepsilon(z) \\ &=\; \int_{\mathbb{R}} \Bigg[\left(\int_{a_k z}^{a_k\varepsilon_{-k}} G'_{k-1,n}(X_0 - X_{0,k} + s)\,\mathrm{d}s \right) \\ &\quad -\; a_k \cdot \varepsilon_{-k} \cdot G'_{k-1,n}(X_0 - X_{0,k}) \;+\; a_k \cdot z \cdot G'_{k-1,n}(X_0 - X_{0,k}) \Bigg]\,\mathrm{d}\mathbf{P}_\varepsilon(z) \\ &=\; \int_{\mathbb{R}} \left(\int_{a_k z}^{a_k\varepsilon_{-k}} [G'_{k-1,n}(X_0 - X_{0,k} + s) \;-\; G'_{k-1,n}(X_0 - X_{0,k})]\,\mathrm{d}s \right) \mathrm{d}\mathbf{P}_\varepsilon(z)\,. \end{aligned} \tag{38}$$

Note that the third equality follows by applying the fundamental theorem of calculus to $G_{k-1,n}$, which is justified by Lemma A.4.(ii). Note also that adding the term

$$a_k \cdot z \cdot G'_{k-1,n}(X_0 - X_{0,k})$$

inside the integral is valid because ε is centered. To proceed, we consider three mutually exclusive cases:

- $1 \;\geq\; (|a_k\varepsilon_{-k}| \vee |a_k z|)$
- $|a_k\varepsilon_{-k}| \;>\; (|a_k z| \vee 1)$
- $|a_k z| \;>\; (|a_k\varepsilon_{-k}| \vee 1)$

These three cases exhaust all possibilities and will be treated separately in the following sections.

Case $1 \geq (|a_k\varepsilon_{-k}| \vee |a_k z|)$ We bound R_1 as follows:

$$|R_1| \;\leq\; \int_{\mathbb{R}} \left| \int_{a_k z}^{a_k\varepsilon_{-k}} \left|G'_{k-1,n}(X_0 - X_{0,k} + s) \;-\; G'_{k-1,n}(X_0 - X_{0,k})\right|\,\mathrm{d}s \right| \mathrm{d}\mathbf{P}_\varepsilon(z)\,. \tag{39}$$

Let s lie in the interval linking $a_k z$ to $a_k\varepsilon_{-k}$, so that $|s| \leq 1$. Let $\gamma \in (\gamma_G, 1)$ be as in Lemma A.2. We can use (15) there and (17) in Lemma A.3 to get for all $x \in \mathbb{R}$

$$\begin{aligned} &\left|f'_{k-1}(x - (X_0 - X_{0,k} + s)) \;-\; f'_{k-1}(x - (X_0 - X_{0,k}))\right| \\ &\quad\lesssim\; |s| \cdot g_\gamma(x - [X_0 - X_{0,k}]) \;\lesssim\; |s|^\gamma \cdot g_\gamma(x) \cdot (1 + |X_0 - X_{0,k}|)^{1+\gamma}\,. \end{aligned} \tag{40}$$

Combining (39) with the representation of $G'_{k-1,n}$ from Lemma A.4.(ii) and then the bound from (40), we bound R_1 as

$$|R_1| \leq (1 + |X_0 - X_{0,k}|)^{1+\gamma} \cdot \int_{\mathbb{R}} G_n(x) g_\gamma(x)\,\mathrm{d}x \cdot \int_{\mathbb{R}} \left| \int_{a_k z}^{a_k\varepsilon_{-k}} |s|^\gamma\,\mathrm{d}s \right| \mathrm{d}\mathbf{P}_\varepsilon(z)\,.$$

Finally, by (37) and using $\varepsilon \in L^{1+\gamma}(\mathbf{P})$ as $1+\gamma < \alpha$, we conclude

$$\begin{aligned} |R_1| &\lesssim (1+|X_0 - X_{0,k}|)^{1+\gamma} \cdot u_n^{-\gamma+\gamma_G} \cdot |a_k|^{1+\gamma} \cdot \left(|\varepsilon_{-k}|^{1+\gamma} + \mathbf{E}|\varepsilon|^{1+\gamma}\right) \\ &\lesssim (1+|X_0 - X_{0,k}|^{1+\gamma}) \cdot u_n^{-\gamma+\gamma_G} \cdot |a_k|^{1+\gamma} \cdot [|\varepsilon_{-k}|^{1+\gamma} \vee 1]\,. \end{aligned} \tag{41}$$

We now show that

$$\mathbf{E}|X_0 - X_{0,k}|^p \lesssim k^{1-(1-d)p} \lesssim 1 \qquad \text{for } p \in \left(\frac{1}{1-d}\,, \alpha\right). \tag{42}$$

To this end, define the array and limiting sequence

$$(Y_{m,i}) = (a_{k+i}\varepsilon_{-(k+i)}\,, i = 1, \ldots, m\,, m \in \mathbb{N}) \qquad \text{and} \qquad (Y_i) = (a_{k+i}\varepsilon_{-(k+i)})_{i\in\mathbb{N}}\,.$$

Note that (10) and (12) are satisfied for $(Y_{m,i})$ and (Y_i). Then we may write

$$X_0 - X_{0,k} = \sum_{i=k+1}^{\infty} a_i \varepsilon_{-i} = \sum_{i=1}^{\infty} a_{k+i}\varepsilon_{-(k+i)} = \sum_{i=1}^{\infty} Y_i\,.$$

The ε_j are independent, centered and belong to $L^p(\mathbf{P})$, so applying Inequality (13) in Lemma A.1.(ii), we obtain

$$\mathbf{E}|X_0 - X_{0,k}|^p \lesssim \sum_{i=k+1}^{\infty} |a_i|^p \mathbf{E}|\varepsilon_{-i}|^p \lesssim \sum_{i=k+1}^{\infty} i^{-(1-d)p} \lesssim k^{1-(1-d)p} \leq 1\,,$$

where we used $|a_i| \asymp i^{-(1-d)}$ and $(1-d)p > 1$. This proves (42). To conclude, note that

$$\frac{1}{1-d} < (1+\gamma)r < \left(1 + \frac{\alpha}{r} - 1\right) r = \alpha\,.$$

Since $X_0 - X_{0,k}$ and ε_{-k} are independent, and using the moment bound from (42) with $p = (1+\gamma)r < \alpha$ along with $|a_i| \asymp i^{-(1-d)}$, we obtain

$$\left\|(1+|X_0 - X_{0,k}|)^{1+\gamma} \cdot u_n^{-\gamma+\gamma_G} \cdot |a_k|^{1+\gamma} \cdot (1 \vee |\varepsilon_{-k}|)^{1+\gamma}\right\|_{L^r(\mathbf{P})} \lesssim u_n^{-\gamma+\gamma_G} \cdot k^{-(1-d)(1+\gamma)}\,.$$

Combining this with the bound from (41), we conclude

$$\|R_1 \mathbb{1}\{1 \geq (|a_k z| \vee |a_k \varepsilon_{-k})\}\|_{L^r(\mathbf{P})} \lesssim k^{-(1-d)(1+\gamma)} u_n^{-\gamma+\gamma_G} \qquad \text{for } k > k_0 + 1\,.$$

Case $|a_k\varepsilon_{-k}| > (1 \vee |a_k z|)$ We now estimate R_1 restricted to the event $|a_k\varepsilon_{-k}| > (1 \vee |a_k z|)$. To do so, we note that on this event $|\varepsilon_{-k}| > |z|$ and we split the integrand in (38), obtaining

$$\begin{aligned} &|R_1 \mathbb{1}\{|a_k\varepsilon_{-k}| > (1 \vee |a_k z|)\}| \\ &\leq \mathbb{1}\{|a_k\varepsilon_{-k}| > 1\} \left| \int_{\mathbb{R}} \left(\int_{a_k z}^{a_k\varepsilon_{-k}} G'_{k-1,n}(X_0 - X_{0,k} + s) \mathbb{1}\{|\varepsilon_{-k}| > |z|\}\, \mathrm{d}s \right) \mathrm{d}\mathbf{P}_\varepsilon(z) \right| \\ &+ \mathbb{1}\{|a_k\varepsilon_{-k}| > 1\} \left| \int_{\mathbb{R}} \left(\int_{a_k z}^{a_k\varepsilon_{-k}} G'_{k-1,n}(X_0 - X_{0,k}) \mathbb{1}\{|\varepsilon_{-k}| > |z|\}\, \mathrm{d}s \right) \mathrm{d}\mathbf{P}_\varepsilon(z) \right|, \end{aligned} \tag{43}$$

and bound each term separately. To bound the first term in (43), we note that necessarily $|s| < |a_k\varepsilon_{-k}|$ and we use successively Lemma A.4.(ii), the bound $|f'_{k-1}(x)| \lesssim g_\gamma(x)$ (from

Lemma A.2), Inequality (18) from Lemma A.3, Inequality (17) from Lemma A.3, and (37), resulting in

$$\begin{aligned}
&\mathbb{1}\{|a_k\varepsilon_{-k}|>1\}\left|\int_{\mathbb{R}}\left(\int_{a_kz}^{a_k\varepsilon_{-k}} G'_{k-1,n}(X_0-X_{0,k}+s)\mathbb{1}\{|\varepsilon_{-k}|>|z|\}\,\mathrm{d}s\right)\mathrm{d}\mathbf{P}_\varepsilon(z)\right|\\
&\le\ \mathbb{1}\{|a_k\varepsilon_{-k}|>1\}\int_{|s|\le|a_k\varepsilon_{-k}|}\left(\int_{\mathbb{R}} G_n(x)|f'_{k-1}(x-(X_0-X_{0,k}+s))|\,\mathrm{d}x\right)\mathrm{d}s\\
&\lesssim\ \mathbb{1}\{|a_k\varepsilon_{-k}|>1\}\int_{\mathbb{R}} G_n(x)\left(\int_{|s|\le|a_k\varepsilon_{-k}|} g_\gamma((x-(X_0-X_{0,k}))-s)\,\mathrm{d}s\right)\mathrm{d}x\\
&\lesssim\ |a_k\varepsilon_{-k}|^{1+\gamma}\int_{\mathbb{R}} G_n(x)g_\gamma(x-(X_0-X_{0,k}))\,\mathrm{d}x\\
&\lesssim\ |a_k\varepsilon_{-k}|^{1+\gamma}(1+|X_0-X_{0,k}|^{1+\gamma})\int_{\mathbb{R}} G_n(x)g_\gamma(x)\,\mathrm{d}x\\
&\lesssim\ |a_k\varepsilon_{-k}|^{1+\gamma}(1+|X_0-X_{0,k}|^{1+\gamma})\cdot u_n^{-\gamma+\gamma_G}\,.
\end{aligned}$$

We now similarly estimate the second term in (43) using that $G'_{k-1,n}(X_0-X_{0,k})$ is constant in s and we obtain

$$\begin{aligned}
&\mathbb{1}\{|a_k\varepsilon_{-k}|>1\}\left|\int_{\mathbb{R}}\left(\int_{a_kz}^{a_k\varepsilon_{-k}} G'_{k-1,n}(X_0-X_{0,k})\mathbb{1}\{|\varepsilon_{-k}|>|z|\}\,\mathrm{d}s\right)\mathrm{d}\mathbf{P}_\varepsilon(z)\right|\\
&\le\ \mathbb{1}\{|a_k\varepsilon_{-k}|>1\}\int_{|s|\le|a_k\varepsilon_{-k}|}\left(\int_{\mathbb{R}} G_n(x)|f'_{k-1}(x-(X_0-X_{0,k}))|\,\mathrm{d}x\right)\mathrm{d}s\\
&\lesssim\ \mathbb{1}\{|a_k\varepsilon_{-k}|>1\}|a_k\varepsilon_{-k}|\int_{\mathbb{R}} G_n(x)g_\gamma(x-(X_0-X_{0,k}))\,\mathrm{d}x\\
&\lesssim\ |a_k\varepsilon_{-k}|^{1+\gamma}(1+|X_0-X_{0,k}|^{1+\gamma})\cdot u_n^{-\gamma+\gamma_G}\,,
\end{aligned}$$

where, in the last step, we also used that $|a_k\varepsilon_{-k}|\le|a_k\varepsilon_{-k}|^{1+\gamma}$ due to $|a_k\varepsilon_{-k}|>1$. Combining the bounds for both terms and applying the moment bound from (42) and $|a_k|\sim k^{-(1-d)}$, we obtain

$$\|R_1\mathbb{1}\{|a_k\varepsilon_{-k}|>(1\vee|a_kz|)\}\|_{L^r(\mathbf{P})}\ \lesssim\ k^{-(1+\gamma)(1-d)}u_n^{-\gamma+\gamma_G}\,.$$

Case $|a_kz|>(1\vee|a_k\varepsilon_{-k}|)$ As in the previous case we split the integrand in (38):

$$\begin{aligned}
&|R_1\mathbb{1}\{|a_kz|>(1\vee|a_k\varepsilon_{-k}|)\}|\\
&\le\ \left|\int_{\mathbb{R}}\mathbb{1}\{|a_kz|>1\}\left(\int_{a_kz}^{a_k\varepsilon_{-k}} G'_{k-1,n}(X_0-X_{0,k}+s)\mathbb{1}\{|z|>|\varepsilon_{-k}|\}\,\mathrm{d}s\right)\mathrm{d}\mathbf{P}_\varepsilon(z)\right|\\
&+\ \left|\int_{\mathbb{R}}\mathbb{1}\{|a_kz|>1\}\left(\int_{a_kz}^{a_k\varepsilon_{-k}} G'_{k-1,n}(X_0-X_{0,k})\mathbb{1}\{|z|>|\varepsilon_{-k}|\}\,\mathrm{d}s\right)\mathrm{d}\mathbf{P}_\varepsilon(z)\right|\,.
\end{aligned} \tag{44}$$

We now bound the first term in (44). The argument proceeds as in the previous case: we apply Lemma A.4.(ii) and estimate the integrals using the smoothness and decay properties of the derivative of the density f'_{k-1}. We find:

$$\begin{aligned}
&\left|\int_{\mathbb{R}} \mathbb{1}\{|a_k z| > 1\} \left(\int_{a_k z}^{a_k \varepsilon_{-k}} G'_{k-1,n}(X_0 - X_{0,k} + s)\mathbb{1}\{|z| > |\varepsilon_{-k}|\}\, \mathrm{d}s\right) \mathrm{d}\mathbf{P}_\varepsilon(z)\right| \\
&\leq \int_{\mathbb{R}} \mathbb{1}\{|a_k z| > 1\} \left(\int_{\mathbb{R}} \left(\int_{|s|\leq|a_k z|} G_n(x)|f'_{k-1}(x - (X_0 - X_{0,k} + s))|\, \mathrm{d}s\right) \mathrm{d}x\right) \mathrm{d}\mathbf{P}_\varepsilon(z) \\
&\lesssim \int_{\mathbb{R}} G_n(x) \left(\int_{\mathbb{R}} \mathbb{1}\{|a_k z| > 1\} \left(\int_{|s|\leq|a_k z|} g_\gamma((x - (X_0 - X_{0,k})) - s)\, \mathrm{d}s\right) \mathrm{d}\mathbf{P}_\varepsilon(z)\right) \mathrm{d}x \\
&\lesssim \left(\int_{\mathbb{R}} G_n(x) g_\gamma(x - (X_0 - X_{0,k}))\, \mathrm{d}x\right) \left(\int_{\mathbb{R}} |a_k z|^{1+\gamma}\, \mathrm{d}\mathbf{P}_\varepsilon(z)\right) \\
&\lesssim |a_k|^{1+\gamma} \mathbf{E}|\varepsilon|^{1+\gamma} (1 + |X_0 - X_{0,k}|)^{1+\gamma} \int_{\mathbb{R}} G_n(x) g_\gamma(x)\, \mathrm{d}x \\
&\lesssim |a_k|^{1+\gamma} (1 + |X_0 - X_{0,k}|)^{1+\gamma} u_n^{-\gamma+\gamma_G}\,.
\end{aligned}$$

We now bound the second term in (44), where $G'_{k-1,n}(X_0 - X_{0,k})$ is constant in s:

$$\begin{aligned}
&\left|\int_{\mathbb{R}} \mathbb{1}\{|a_k z| > 1\} \left(\int_{a_k z}^{a_k \varepsilon_{-k}} G'_{k-1,n}(X_0 - X_{0,k})\mathbb{1}\{|z| > |\varepsilon_{-k}|\}\, \mathrm{d}s\right) \mathrm{d}\mathbf{P}_\varepsilon(z)\right| \\
&\leq \int_{\mathbb{R}} \mathbb{1}\{|a_k z| > 1\} \left(\int_{|s|\leq|a_k z|} \left(\int_{\mathbb{R}} G_n(x)|f'_{k-1}(x - (X_0 - X_{0,k}))|\, \mathrm{d}x\right) \mathrm{d}s\right) \mathrm{d}\mathbf{P}_\varepsilon(z) \\
&\lesssim \left(\int_{|a_k z|>1} |a_k z| \mathrm{d}\mathbf{P}_\varepsilon(z)\right) \left(\int_{\mathbb{R}} G_n(x) g_\gamma(x - (X_0 - X_{0,k}))\, \mathrm{d}x\right) \\
&\lesssim \left(\int_{|a_k z|>1} |a_k z|^{1+\gamma} \mathrm{d}\mathbf{P}_\varepsilon(z)\right) (1 + |X_0 - X_{0,k}|)^{1+\gamma} \left(\int_{\mathbb{R}} G_n(x) g_\gamma(x)\, \mathrm{d}x\right) \\
&\lesssim |a_k|^{1+\gamma} (1 + |X_0 - X_{0,k}|)^{1+\gamma} u_n^{-\gamma+\gamma_G}\,.
\end{aligned}$$

Combining both bounds with the moment bound in (42), we conclude analogously to the previous case that

$$\|R_1 \mathbb{1}\{|a_k z| > (1 \vee |a_k \varepsilon_{-k}|)\}\|_{L^r(\mathbf{P})} \lesssim k^{-(1+\gamma)(1-d)} u_n^{-\gamma+\gamma_G}\,.$$

This completes the proof of the control of R_1.

B.2.2 Analysis of R_2

We begin with the estimate

$$\begin{aligned}
&|f'_\infty(x - (X_0 - X_{0,k})) - f'_\infty(x)| \\
&\leq |f'_\infty(x - (X_0 - X_{0,k})) - f'_\infty(x)| \cdot \mathbb{1}\{|X_0 - X_{0,k}| \leq 1\} \\
&+ \big(|f'_\infty(x - (X_0 - X_{0,k}))| + |f'_\infty(x)|\big)\, \mathbb{1}\{|X_0 - X_{0,k}| > 1\}\,.
\end{aligned} \tag{45}$$

Note that by Inequality (15) from Lemma A.2, we have

$$\begin{aligned}
&|f'_\infty(x - (X_0 - X_{0,k})) - f'_\infty(x)| \cdot \mathbb{1}\{|X_0 - X_{0,k}| \leq 1\} \\
&\lesssim |X_0 - X_{0,k}| \cdot g_\gamma(x)\,.
\end{aligned} \tag{46}$$

Moreover, using (14) from Lemma A.2 and (17) from Lemma A.3, it follows that

$$\begin{aligned}&\big(|f'_\infty(x-(X_0-X_{0,k}))| \;+\; |f'_\infty(x)|\big)\,\mathbb{1}\{|X_0-X_{0,k}|>1\}\\ &\lesssim\; (g_\gamma(x-(X_0-X_{0,k}))+g_\gamma(x))\,\mathbb{1}\{|X_0-X_{0,k}|>1\}\\ &\lesssim\; g_\gamma(x)\cdot|X_0-X_{0,k}|^{1+\gamma}\,.\end{aligned}\tag{47}$$

Therefore, combining Inequalities (45)-(47), we get

$$|f'_\infty(x-(X_0-X_{0,k})) \;-\; f'_\infty(x)| \;\lesssim\; g_\gamma(x)\left(|X_0-X_{0,k}|\vee|X_0-X_{0,k}|^{1+\gamma}\right).\tag{48}$$

Applying successively Lemma A.4.(ii), Inequality (48), and (37), we obtain

$$\begin{aligned}&|G'_{\infty,n}(X_0-X_{0,k}) \;-\; G'_{\infty,n}(0)|\\ &\le\; \int_{\mathbb{R}} G_n(x)\cdot|f'_\infty(x-(X_0-X_{0,k})) \;-\; f'_\infty(x)|\,\mathrm{d}x\\ &\lesssim\; \left(|X_0-X_{0,k}|\vee|X_0-X_{0,k}|^{1+\gamma}\right)\int_{\mathbb{R}} G_n(x)g_\gamma(x)\,\mathrm{d}x\\ &\lesssim\; \left(|X_0-X_{0,k}|\vee|X_0-X_{0,k}|^{1+\gamma}\right)u_n^{-\gamma+\gamma_G}\,.\end{aligned}\tag{49}$$

From the definition of R_2, together with the independence of $X_0-X_{0,k}$ and ε_{-k}, and Inequality (49), we find

$$\|R_2\|_{L^r(\mathbf{P})} \;\lesssim\; k^{-(1-d)}\left(\|X_0-X_{0,k}\|_{L^r(\mathbf{P})} \;+\; \big\||X_0-X_{0,k}|^{1+\gamma}\big\|_{L^r(\mathbf{P})}\right)u_n^{-\gamma+\gamma_G}\,,\tag{50}$$

since $|a_k|\asymp k^{-(1-d)}$, and $\varepsilon\in L^r(\mathbf{P})$ for $r<\alpha$. We now apply Estimate (42) for both $p=r$ and $p=r(1+\gamma)$. Since $1/(1-d)<r<r(1+\gamma)<\alpha$ by the condition in Lemma (B.3)(ii), it follows

$$\begin{aligned}\|X_0-X_{0,k}\|_{L^r(\mathbf{P})} \;+\; \big\||X_0-X_{0,k}|^{1+\gamma}\big\|_{L^r(\mathbf{P})} \;&\lesssim\; k^{1/r-(1-d)} \;+\; k^{1/r-(1-d)(1+\gamma)}\\ &\lesssim\; k^{1/r-(1-d)}\,.\end{aligned}\tag{51}$$

Note finally that

$$\begin{aligned}\frac{1}{r}-2(1-d) \;&=\; -(1-d)\left(1-\frac{1}{r(1-d)}\right) \;-\; (1-d)\\ &\le\; -(1-d)\gamma \;-\; (1-d)\\ &=\; -(1-d)(1+\gamma)\,.\end{aligned}$$

Combining this inequality with Inequalities (50) and (51), we obtain

$$\|R_2\|_{L^r(\mathbf{P})} \;\lesssim\; k^{1/r-2(1-d)}u_n^{-\gamma+\gamma_G} \;\lesssim\; k^{-(1+\gamma)(1-d)}u_n^{-\gamma+\gamma_G}\,.\tag{52}$$

This completes the proof of the control of R_2.

B.2.3 Analysis of R_3

Using Inequality (16) from Lemma A.2 and Inequality (17) from Lemma A.3, we obtain

$$\begin{aligned}&|f'_\infty(x-(X_0-X_{0,k})) \;-\; f'_{k-1}(x-(X_0-X_{0,k}))|\\ &\lesssim\; k^{-(1-d)+1/r}g_\gamma(x-(X_0-X_{0,k}))\\ &\lesssim\; k^{-(1-d)+1/r}g_\gamma(x)(1\vee|X_0-X_{0,k}|)^{1+\gamma}\,.\end{aligned}\tag{53}$$

Applying Lemma A.4.(ii) together with Inequality (53) and then (37), we obtain

$$\begin{aligned} &\left|G'_{k-1,n}(X_0 - X_{0,k}) \;-\; G'_{\infty,n}(X_0 - X_{0,k})\right| \\ &\lesssim \int_{\mathbb{R}} G_n(x)|f'_\infty(x - (X_0 - X_{0,k})) \;-\; f'_{k-1}(x - (X_0 - X_{0,k}))|\,\mathrm{d}x \\ &\lesssim k^{-(1-d)+1/r}\left(1 \vee |X_0 - X_{0,k}|^{1+\gamma}\right) u_n^{-\gamma+\gamma_G} . \end{aligned} \tag{54}$$

Using the independence of $X_0 - X_{0,k}$ and $\varepsilon_{-k} \stackrel{d}{=} \varepsilon \in L^r(\mathbf{P})$, together with Inequality (54) and the fact that $|X_0 - X_{0,k}|^{1+\gamma} \in L^r(\mathbf{P})$ by (42), we obtain

$$\|R_3\|_{L^r(\mathbf{P})} \;\lesssim\; |a_k| u_n^{-\gamma+\gamma_G} k^{-(1-d)+1/r}\left(1 + \left\||X_0 - X_{0,k}|^{1+\gamma}\right\|_{L^r(\mathbf{P})}\right) .$$

Since $|a_k| \asymp k^{-(1-d)}$, we conclude, using the same argument as in (52), that

$$\|R_3\|_{L^r(\mathbf{P})} \;\lesssim\; k^{-2(1-d)+1/r} u_n^{-\gamma+\gamma_G} \;\lesssim\; k^{-(1-d)(1+\gamma)} u_n^{-\gamma+\gamma_G} .$$

This completes the proof of the control of R_3, and therefore the proof of Lemma B.3.(ii). □

We have bounded the individual summands in Lemma B.2. To advance the proof and finally show Equation (36), we now require the following lemma, which controls the rate of divergence of a convolution-type sum that arises from applying the individual bounds.

Lemma B.4. *Let $\gamma, r \in \mathbb{R}$ satisfy the assumptions of Lemma B.3(ii). Then for all $n \in \mathbb{N}$, we have*

$$\sum_{k=-\infty}^{n}\left(\sum_{t=k\vee 1}^{n}\left(1 \wedge (t-k)^{-(1-d)(1+\gamma)}\right)\right)^r \;\lesssim\; n^{r+1-(1-d)(1+\gamma)r} ,$$

where $\lesssim$ denotes inequality up to a constant $C_{\gamma,\alpha,d} > 0$, depending only on γ, α, and d. We adopt the convention $1/0 = \infty$.

Remark B.5. The above lemma actually holds with the relaxed conditions

$$0 \;<\; \gamma \;<\; \frac{d}{1-d} \qquad \text{and} \qquad \frac{1}{1-d} \;<\; r .$$

Proof. The sum in question can be decomposed and reindexed as

$$\begin{aligned} &\sum_{k=-\infty}^{0}\left(\sum_{t=1}^{n}(t-k)^{-(1-d)(1+\gamma)}\right)^r \;+\; \sum_{k=1}^{n}\left(1 \;+\; \sum_{t=k+1}^{n}(t-k)^{-(1-d)(1+\gamma)}\right)^r \\ &= \sum_{k=0}^{\infty}\left(\sum_{t=k+1}^{n+k} t^{-(1-d)(1+\gamma)}\right)^r \;+\; \sum_{k=1}^{n}\left(1 \;+\; \sum_{t=1}^{n-k} t^{-(1-d)(1+\gamma)}\right)^r \\ &= \left(\sum_{t=1}^{n} t^{-(1-d)(1+\gamma)}\right)^r \;+\; \sum_{k=n+1}^{\infty}\left(\sum_{t=k+1}^{n+k} t^{-(1-d)(1+\gamma)}\right)^r \\ &\quad + \sum_{k=1}^{n}\left(\left(1 \;+\; \sum_{t=1}^{n-k} t^{-(1-d)(1+\gamma)}\right)^r \;+\; \left(\sum_{t=k+1}^{n+k} t^{-(1-d)(1+\gamma)}\right)^r\right) . \end{aligned}$$

The first and second terms correspond to $k = 0$ and $k > n$, respectively, in the first double sum of the previous equation. The third term arises from a combination of the remaining terms $k \in \{1, \dots, n\}$ in the first double sum and all terms in the second double sum of the previous equation. From the assumptions on γ and r, we observe that

$$(1-d)(1+\gamma) < 1 \qquad \text{and} \qquad r(1-d)(1+\gamma) > 1\,.$$

We begin by estimating the first term in the decomposition above. Since $(1-d)(1+\gamma) < 1$, we have

$$\left(\sum_{t=1}^{n} t^{-(1-d)(1+\gamma)}\right)^{r} \lesssim n^{r-(1-d)(1+\gamma)r} \le n^{r+1-(1-d)(1+\gamma)r}\,. \tag{55}$$

For $k \geq 1$, again since $(1-d)(1+\gamma) < 1$, we also have

$$\begin{aligned} \sum_{t=k+1}^{n+k} t^{-(1-d)(1+\gamma)} &\le \min\left(n \cdot k^{-(1-d)(1+\gamma)}, \int_{k}^{n+k} t^{-(1-d)(1+\gamma)}\,\mathrm{d}t\right) \\ &\lesssim \min\left(n \cdot k^{-(1-d)(1+\gamma)}, (n+k)^{1-(1-d)(1+\gamma)}\right)\,. \end{aligned} \tag{56}$$

Using the first bound in the Estimate (56) and the fact that $(1-d)(1+\gamma)r > 1$, we obtain

$$\begin{aligned} &\sum_{k=n+1}^{\infty}\left(\sum_{t=k+1}^{n+k} t^{-(1-d)(1+\gamma)}\right)^{r} \\ &\lesssim n^{r}\sum_{k=n+1}^{\infty} k^{-(1-d)(1+\gamma)r} \lesssim n^{r}\int_{n}^{\infty} t^{-(1-d)(1+\gamma)r}\,\mathrm{d}t \lesssim n^{r+1-(1-d)(1+\gamma)r}\,. \end{aligned}$$

Applying the second bound in Estimate (56), we obtain

$$\sum_{k=1}^{n}\left(\sum_{t=k+1}^{n+k} t^{-(1-d)(1+\gamma)}\right)^{r} \lesssim \sum_{k=1}^{n}(n+k)^{r-(1-d)(1+\gamma)r} \lesssim n^{r+1-(1-d)(1+\gamma)r}\,.$$

Finally, recalling the first bound in (55), we obtain

$$\sum_{k=1}^{n}\left(1 + \sum_{t=1}^{n-k} t^{-(1-d)(1+\gamma)}\right)^{r} \lesssim n\left(\sum_{t=1}^{n} t^{-(1-d)(1+\gamma)}\right)^{r} \lesssim n^{r+1-(1-d)(1+\gamma)r}\,.$$

This completes the proof of Lemma B.4. □

With all necessary tools in hand, we now turn to the final part of the proof.

B.3 Proof of Theorem 3.3

Let $n > k_0$. We are going to apply Lemmas B.2, B.3, and B.4. We begin the proof. By Lemma B.2, it follows

$$\mathbf{E}\left|\sum_{t=1}^{n} G_n(X_t) - \mathbf{E}[G_n(X_t)] - G'_{\infty,n}(0)X_t\right|^{r} \lesssim \sum_{k=-\infty}^{n}\left(\sum_{t=k\vee 1}^{n} \|P_{k-t}\mathcal{U}_n(X_0)\|_{L^r(\mathbf{P})}\right)^{r},$$

where $\mathcal{U}_n(X_t) := G_n(X_t) - \mathbf{E}[G_n(X_t)] - G'_{\infty,n}(0)X_t$. We split the inner sum based on whether $t-k > k_0+1$ and use the convexity of $x \mapsto x^r$ (due to $r>1$) to get

$$\begin{aligned}
&\mathbf{E}\left|\sum_{t=1}^{n} G_n(X_t) \;-\; \mathbf{E}[G_n(X_t)] \;-\; G'_{\infty,n}(0)X_t\right|^r \\
&\lesssim \sum_{k=-\infty}^{n}\left(\sum_{\substack{(1\vee k)\le t\le n\\ t-k>k_0+1}} \|P_{k-t}\mathcal{U}_n(X_0)\|_{L^r(\mathbf{P})}\right)^r \\
&\quad + \sum_{k=-\infty}^{n}\left(\sum_{\substack{(1\vee k)\le t\le n\\ t-k\le k_0+1}} \|P_{k-t}\mathcal{U}_n(X_0)\|_{L^r(\mathbf{P})}\right)^r .
\end{aligned}$$

For the first term, we apply Lemma B.3.(ii) and Lemma B.4. This yields

$$\begin{aligned}
&\sum_{k=-\infty}^{n}\left(\sum_{\substack{(1\vee k)\le t\le n\\ t-k>k_0+1}} \|P_{k-t}\mathcal{U}_n(X_0)\|_{L^r(\mathbf{P})}\right)^r \\
&\lesssim u_n^{(\gamma_G-\gamma)r}\sum_{k=-\infty}^{n}\left(\sum_{t=1\vee k}^{n}\left(1\wedge (t-k)^{-(1-d)(1+\gamma)}\right)\right)^r \\
&\lesssim u_n^{(\gamma_G-\gamma)r} n^{r+1-(1-d)(1+\gamma)r} .
\end{aligned}$$

For the second term, note that the constraints $t \le k_0+k+1$ and $t \ge (1\vee k)$ ensure that the inner sum is empty when $k < -k_0$. Therefore, applying Lemma B.1.(ii), we obtain for sufficiently large $n \in \mathbb{N}$ that

$$\begin{aligned}
&\sum_{k=-\infty}^{n}\left(\sum_{\substack{(1\vee k)\le t\le n\\ t-k\le k_0+1}} \|P_{k-t}\mathcal{U}_n(X_0)\|_{L^r(\mathbf{P})}\right)^r \\
&\lesssim \left(\|G_n(X_0)\|_{L^r(\mathbf{P})} \vee |G'_{\infty,n}(0)|\right)^r \sum_{k=-k_0}^{n}\left(\#\{t\in\mathbb{N}:\ (1\vee k)\le t\le (n\wedge(k+k_0+1))\}\right)^r \\
&\le \left(\|G_n(X_0)\|_{L^r(\mathbf{P})} \vee |G'_{\infty,n}(0)|\right)^r \sum_{k=-k_0}^{n}\left(\#\{t\in\mathbb{N}:\ k\le t\le k+k_0+1\}\right)^r \\
&= \left(\|G_n(X_0)\|_{L^r(\mathbf{P})} \vee |G'_{\infty,n}(0)|\right)^r (n+k_0+1)(k_0+2)^r \\
&\lesssim n\left(\mathbf{E}|G_n(X_0)|^r \vee |G'_{\infty,n}(0)|^r\right) .
\end{aligned}$$

We conclude that

$$\mathbf{E}\left|\sum_{t=1}^{n}\mathcal{U}_n(X_t)\right|^r \;\lesssim\; u_n^{(\gamma_G-\gamma)r} n^{r+1-(1-d)(1+\gamma)r} \;+\; n\left(\mathbf{E}|G_n(X_0)|^r \vee |G'_{\infty,n}(0)|^r\right) .$$

This is the first inequality we wanted to prove. For the second inequality, observe that $1-(1-d)(1+\gamma) \ge 0$ due to $\gamma \le d/(1-d)$. Therefore the n rate of the first term dominates.

Since $\gamma_G < \gamma$, it holds that $u_n^{\gamma_G-\gamma} < \infty$ for all $n \in \mathbb{N}$. Applying Lemma A.4 gives that $\mathbf{E}|G_n(X_0)|^r < \infty$ and $|G'_{\infty,n}(0)| < \infty$ for all $n \in \mathbb{N}$, such that we may conclude

$$u_n^{(\gamma_G-\gamma)r} n^{r+1-(1-d)(1+\gamma)r} \; + \; n\left(\mathbf{E}|G_n(X_0)|^r \vee |G'_{\infty,n}(0)|^r\right) \;\lesssim\; n^{r+1-(1-d)(1+\gamma)r}\,.$$

Taking power $1/r$ on both sides of the inequality finishes the proof of Theorem 3.3. $\square$

Having established a $L^r(\mathbf{P})$ moment bound in Theorem 3.3, we now turn to the proof of Theorem 3.5 which sharpens the result by identifying the optimal rate of convergence. This requires a careful balance of the parameters γ and r.

B.4 Proof of Theorem 3.5

Combining Theorem 3.3 with the following proposition immediately gives the desired result.

Proposition B.6. *Denote* $\kappa(\gamma, r) = 1 + 1/r - (1-d)(1+\gamma)$.

(i) There exists a pair of unique minimizers (γ_0, r_0) *of* $\kappa(\gamma, r)$ *over the compact domain*

$$0 \;\le\; \gamma \;\le\; \min\left\{\frac{d}{1-d}\,, 1 - \frac{1}{r(1-d)}\,, \frac{\alpha}{r} - 1\right\} \qquad \textit{and} \qquad \frac{1}{1-d} \;\le\; r \;\le\; \alpha\,,$$

namely

$$\gamma_0 := \begin{cases} \frac{d}{1-d} & \textit{if } \frac{1}{(1-d)(1-2d)} < \alpha\,, \\ \frac{\alpha(1-d)-1}{\alpha(1-d)+1} & \textit{else,} \end{cases} \qquad r_0 := \begin{cases} \alpha(1-d) & \textit{if } \frac{1}{(1-d)(1-2d)} < \alpha\,, \\ \frac{1}{2}\left(\frac{1}{1-d} \;+\; \alpha\right) & \textit{else,} \end{cases}$$

yielding

$$\kappa_0 \;=\; \kappa(\gamma_0, r_0) \;=\; \begin{cases} \frac{1}{\alpha(1-d)} \quad \textit{if} \quad \frac{1}{(1-d)(1-2d)} < \alpha\,, \\[1ex] \frac{2(1-d)+(1-\alpha(1-d)(1-2d))}{\alpha(1-d)+1} = d \;+\; (1-d)\frac{3-\alpha(1-d)}{\alpha(1-d)+1} \quad \textit{else.} \end{cases}$$

(ii) It holds

$$\gamma_0 \;=\; \min\left\{\frac{d}{1-d}\,, 1 - \frac{1}{r_0(1-d)}\,, \frac{\alpha}{r_0} - 1\right\}, \qquad r_0 \in \left(\frac{1}{1-d}\,, \alpha\right),$$

and

$$\kappa_0 \;-\; (d + 1/\alpha) \;<\; 0\,.$$

Proof. **Proof of Part (i):** The goal is to minimize

$$\kappa(\gamma, r) \;=\; 1 \;+\; \frac{1}{r} \;-\; (1-d)(1+\gamma)$$

over γ and r from the closure of the range in Theorem 3.3(i), namely such that

$$0 \;\le\; \gamma \;\le\; \gamma(r) \qquad \text{and} \qquad \frac{1}{1-d} \;\le\; r \;\le\; \alpha\,,$$

where the function $\gamma(r)$ is defined as

$$\gamma(r) \;:=\; \gamma_1(r) \;\wedge\; \gamma_2(r) \;\wedge\; \gamma_3(r)\,,$$

with

$$\gamma_1(r) \;:=\; \frac{d}{1-d}\,, \qquad \gamma_2(r) \;:=\; 1-\frac{1}{r(1-d)}\,, \qquad \gamma_3(r) \;:=\; \frac{\alpha}{r}-1\,.$$

The function κ is continuous and strictly decreasing in both γ and r, which implies that

$$(\gamma_0, r_0) \;=\; \operatorname{argmin}\left\{\kappa(\gamma,r)\,\middle|\,(\gamma,r)\in K\right\} \quad \text{with} \quad K := \left\{(\gamma(r),r)\,\middle|\,\frac{1}{1-d} \;\leq\; r \;\leq\; \alpha\right\}.$$

Moreover, for any admissible value of r, the value of $\kappa(\gamma(r),r)$ is determined by the smallest among $\gamma_1(r),\gamma_2(r),\gamma_3(r)$. Specifically,

$$\kappa(\gamma(r),r) \;=\; \begin{cases} \kappa(\gamma_1(r),r)\,, & \gamma_1(r) \leq (\gamma_2(r)\wedge\gamma_3(r))\,,\\ \kappa(\gamma_2(r),r)\,, & \gamma_2(r) \leq (\gamma_1(r)\wedge\gamma_3(r))\,,\\ \kappa(\gamma_3(r),r)\,, & \gamma_3(r) \leq (\gamma_1(r)\wedge\gamma_2(r))\,. \end{cases}$$

Note that $r\mapsto\kappa(\gamma(r),r)$ is not differentiable everywhere. We shall therefore argue that on the set where it is differentiable the derivative does not vanish and therefore the optimal values have to lie on the boundary. Note that this set is essentially

$$\{r>0 \mid (\gamma_1(r),\gamma_2(r),\gamma_3(r)) \text{ are pairwise distinct}\}\,,$$

and it is nonempty because $\alpha > 1/(1-d)$, and open since each $\gamma_j(r)$ is continuous. Next, we compute the derivative of $r\mapsto\kappa(\gamma(r),r)$ on this set. Clearly, for all $r>0$ it holds that

$$\gamma_1'(r) \;=\; 0\,, \qquad \gamma_2'(r) \;=\; \frac{1}{r^2(1-d)}\,, \qquad \gamma_3'(r) \;=\; -\frac{\alpha}{r^2}\,.$$

Applying the chain rule yields

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}r}\kappa(\gamma_j(r),r) \;&=\; \gamma_j'(r)\frac{\partial}{\partial\gamma}\kappa(\gamma_j(r),r) \;+\; \frac{\partial}{\partial r}\kappa(\gamma_j(r),r)\\ &=\; -\gamma_j'(r)(1-d) \;-\; \frac{1}{r^2}\,.\end{aligned}$$

Setting the derivative equal to zero gives

$$\gamma_j'(r) \;=\; -\frac{1}{r^2(1-d)}\,.$$

This equation clearly has no solution for $j=1,2$. For $j=3$, solving yields $\alpha = 1/(1-d)$, which contradicts our assumption that $\alpha > 1/(1-d)$. Therefore, we conclude that for all $r>0$ it holds that

$$\frac{\mathrm{d}}{\mathrm{d}r}\kappa(\gamma_j(r),r) \;\neq\; 0 \qquad \text{for } j=1,2,3\,.$$

To complete the analysis, we examine the behavior of $\kappa(\gamma(r),r)$ at values of r where two or more of the functions $\gamma_j(r)$ coincide.

We begin by identifying the intersection points of the functions $\gamma_1(r)$, $\gamma_2(r)$, and $\gamma_3(r)$. Note that γ_1 is constant, γ_2 is strictly increasing, and γ_3 is strictly decreasing. Hence, each pair of functions can intersect at most once. These intersections occur at

$$\gamma_1 \;=\; \gamma_2(r) \;=\; \frac{d}{1-d} \qquad \text{if and only if} \qquad r \;=\; r_{1,2} \;:=\; \frac{1}{1-2d}\,,$$

$$\gamma_1 \;=\; \gamma_3(r) \;=\; \frac{d}{1-d} \qquad \text{if and only if} \qquad r \;=\; r_{1,3} \;:=\; \alpha(1-d)\,,$$

$$\gamma_2(r) \;=\; \gamma_3(r) \;=\; \frac{\alpha(1-d)-1}{\alpha(1-d)+1} \qquad \text{if and only if} \qquad r \;=\; r_{2,3} \;:=\; \frac{1}{2}\left(\frac{1}{1-d} \;+\; \alpha\right)\,.$$

We next need to calculate $\kappa(\gamma(r), r)$ at $r \in \{r_{1,2}, r_{1,3}, r_{2,3}\}$, which requires computing $\gamma(r) = \gamma_1 \wedge \gamma_2(r) \wedge \gamma_3(r)$ at these points. To determine whether $\gamma_2(r_{2,3})$ lies above or below γ_1, we examine the sign of the difference:

$$\operatorname{sgn}\left(\gamma_2(r_{2,3}) \;-\; \frac{d}{1-d}\right) \;=\; \operatorname{sgn}\left(\alpha(1-d)(1-2d) \;-\; 1\right)\,.$$

Thus,

$$\gamma_2(r_{2,3}) \;>\; \frac{d}{1-d} \;=\; \gamma_1 \qquad \text{if and only if} \qquad \alpha \;>\; \frac{1}{(1-d)(1-2d)}\,.$$

We now distinguish between the two cases:

- $\alpha(1-d)(1-2d) > 1$
- $\alpha(1-d)(1-2d) \leq 1$

In the first case, $\alpha(1-d)(1-2d) > 1$, the ordering of the intersection points is

$$r_{1,2} \;<\; r_{2,3} \;<\; r_{1,3}\,,$$

and we therefore have $\gamma(r) = d/(1-d)$ for all $r \in \{r_{1,2}, r_{2,3}, r_{1,3}\}$ due to the monotonicity properties of γ_2 and γ_3, see the left panel of Figure 4. Since κ is decreasing in both γ and r,

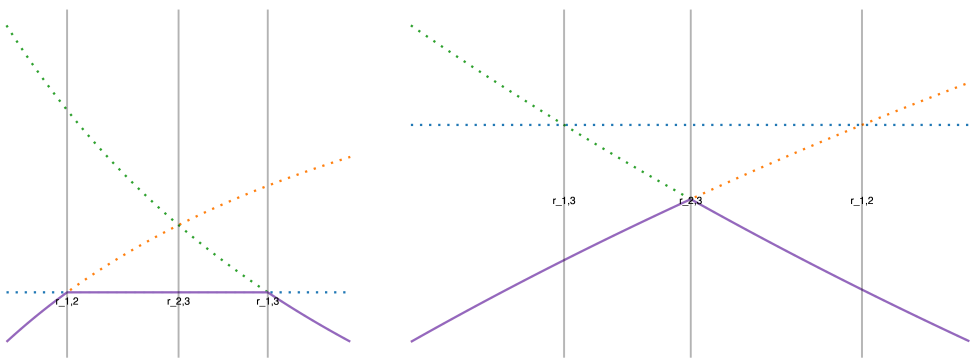

Figure 4: General position of the curves representing the functions γ_1 (blue dotted line), γ_2 (orange dotted line) and γ_3 (green dotted line) in the cases $\alpha(1-d)(1-2d) > 1$ (left panel) and $\alpha(1-d)(1-2d) \leq 1$ (right panel). The purple curve represents the function $\gamma_1 \wedge \gamma_2 \wedge \gamma_3$.

the optimal value is attained at the largest feasible r, which in this case is $r_{1,3}$, provided it lies within the admissible interval. Note that the inequality chain

$$\frac{1}{1-d} \;<\; \alpha(1-2d) \;<\; \alpha(1-d) \;=\; r_{1,3} \;<\; \alpha\,,$$

confirms that $r_{1,3}$ lies within the range $(1/(1-d), \alpha)$, and is therefore an admissible choice for minimizing κ over K. We conclude that, when $\alpha(1-d)(1-2d) > 1$, the optimal choice is $(\gamma_0, r_0) = (d/(1-d), \alpha(1-d))$. If on the contrary $\alpha(1-d)(1-2d) \leq 1$, we have the ordering

$$r_{1,3} \;\leq\; r_{2,3} \;\leq\; r_{1,2} \qquad \text{and} \qquad \gamma(r_{1,2}), \gamma(r_{1,3}) \;\leq\; \gamma(r_{2,3})$$

as the right panel of Figure 4 indicates. For $r \in [r_{2,3}, r_{1,2}]$, we have $\gamma(r) = \gamma_3(r) = \alpha/r - 1$, and thus

$$\kappa(\gamma(r), r) \;=\; \kappa\left(\frac{\alpha}{r} - 1, r\right) \;=\; 1 \;+\; \frac{1}{r}(1 - \alpha(1-d))\,.$$

Since $\alpha(1-d) > 1$ by assumption, this function is increasing in r on the interval $[r_{2,3}, r_{1,2}]$, so $\kappa(\gamma(r), r)$ is minimal at $r = r_{2,3}$ on this interval. As both $r_{1,3} \leq r_{2,3}$ and $\gamma(r_{1,3}) \leq \gamma(r_{2,3})$ (see again the right panel of Figure 4), and since κ is decreasing in both γ and r, the minimum of $\kappa(\gamma(r), r)$ over the full interval $[r_{1,3}, r_{2,3}]$ is attained at $r = r_{2,3}$. Finally, given that $r_{2,3}$ lies in the center of the interval $[1/(1-d), \alpha]$, it is therefore an admissible choice for minimizing κ over K. Therefore, the optimal choice in the case $\alpha(1-d)(1-2d) \leq 1$ is

$$(\gamma_0, r_0) = \left(\frac{\alpha(1-d) - 1}{\alpha(1-d) + 1}, \frac{1}{2}\left(\frac{1}{1-d} \;+\; \alpha\right)\right)\,.$$

Plugging the values of γ_0 and r_0 into $\kappa(\gamma, r)$ finishes the proof of (i).
Proof of Part (ii): Note that we already showed that $r_0 \in (1/(1-d), \alpha)$, and the statement about γ_0 follows from $\gamma_0 = \gamma_1(r_0) \wedge \gamma_2(r_0) \wedge \gamma_3(r_0)$. From what we showed it also follows that γ_0 lies on the upper bound of the admissible interval for γ. It remains to show $\kappa_0 - (d + 1/\alpha) < 0$. In the case $\alpha(1-d)(1-2d) > 1$ this follows from

$$\frac{1}{\alpha(1-d)} \;-\; d \;-\; \frac{1}{\alpha} \;=\; \frac{1 - (1-d)(\alpha d + 1)}{\alpha(1-d)} \;=\; \frac{d(1 - \alpha(1-d))}{\alpha(1-d)} \;<\; 0\,,$$

since $\alpha(1-d) > 1$. In the opposite case $\alpha(1-d)(1-2d) \leq 1$, it holds that

$$\begin{aligned}
& d \;+\; (1-d)\frac{3 - \alpha(1-d)}{\alpha(1-d) + 1} \;-\; d \;-\; \frac{1}{\alpha} \\
&= \frac{1}{\alpha(\alpha(1-d) + 1)}\,(\alpha(1-d)(3 - \alpha(1-d)) \;-\; \alpha(1-d) - 1) \\
&= \frac{1}{\alpha(\alpha(1-d) + 1)}\,(\alpha(1-d)(2 - \alpha(1-d)) \;-\; 1)\,.
\end{aligned}$$

By $\alpha(1-d)(1-2d) \leq 1$ and $\alpha(1-d) \leq \alpha \leq 2$ this is bounded above by

$$\frac{1}{\alpha(\alpha(1-d) + 1)}\left(\frac{2 - \alpha(1-d)}{1 - 2d} \;-\; 1\right),$$

which in turn by $\alpha(1-d) > 1$ is bounded above by

$$\frac{1}{\alpha(\alpha(1-d) + 1)}\left(\frac{1}{1 - 2d} \;-\; 1\right) \;<\; 0\,.$$

It follows (ii). $\square$

B.5 Proof of Corollaries 3.8 and 3.12

Proof of Corollary 3.8. We want to apply Corollary 3.6 for $G_n(x) = \mathbb{1}\{x > u_n\}$ and $G_n(x) = \log(x/u_n)\mathbb{1}\{x > u_n\}$, respectively. Note that both choices of G_n satisfy Assumption 2 for all $\gamma_G > 0$ arbitrarily small. By Lemma A.4 it holds $\mathbf{E}[G_n(X_0)] = \mathbf{P}[X_0 > u_n]$ and $G'_{\infty,n}(0) = f_{X_0}(u_n)$ for $G_n(x) = \mathbb{1}\{x > u_n\}$. Besides, for $G_n(x) = \log(x/u_n)\mathbb{1}\{x > u_n\}$, one has

$$\mathbf{E}[G_n(X_0)] \;=\; \int_{u_n}^{\infty} \log(x/u_n) f_{X_0}(x)\,\mathrm{d}x \;\sim\; \frac{\mathbf{P}[X_0 > u_n]}{\nu}$$

by the regular variation property of f_{X_0} (see *e.g.* [12, Equation (3.2.1)], and

$$G'_{\infty,n}(0) \;=\; -\int_{u_n}^{\infty} \log(x/u_n) f'_{X_0}(x)\,\mathrm{d}x \;=\; \int_{u_n}^{\infty} \frac{f_{X_0}(x)}{x}\,\mathrm{d}x \;\sim\; \frac{f_{X_0}(u_n)}{\nu+1}\,,$$

by partial integration and Karamata's theorem [12, Theorem B.1.5]. Therefore it holds for $G_n(x) = \mathbb{1}\{x > u_n\}$ that

$$\frac{\mathbf{E}[G_n(X_0)]}{G'_{\infty,n}(0)} \;=\; \frac{\mathbf{P}[X_0 > u_n]}{f_{X_0}(u_n)} \;\sim\; \frac{u_n}{\nu}\,,$$

by Karamata's theorem again, and for $G_n(x) = \log(x/u_n)\mathbb{1}\{x > u_n\}$ that

$$\frac{\mathbf{E}[G_n(X_0)]}{G'_{\infty,n}(0)} \sim \frac{\mathbf{P}[X_0 > u_n]}{f_{X_0}(u_n)}\frac{\nu+1}{\nu} \;\sim\; u_n\frac{1}{\nu}\frac{\nu+1}{\nu}\,.$$

We finally check that $n^{\kappa_0+\delta/2-1/(\nu\wedge 2)-d}/G'_{\infty,n}(0) \to 0$. For this it suffices to consider

$$\frac{n^{\kappa_0+\delta/2-1/(\nu\wedge 2)-d}}{f_{X_0}(u_n)} \;\sim\; L(u_n) n^{\kappa_0+\delta/2-1/(\nu\wedge 2)-d} u_n^{\nu+1}\,,$$

where $L : x \mapsto x^{-\nu-1}/f_{X_0}(x)$ is slowly varying. Note that the term in the last display converges to zero since $u_n^{\nu+1} = o(n^{-\kappa_0-\delta+d+1/(\nu\wedge 2)})$ and $L(u_n) = o(n^{\delta/2})$. Therefore, we may apply Corollary 3.6 to yield the desired results. □

For the proof of Corollary 3.12 we require the following auxiliary result.

Lemma B.7. *If $\omega \in C^1(\mathbb{R})$ is positive and satisfies, for some $\beta > 0$, that $x^{1-\beta}\omega'(x)/\omega(x)$ converges to 0 as $x \to \infty$, then for all $p \in \mathbb{R}$ it holds that $x^p\omega(x)\exp(-x^\beta/\beta) \to 0$ as $x \to \infty$.*

Proof. Since $x^{1-\beta}\omega'(x)/\omega(x) \to 0$ as $x \to \infty$, there exists $x_0 > 0$ with $\omega'(x) \le \frac{1}{2}x^{\beta-1}\omega(x)$ for all $x > x_0$. By Grönwall's lemma it holds, for $x > x_0$,

$$\omega(x) \;\le\; \omega(x_0)\exp(-x_0^\beta/(2\beta))\exp(x^\beta/(2\beta))\,.$$

Therefore, it holds that for any $p \in \mathbb{R}$, as $x \to \infty$,

$$x^p\omega(x)\exp(-x^\beta/\beta) \;\lesssim\; x^p\exp(-x^\beta/(2\beta)) \;\to\; 0\,,$$

which immediately yields the result since ω is assumed to be non-negative. □

Proof of Corollary 3.12. We proceed as in the proof of Corollary 3.8. There we checked that Assumption 2 holds. We next establish the limit

$$\lim_{n\to\infty} \frac{u_n^{1-\beta} f_{X_0}(u_n)}{\mathbf{P}[X_0 > u_n]} = 1. \tag{57}$$

Let $\phi : x \mapsto x^{1-\beta} f_{X_0}(x)$ and $\psi : x \mapsto \mathbf{P}[X_0 > x]$. The functions ϕ and ψ are differentiable in a neighborhood of infinity, converge to 0 at infinity, and $\psi'(x) = -f_{X_0}(x) < 0$ for x large enough. Moreover $x^{1-\beta}\omega'(x)/\omega(x) \to 0$ as $x \to \infty$ since $\beta > 0$ and ω' is either identically zero far from the origin or regularly varying at infinity, so that

$$\lim_{x\to\infty} \frac{\phi'(x)}{\psi'(x)} = \lim_{x\to\infty} -\frac{(1-\beta)x^{-\beta} f_{X_0}(x) + x^{1-\beta} f'_{X_0}(x)}{f_{X_0}(x)} = \lim_{x\to\infty} \left(1 - x^{1-\beta}\frac{\omega'(x)}{\omega(x)}\right) = 1.$$

Then (57) follows from l'Hôpital's rule, that is,

$$\lim_{n\to\infty} \frac{u_n^{1-\beta} f_{X_0}(u_n)}{\mathbf{P}[X_0 > u_n]} = \lim_{n\to\infty} \frac{\phi(u_n)}{\psi(u_n)} = \lim_{n\to\infty} \frac{\phi'(u_n)}{\psi'(u_n)} = 1.$$

Therefore $\mathbf{P}[X_0 > u_n] \sim u_n^{1-\beta} f_{X_0}(u_n)$. Next, note that by partial integration it holds

$$\begin{aligned}\mathbf{E}[G_n(X_0)] &= \int_{u_n}^{\infty} \log(x/u_n) f_{X_0}(x)\,\mathrm{d}x \\ &= -\lim_{x\to\infty} \log(x/u_n)\mathbf{P}[X_0 > x] \;+\; \int_{u_n}^{\infty} \frac{\mathbf{P}[X_0 > x]}{x}\,\mathrm{d}x \;=\; \int_{u_n}^{\infty} \frac{\mathbf{P}[X_0 > x]}{x}\,\mathrm{d}x\end{aligned}$$

since the boundary term vanishes due to Lemma B.7. Similarly we get

$$G'_{\infty,n}(0) = -\int_{u_n}^{\infty} \log(x/u_n) f'_{X_0}(x)\,\mathrm{d}x \;=\; \int_{u_n}^{\infty} \frac{f_{X_0}(x)}{x}\,\mathrm{d}x\,.$$

Note that

$$\lim_{n\to\infty} \frac{u_n^{-\beta}\mathbf{P}[X_0 > u_n]}{\int_{u_n}^{\infty} \frac{\mathbf{P}[X_0 > x]}{x}\,\mathrm{d}x} \;=\; \lim_{n\to\infty} \frac{\beta u_n^{-\beta-1}\mathbf{P}[X_0 > u_n] + u_n^{-\beta} f_{X_0}(u_n)}{\mathbf{P}[X_0 > u_n]/u_n} \;=\; 1\,,$$

and

$$\lim_{n\to\infty} \frac{u_n^{-1}\mathbf{P}[X_0 > u_n]}{\int_{u_n}^{\infty} \frac{f_{X_0}(x)}{x}\,\mathrm{d}x} \;=\; \lim_{n\to\infty} \frac{u_n^{-2}\mathbf{P}[X_0 > u_n] + u_n^{-1} f_{X_0}(u_n)}{f_{X_0}(u_n)/u_n} \;=\; 1\,,$$

where we used twice that $\mathbf{P}[X_0 > u_n] \sim u_n^{1-\beta} f_{X_0}(u_n)$ and l'Hôpital's rule in a way similar to what was done above. This yields $\mathbf{E}[G_n(X_0)]/G'_{\infty,n}(0) \sim u_n^{1-\beta}$. Finally we check that $n^{\kappa_0+\delta/2-d-1/2}/G'_{\infty,n}(0) \to 0$. Since $G'_{\infty,n}(0) \sim u_n^{-\beta} f_{X_0}(u_n)$, this boils down to

$$\frac{n^{\kappa_0+\delta/2-d-1/2}}{u_n^{-\beta}\omega(u_n)\exp(-u_n^{\beta}/\beta)} \;=\; \frac{n^{-\delta/2}}{u_n^{-\beta}\omega(u_n)} \;\to\; 0 \tag{58}$$

due to the regular variation property of ω. Then, as before, applying Corollary 3.6 yields the results. □

C Derandomization device

We state and prove a lemma which is the main tool for obtaining convergence of PoT estimators with random thresholds (Corollaries 3.10 and 3.15) from their convergence with deterministic thresholds (Corollaries 3.8 and 3.12).

Lemma C.1. *Let (X_t) be a stationary time series, and let g be an increasing, continuously differentiable function whose derivative g' is regularly varying at infinity. For $x > 0$ define*

$$\xi(x) \;:=\; \frac{x f_{X_0}(x)}{\mathbf{P}[X_0 > x]}\,.$$

Assume the following:

- *There exists a threshold sequence (u_n) with $u_n \to \infty$, and a rate sequence (r_n) with $r_n \to \infty$, such that $r_n/\xi(u_n) \to \infty$, and there exists a constant $c > 0$ such that*

$$u_n \cdot g'(u_n) \frac{\mathbf{P}[X_0 > u_n]}{\mathbf{E}\left[\, (g(X_0) - g(u_n))\, \mathbb{1}\left\{X_0 > u_n\right\}\right]} \;\sim\; c \cdot \xi(u_n)\,, \tag{59}$$

  and, for all $t \in \mathbb{R}$,

$$r_n \left(\frac{\mathbf{P}[X_0 > u_n]}{\mathbf{P}[X_0 > (1 + t/r_n) \cdot u_n]} \;-\; 1 \right) \;\sim\; t \cdot \xi(u_n)\,. \tag{60}$$

- *There exists $\delta_\epsilon \in (0, 1)$ such that for all sequences (ϵ_n) converging to 0 and satisfying*

$$|\epsilon_n| = \mathcal{O}(r_n^{\delta_\epsilon - 1}) \tag{61}$$

  the following holds:

$$\frac{r_n/\xi(u_n)}{n} \cdot \sum_{t=1}^{n} \left[\left(\frac{(g(X_t) - g(u_n))\, \mathbb{1}\left\{X_t > u_n\right\}}{\mathbf{E}\left[\, (g(X_0) - g(u_n))\, \mathbb{1}\left\{X_0 > u_n\right\}\right]} \;-\; 1 \right), \left(\frac{\mathbb{1}\left\{X_t > u_n(1+\epsilon_n)\right\}}{\mathbf{P}[X_0 > u_n(1+\epsilon_n)]} \;-\; 1 \right) \right] \tag{62}$$

  converges in distribution, as $n \to \infty$, to a pair of non-degenerate continuous random variables $[\Phi, \Psi]$.

- *There exists a positive sequence (τ_n) such that*

$$\tau_n \;=\; \mathcal{O}(r_n^{\delta_\epsilon - 1})\,, \tag{63}$$

$$\tau_n r_n \to \infty\,, \tag{64}$$

  and

$$\frac{\mathbf{P}[X_0 > u_n]}{\mathbf{P}[X_0 > u_n(1 \pm \tau_n)]} \;\to\; 1\,, \tag{65}$$

  as $n \to \infty$.

If additionally $n\mathbf{P}[X_0 > u_n] \to \infty$*, then under the above assumptions, for*

$$k \;=\; k(n, u_n) \;:=\; \lfloor n\mathbf{P}[X_0 > u_n] \rfloor \,,$$

it holds that

$$\left[\frac{r_n}{\xi(u_n)} \frac{1}{k} \sum_{i=1}^{k} \left(\frac{g(X_{n-i+1:n}) \;-\; g(X_{n-k:n})}{\mathbf{E}\left[(\, g(X_0) - g(u_n))\, \mathbb{1}\left\{ X_0 > u_n \right\} \right]} \frac{k}{n} \;-\; 1 \right) \,,\; r_n \left(\frac{X_{n-k:n}}{u_n} \;-\; 1 \right) \right]$$

converges in distribution to $[\Phi \;-\; c \cdot \Psi \,, \Psi]$ *as* $n \to \infty$*.*

To prove this result, we need the following lemma, which establishes a connection between random thresholds and their deterministic counterparts.

Lemma C.2. (Koenker-type inversion lemma) *Let* $k(u_n, n) := \lfloor n\mathbf{P}[X_0 > u_n] \rfloor$*, and assume that Conditions* (60) *and* (62) *from Lemma C.1 hold. Then:*

(i) The inequality

$$r_n \left(\frac{X_{n-k:n}}{u_n} \;-\; 1 \right) \;\leq\; s$$

is equivalent to

$$r_n \left(\frac{1}{n} \sum_{t=1}^{n} \frac{\mathbb{1}\left\{ X_t > u_n(1 + s/r_n) \right\}}{\mathbf{P}[X_0 > u_n(1 + s/r_n)]} \;-\; 1 \right) \;\leq\; r_n \left(\frac{\mathbf{P}[X_0 > u_n]}{\mathbf{P}[X_0 > u_n(1 + s/r_n)]} \;-\; 1 \right) .$$

(ii) As $n \to \infty$*, it holds that*

$$r_n \left[\frac{1/\xi(u_n)}{n} \sum_{t=1}^{n} \left(\frac{(g(X_t) - g(u_n))\, \mathbb{1}\left\{ X_t > u_n \right\}}{\mathbf{E}\left[\, (g(X_0) - g(u_n))\, \mathbb{1}\left\{ X_0 > u_n \right\} \right]} \;-\; 1 \right) , \left(\frac{X_{n-k:n}}{u_n} \;-\; 1 \right) \right]$$
$$\xrightarrow{\mathrm{d}} \;[\Phi \,, \Psi].$$

Proof. **Proof of Part (i):** For any $x \in \mathbb{R}$, it holds that

$$X_{\lceil n\alpha \rceil : n} \;\leq\; x \qquad \text{if and only if} \qquad \alpha \;\leq\; \frac{1}{n} \sum_{t=1}^{n} \mathbb{1}\left\{ X_t \leq x \right\} .$$

This equivalence follows from a general property of quantile functions: for any distribution function G and its associated quantile function Q, we have $Q(\alpha) \leq x$ if and only if $\alpha \leq G(x)$, where Q is defined as the left-continuous inverse of G.

Proof of Part (ii): Let $s_1, s_2 \in \mathbb{R}$ and define the event

$$E_n(s_1) = \left\{ \frac{r_n/\xi(u_n)}{n} \sum_{t=1}^{n} \left(\frac{(g(X_t) - g(u_n))\, \mathbb{1}\left\{ X_t > u_n \right\}}{\mathbf{E}\left[\, (g(X_0) - g(u_n))\, \mathbb{1}\left\{ X_0 > u_n \right\} \right]} \;-\; 1 \right) \;\leq\; s_1 \right\}.$$

By Part *(i)*, together with Conditions (60) and (62) and the continuity of Ψ, we obtain

$$\begin{aligned}
&\mathbf{P}\left[E_n(s_1),\, r_n \left(\frac{X_{n-k:n}}{u_n} \;-\; 1 \right) \;\leq\; s_2 \right] \\
&= \mathbf{P}\left[E_n(s_1),\, r_n \left(\frac{1}{n} \sum_{t=1}^{n} \frac{\mathbb{1}\left\{ X_t > u_n(1 + s_2/r_n) \right\}}{\mathbf{P}[X_0 > u_n(1 + s_2/r_n)]} \;-\; 1 \right) \;\leq\; r_n \left(\frac{\mathbf{P}[X_0 > u_n]}{\mathbf{P}[X_0 > u_n(1 + s_2/r_n)]} \;-\; 1 \right) \right] \\
&= \mathbf{P}\left[E_n(s_1),\, r_n \left(\frac{1}{n} \sum_{t=1}^{n} \frac{\mathbb{1}\left\{ X_t > u_n(1 + s_2/r_n) \right\}}{\mathbf{P}[X_0 > u_n(1 + s_2/r_n)]} \;-\; 1 \right) \;\leq\; s_2 \cdot \xi(u_n)(1 + o(1)) \right] \\
&\to \mathbf{P}\left[\Phi \leq s_1 \,, \Psi \leq s_2 \right] ,
\end{aligned}$$

from which the convergence in distribution follows by Slutsky's lemma (note that the quantity $\epsilon_n := s_2/r_n$ indeed satisfies Condition (61)). □

Proof of Lemma C.1. Define

$$\widehat{e}_{g,n}(k, u_n) \ := \ \frac{1}{k}\sum_{t=1}^{n}(g(X_t) \ - \ g(u_n))\,\mathbb{1}\left\{X_t > u_n\right\}.$$

Note that

$$\widehat{e}_{g,n}(k, X_{n-k:n}) \ = \ \frac{1}{k}\sum_{t=1}^{n}(g(X_t) \ - \ g(X_{n-k:n}))\,\mathbb{1}\{X_t > X_{n-k:n}\},$$

and therefore

$$\widehat{e}_{g,n}(k, X_{n-k:n}) \ - \ \widehat{e}_{g,n}(k, u_n) \ = \ g(u_n) \ - \ g(X_{n-k:n}) \ + \ R_1 \ + \ R_2\,, \tag{66}$$

where

$$R_1 \ := \ (g(X_{n-k:n}) \ - \ g(u_n))\left(1 \ - \ \frac{1}{k}\sum_{t=1}^{n}\mathbb{1}\{X_t > X_{n-k:n}\}\right),$$

$$R_2 \ := \ \frac{1}{k}\sum_{t=1}^{n}(g(X_t) \ - \ g(u_n))\left(\mathbb{1}\left\{X_t > X_{n-k:n}\right\} - \mathbb{1}\left\{X_t > u_n\right\}\right).$$

We will show that $R_j = o_{\mathbf{P}}(g(u_n) - g(X_{n-k:n}))$ for $j = 1, 2$, that is,

$$\frac{R_j}{g(u_n) - g(X_{n-k:n})} \ \xrightarrow{\mathbf{P}} \ 0\,. \tag{67}$$

Combining (67) with (66), it will follow that

$$\widehat{e}_{g,n}(k, X_{n-k:n}) \ - \ \widehat{e}_{g,n}(k, u_n) \ = \ (g(u_n) \ - \ g(X_{n-k:n}))(1 \ + \ o_{\mathbf{P}}(1))\,. \tag{68}$$

We begin with the analysis of R_1. By Lemma C.2.(ii), we have

$$r_n\left(\frac{X_{n-k:n}}{u_n} \ - \ 1\right) \ \xrightarrow{\mathrm{d}} \ \Psi. \tag{69}$$

Let (τ_n) be as in Condition (63) and (64). Then, we obtain

$$\frac{1}{\tau_n}\left(\frac{X_{n-k:n}}{u_n} \ - \ 1\right) \ = \ \frac{1}{r_n\tau_n}r_n\left(\frac{X_{n-k:n}}{u_n} \ - \ 1\right) \ \xrightarrow{\mathbf{P}} \ 0\,, \tag{70}$$

since $r_n\tau_n \to \infty$ as $n \to \infty$. Consequently, with arbitrarily high probability for large n, we have

$$(1 \ - \ \tau_n)\,u_n \ \le \ X_{n-k:n} \ \le \ (1 \ + \ \tau_n)\,u_n\,. \tag{71}$$

This implies that

$$\begin{aligned}\left|\frac{R_1}{g(u_n) - g(X_{n-k:n})}\right| \ &= \ \left|\left(\frac{1}{k}\sum_{t=1}^{n}\mathbb{1}\{X_t > X_{n-k:n}\}\right) \ - \ 1\right| \\ \le \ &\left|\left(\frac{1}{k}\sum_{t=1}^{n}\mathbb{1}\left\{X_t > (1 \ - \ \tau_n)\,u_n\right\}\right) \ - \ 1\right| \ \vee \ \left|\left(\frac{1}{k}\sum_{t=1}^{n}\mathbb{1}\left\{X_t > (1 \ + \ \tau_n)\,u_n\right\}\right) \ - \ 1\right|,\end{aligned} \tag{72}$$

with arbitrarily high probability as $n \to \infty$. By Assumption (65) and definition $k = \lfloor n\mathbf{P}[X_0 > u_n] \rfloor$, we obtain

$$\begin{aligned}
&\left| \left(\frac{1}{k} \sum_{t=1}^{n} \mathbb{1}\left\{X_t > u_n(1 \pm \tau_n)\right\} \right) \; - \; 1 \right| \\
&\leq \; \frac{n\mathbf{P}[X_0 > u_n]}{\lfloor n\mathbf{P}[X_0 > u_n] \rfloor} \cdot \frac{\mathbf{P}[X_0 > u_n(1 \pm \tau_n)]}{\mathbf{P}[X_0 > u_n]} \cdot \frac{1}{r_n/\xi(u_n)} \cdot \left| \frac{r_n/\xi(u_n)}{n} \sum_{t=1}^{n} \left(\frac{\mathbb{1}\left\{X_t > u_n(1 \pm \tau_n)\right\}}{\mathbf{P}[X_0 > u_n(1 \pm \tau_n)]} \; - \; 1 \right) \right| \\
&\quad + \; \left| \frac{n\mathbf{P}[X_0 > u_n]}{\lfloor n\mathbf{P}[X_0 > u_n] \rfloor} \cdot \frac{\mathbf{P}[X_0 > u_n(1 \pm \tau_n)]}{\mathbf{P}[X_0 > u_n]} \; - \; 1 \right| \\
&\xrightarrow{\mathbf{P}} \; 0 \,,
\end{aligned} \tag{73}$$

as $n \to \infty$, since each term on the right-hand side converges to zero in probability due to

$$\frac{n\mathbf{P}[X_0 > u_n]}{\lfloor n\mathbf{P}[X_0 > u_n] \rfloor} \,, \; \frac{\mathbf{P}[X_0 > u_n(1 \pm \tau_n)]}{\mathbf{P}[X_0 > u_n]} \; \to \; 1 \qquad \text{and} \qquad \frac{1}{r_n/\xi(u_n)} \; \to \; 0 \,,$$

by $n\mathbf{P}[X_0 > u_n] \to \infty$, Assumption (65) and $r_n/\xi(u_n) \to \infty$, and

$$\left| \frac{r_n/\xi(u_n)}{n} \sum_{t=1}^{n} \left(\frac{\mathbb{1}\left\{X_t > u_n(1 \pm \tau_n)\right\}}{\mathbf{P}[X_0 > u_n(1 \pm \tau_n)]} \; - \; 1 \right) \right| \; \xrightarrow{\mathrm{d}} \; |\Psi| \,,$$

which is guaranteed by (62) and the fact that (τ_n) also satisfies (63). Using (72) and (73) we get (67) for $j = 1$.

We now analyze R_2. Since g is increasing, we can write

$$\begin{aligned}
&|g(u_n) \; - \; g(X_t)| \left| \mathbb{1}\left\{X_t > X_{n-k:n}\right\} - \mathbb{1}\left\{X_t > u_n\right\} \right| \\
&= \; (g(u_n) \; - \; g(X_t))\, \mathbb{1}\{u_n > X_t > X_{n-k:n}\} \\
&\quad + \; (g(X_t) \; - \; g(u_n))\, \mathbb{1}\{X_{n-k:n} \geq X_t > u_n\} \,.
\end{aligned}$$

By monotonicity of g, we may further bound this as

$$\begin{aligned}
&\leq \; (g(u_n) \; - \; g(X_{n-k:n}))\, \mathbb{1}\{u_n > X_t > X_{n-k:n}\} \\
&\quad + \; (g(X_{n-k:n}) \; - \; g(u_n))\, \mathbb{1}\{X_{n-k:n} > X_t > u_n\} \\
&= \; (g(X_{n-k:n}) \; - \; g(u_n))\left(\mathbb{1}\{X_{n-k:n} > X_t > u_n\} \; - \; \mathbb{1}\{u_n > X_t > X_{n-k:n}\}\right) \,.
\end{aligned}$$

It follows, rewriting R_2 and carefully considering the sign in the absolute value, that

$$\begin{aligned}
|R_2| \; &= \; \left| \frac{1}{k} \sum_{t=1}^{n} (g(X_t) \; - \; g(u_n))\left(\mathbb{1}\left\{X_t > X_{n-k:n}\right\} - \mathbb{1}\left\{X_t > u_n\right\}\right) \right| \\
&\leq \; (g(X_{n-k:n}) \; - \; g(u_n)) \cdot \frac{1}{k} \sum_{t=1}^{n} \left(\mathbb{1}\{X_{n-k:n} > X_t > u_n\} - \mathbb{1}\{u_n > X_t > X_{n-k:n}\}\right) \\
&\leq \; (g(X_{n-k:n}) \; - \; g(u_n)) \cdot \frac{1}{k} \sum_{t=1}^{n} \Big(\mathbb{1}\left\{X_t > u_n\right\} \; - \; \mathbb{1}\left\{X_t > X_{n-k:n}\right\} \Big) \\
&\leq \; |g(u_n) \; - \; g(X_{n-k:n})| \cdot \left| \frac{1}{k} \sum_{t=1}^{n} \Big(\mathbb{1}\left\{X_t > X_{n-k:n}\right\} \; - \; \mathbb{1}\left\{X_t > u_n\right\} \Big) \right| \,.
\end{aligned} \tag{74}$$

We now aim to show that

$$\left|\frac{1}{k}\sum_{t=1}^{n}\Big(\mathbb{1}\{X_t > X_{n-k:n}\} \;-\; \mathbb{1}\{X_t > u_n\}\Big)\right| \;\xrightarrow{\mathbf{P}}\; 0\,, \tag{75}$$

which, when combined with (74), establishes (67) for $j = 2$. To prove this convergence, note that by (71), with arbitrarily high probability as $n \to \infty$, the following holds:

$$\begin{aligned}
&\left|\frac{1}{k}\sum_{t=1}^{n}\Big(\mathbb{1}\{X_t > X_{n-k:n}\} \;-\; \mathbb{1}\{X_t > u_n\}\Big)\right| \\
&\le \left|\left(\frac{1}{k}\sum_{t=1}^{n}\mathbb{1}\{X_t > X_{n-k:n}\}\right) - 1\right| + \left|\left(\frac{1}{k}\sum_{t=1}^{n}\mathbb{1}\{X_t > u_n\}\right) - 1\right| \\
&\le \left|\left(\frac{1}{k}\sum_{t=1}^{n}\mathbb{1}\{X_t > (1 - \tau_n)\, u_n\}\right) - 1\right| \vee \left|\left(\frac{1}{k}\sum_{t=1}^{n}\mathbb{1}\{X_t > (1 + \tau_n)\, u_n\}\right) - 1\right| \\
&\quad + \left|\left(\frac{1}{k}\sum_{t=1}^{n}\mathbb{1}\{X_t > u_n\}\right) - 1\right|.
\end{aligned} \tag{76}$$

Applying Inequalities (73) and (76), we obtain the desired convergence in (75), and hence conclude (67), and therefore (68). We now show that

$$g(X_{n-k:n}) \;-\; g(u_n) \;=\; u_n g'(u_n)\left(\frac{X_{n-k:n}}{u_n} \;-\; 1\right) \;+\; o_{\mathbf{P}}\left(\frac{u_n \cdot g'(u_n)}{r_n}\right), \tag{77}$$

and, using this together with (68), (69), and (77), we will conclude that

$$\widehat{e}_{g,n}(k, X_{n-k:n}) \;=\; \widehat{e}_{g,n}(k, u_n) \;-\; u_n \cdot g'(u_n)\left(\frac{X_{n-k:n}}{u_n} \;-\; 1\right) \;+\; o_{\mathbf{P}}\left(\frac{u_n \cdot g'(u_n)}{r_n}\right). \tag{78}$$

To establish Equation (77), observe that

$$\begin{aligned}
&g(X_{n-k:n}) \;-\; g(u_n) \;-\; u_n g'(u_n)\left(\frac{X_{n-k:n}}{u_n} \;-\; 1\right) \\
&\quad = \; u_n g'(u_n)\int_1^{X_{n-k:n}/u_n}\left(\frac{g'(xu_n)}{g'(u_n)} \;-\; 1\right)\mathrm{d}x\,.
\end{aligned} \tag{79}$$

Since g' is regularly varying and $X_{n-k:n}/u_n \xrightarrow{\mathbf{P}} 1$ by (70), it follows from [12, Proposition B.1.10] that

$$\int_1^{X_{n-k:n}/u_n}\left(\frac{g'(xu_n)}{g'(u_n)} \;-\; 1\right)\mathrm{d}x \;=\; o_{\mathbf{P}}\left(\left|\frac{X_{n-k:n}}{u_n} \;-\; 1\right|\right) \;=\; o_{\mathbf{P}}\left(\frac{1}{r_n}\right),$$

where the final equality follows from (69). Combining this with (79), we obtain the desired result in (77). We now prove the stated result using Equation (78). To this end, note that by the stationarity of the time series, $k/n \sim \mathbf{P}[X_0 > u_n]$, and Assumption (59) it follows that

$$\begin{aligned}
\mathbf{E}[\widehat{e}_{g,n}(k, u_n)] \;&=\; \frac{n}{k}\mathbf{E}[(g(X_0) - g(u_n))\,\mathbb{1}\{X_0 > u_n\}] \\
&\sim\; \frac{\mathbf{E}[(g(X_0) - g(u_n))\,\mathbb{1}\{X_0 > u_n\}]}{\mathbf{P}[X_0 > u_n]} \\
&\sim\; \frac{u_n g'(u_n)}{c \cdot \xi(u_n)}\,.
\end{aligned} \tag{80}$$

Consequently, we can write

$$\begin{aligned}
&\frac{r_n}{\xi(u_n)}\left(\frac{\widehat{e}_{g,n}(k,u_n)}{\mathbf{E}[\widehat{e}_{g,n}(k,u_n)]}\ -\ 1\right)\\
&\quad=\ \frac{r_n/\xi(u_n)}{n}\sum_{t=1}^{n}\left(\frac{(g(X_t)-g(u_n))\,\mathbb{1}\{X_t>u_n\}}{\mathbf{E}[(g(X_0)-g(u_n))\,\mathbb{1}\{X_0>u_n\}]}\ -\ 1\right).
\end{aligned}$$

Substituting the approximation for $\widehat{e}_{g,n}(k,X_{n-k:n})$ from (78) and combining with the joint convergence established in Lemma C.2.(ii), we conclude

$$\begin{aligned}
&\frac{r_n}{\xi(u_n)}\left(\frac{\widehat{e}_{g,n}(k,X_{n-k:n})}{\mathbf{E}[\widehat{e}_{g,n}(k,u_n)]}\ -\ 1\right)\\
&\quad=\ \frac{r_n}{\xi(u_n)}\left(\frac{\widehat{e}_{g,n}(k,u_n)}{\mathbf{E}[\widehat{e}_{g,n}(k,u_n)]}\ -\ 1\right)\ -\ c\cdot r_n\left(\frac{X_{n-k:n}}{u_n}\ -\ 1\right)\ +\ o_{\mathbf{P}}(1)\,.
\end{aligned}$$

Therefore, by Lemma C.2.(ii), it follows that

$$\left[\frac{r_n}{\xi(u_n)}\left(\frac{\widehat{e}_{g,n}(k,X_{n-k:n})}{\mathbf{E}\left[\widehat{e}_{g,n}(k,u_n)\right]}\ -\ 1\right),\ r_n\left(\frac{X_{n-k:n}}{u_n}\ -\ 1\right)\right]\ \xrightarrow{\mathrm{d}}\ [\Phi\ -\ c\cdot\Psi,\Psi]\,.$$

To complete the proof, we use Equation (80) and the fact that

$$\widehat{e}_{g,n}(k,X_{n-k:n})=\frac{1}{k}\sum_{i=1}^{k}g(X_{n-i+1:n})-g(X_{n-k:n})$$

to express

$$\frac{\widehat{e}_{g,n}(k,X_{n-k:n})}{\mathbf{E}\left[\widehat{e}_{g,n}(k,u_n)\right]}\ -\ 1\ =\ \frac{1}{k}\sum_{i=1}^{k}\left(\frac{g(X_{n-i+1:n})-g(X_{n-k:n})}{\mathbf{E}[(g(X_0)-g(u_n))\mathbb{1}\{X_0>u_n\}]}\frac{k}{n}\ -\ 1\right)$$

from which the stated convergence result follows immediately. □

D Verifying the conditions of Lemma C.1 in the settings of Corollary 3.8 and 3.12

In the next lemma we check the assumptions of Lemma C.1 in the settings of Corollary 3.8 and Corollary 3.12. Applying Lemma C.1 then directly results in Corollary 3.10 and 3.15.

Lemma D.1. *Let $g=\log$. Then the assumptions of Lemma C.1 hold in the settings of (i) Corollary 3.8 and (ii) Corollary 3.12, respectively, with the following parameters:*

(i) In Corollary 3.8, $r_n=n^{1-1/(\nu\wedge 2)-d}u_n$,

$$\xi(u_n)\ =\ \frac{u_n f_{X_0}(u_n)}{\mathbf{P}[X_0>u_n]}\ \to\ \nu\,,$$

and the joint limit is

$$(\Phi,\Psi)\ =\ \left(\frac{\nu}{1+\nu}Z_{2\wedge\nu},\ Z_{2\wedge\nu}\right),$$

with constant $c=1$ and (τ_n) any sequence satisfying (63) and (64).

(ii) In Corollary 3.8, $r_n = n^{1/2-d}u_n$*,*

$$\xi(u_n) \;=\; \frac{u_n f_{X_0}(u_n)}{\mathbf{P}[X_0 > u_n]} \;\sim\; u_n^{\beta}\,,$$

and the joint limit is $(\Phi, \Psi) = (Z_2, Z_2)$*, with constant* $c = 1$ *and* (τ_n) *any sequence satisfying* (63) *and* (64)*.*

Proof of Lemma D.1. We begin by showing all the assumptions of Lemma C.1 except for the joint convergence Assumption (62). We do this separately for both settings.

In Setting (i), it follows from Karamata's theorem [12, Theorem B.1.5], that

$$\xi(u_n) \;=\; \frac{u_n f_{X_0}(u_n)}{\mathbf{P}[X_0 > u_n]} \;\to\; \nu\,.$$

This implies $r_n/\xi(u_n) \to \infty$. Note now that, by Karamata's theorem, we have

$$\begin{aligned}
\mathbf{E}\left[\log{(X_0/u_n)_+}\right] \;&=\; u_n f_{X_0}(u_n) \int_1^\infty \log(x) \frac{f_{X_0}(u_n x)}{f_{X_0}(u_n)}\,\mathrm{d}x \\
&\sim\; \nu \mathbf{P}[X_0 > u_n] \int_1^\infty \log(x) \frac{1}{x^{1+\nu}}\,\mathrm{d}x \\
&=\; \frac{\mathbf{P}[X_0 > u_n]}{\nu}\,.
\end{aligned} \tag{81}$$

Since $g'(u_n) = 1/u_n$, and using Equation (81), we find that

$$u_n g'(u_n) \frac{\mathbf{P}[X_0 > u_n]}{\mathbf{E}[(g(X_0) - g(u_n))\mathbb{1}\{X_0 > u_n\}]} = \frac{\mathbf{P}[X_0 > u_n]}{\mathbf{E}\left[\log{(X_0/u_n)_+}\right]} \;\sim\; \nu \;\sim\; \xi(u_n)\,,$$

which shows that Assumption (59) holds with constant $c = 1$. Applying the mean value theorem and the regular variation property of f_{X_0}, along with the local uniformity property of regularly varying functions, we conclude that for each $t \in \mathbb{R}$, there exists a sequence $(\theta_n) = (\theta_n(t)) \subset (0, 1)$ such that

$$\begin{aligned}
& r_n \left(\frac{\mathbf{P}[X_0 > u_n]}{\mathbf{P}[X_0 > (1 + t/r_n) \cdot u_n]} \;-\; 1 \right) \\
&=\; \frac{r_n}{\mathbf{P}[X_0 > (1 + t/r_n)u_n]} \frac{u_n t}{r_n} f_{X_0}((1 + \theta_n t/r_n)u_n) \\
&=\; t \cdot \frac{u_n}{u_n(1 + t/r_n)} \frac{u_n(1 + t/r_n) \cdot f_{X_0}(u_n(1 + t/r_n))}{\mathbf{P}[X_0 > u_n(1 + t/r_n)]} \cdot \frac{f_{X_0}(u_n(1 + \theta_n t/r_n))}{f_{X_0}(u_n(1 + t/r_n))} \\
&\sim\; t \cdot \xi(u_n(1 + t/r_n)) \\
&\sim\; t \cdot \xi(u_n)\,,
\end{aligned}$$

where the last approximation uses that $\xi(x) \to \nu$ as $x \to \infty$, as shown above. Thus, Assumption (60) is satisfied. Moreover, by the local uniformity of regularly varying functions, we have that for any sequence (τ_n) with $\tau_n \to 0$ as $n \to \infty$,

$$\frac{\mathbf{P}[X_0 > u_n]}{\mathbf{P}[X_0 > u_n(1 \pm \tau_n)]} \;\to\; 1\,.$$

Therefore, Assumption (65) is also satisfied.

Now we turn to Setting (ii). We use again

$$\xi(u_n) \;=\; \frac{u_n f_{X_0}(u_n)}{\mathbf{P}[X_0 > u_n]} \;\sim\; u_n^{1-(1-\beta)} = u_n^{\beta}$$

such that $\xi(u_n)/r_n \sim u_n^{\beta}/r_n \to 0$, since u_n grows only logarithmically while r_n grows polynomially in n. As shown in the proof of Corollary 3.12, we have

$$u_n g'(u_n) \frac{\mathbf{P}[X_0 > u_n]}{\mathbf{E}[(g(X_0) - g(u_n))\mathbb{1}\{X_0 > u_n\}]} \;=\; \frac{\mathbf{P}[X_0 > u_n]}{\int_{u_n}^{\infty} \frac{\mathbf{P}[X_0 > x]}{x}\,\mathrm{d}x} \;\sim\; u_n^{\beta} \;\sim\; \xi(u_n)\,,$$

which confirms that Assumption (59) holds with constant $c = 1$. Applying the mean value theorem, we find that for each $t \in \mathbb{R}$, there exists a sequence $(\theta_n(t)) \subset (0,1)$ such that

$$\begin{aligned} r_n\left(\frac{\mathbf{P}[X_0 > u_n]}{\mathbf{P}[X_0 > (1+t/r_n)u_n]} \;-\; 1\right) \;&=\; \frac{r_n}{\mathbf{P}[X_0 > (1+t/r_n)u_n]} \frac{u_n t}{r_n} f_{X_0}((1+\theta_n t/r_n)u_n) \\ &\sim\; \frac{u_n^{\beta}}{f_{X_0}((1+t/r_n)u_n)} t f_{X_0}((1+\theta_n t/r_n)u_n) \\ &\sim\; t u_n^{\beta} \;\sim\; t\xi(u_n)\,. \end{aligned}$$

This confirms that Assumption (60) holds in Case (ii). For any sequence (τ_n) satisfying Conditions (63) and (64), we have

$$\begin{aligned} \frac{\mathbf{P}[X_0 > u_n]}{\mathbf{P}[X_0 > u_n(1 \pm \tau_n)]} \;&\sim\; \frac{f_{X_0}(u_n)}{f_{X_0}(u_n(1 \pm \tau_n))} \frac{1}{(1 \pm \tau_n)^{1-\beta}} \\ &\sim\; \frac{\omega(u_n)}{\omega(u_n(1 \pm \tau_n))} \exp(-u_n^{\beta}(1 - (1 \pm \tau_n)^{\beta})/\beta) \;\to\; 1\,, \end{aligned}$$

as $n \to \infty$. This holds because τ_n decreases polynomially in n (since r_n grows polynomially), whereas u_n grows only logarithmically. Thus Assumption (65) is satisfied.

It remains to establish the joint convergence stated in (62). To this end, we apply the Cramér-Wold device with coefficients $\lambda_1, \lambda_2 \in \mathbb{R}$. Our goal is to show that for any sequence (ϵ_n), satisfying (61),

$$\frac{r_n/\xi(u_n)}{n} \sum_{t=1}^{n} \left[\lambda_1 \left(\frac{(g(X_t) - g(u_n))_+}{\mathbf{E}\left[(g(X_0) - g(u_n))_+\right]} \;-\; 1\right) \;+\; \lambda_2 \left(\frac{\mathbb{1}\{X_t > u_n(1+\epsilon_n)\}}{\mathbf{P}[X_0 > u_n(1+\epsilon_n)]} \;-\; 1\right)\right]$$

converges in distribution to

$$\begin{cases} Z_{\nu\wedge 2}\left(\lambda_1 \dfrac{\nu}{1+\nu} \;+\; \lambda_2\right) & \text{in Case (i), and} \\ Z_2\left(\lambda_1 \;+\; \lambda_2\right) & \text{in Case (ii).} \end{cases} \tag{82}$$

The proof consists in reducing (82) to an application of Theorem 3.5, and it is organized in two consecutive steps. Define the function

$$G_n(x) \;=\; \lambda_1 \cdot \frac{(g(x) - g(u_n))_+ \,\mathbf{P}[X_0 > u_n]}{\mathbf{E}[(g(X_0) - g(u_n))_+]} \;+\; \lambda_2 \cdot \mathbb{1}\{x > u_n\}\,.$$

Step 1: Observe that by construction

$$\mathbf{E}[G_n(X_0)] \;=\; (\lambda_1 \;+\; \lambda_2)\,\mathbf{P}[X_0 > u_n]\,,$$

which yields

$$\frac{1}{\mathbf{P}[X_0 > u_n]} \sum_{t=1}^n (G_n(X_t) \ - \ \mathbf{E}[G_n(X_0)]) = \sum_{t=1}^n \left(\frac{G_n(X_t)}{\mathbf{P}[X_0 > u_n]} \ - \ (\lambda_1 + \lambda_2) \right).$$

We now write the decomposition

$$\begin{aligned}
&\sum_{t=1}^n \left(\frac{G_n(X_t)}{\mathbf{P}[X_0 > u_n]} \ - \ (\lambda_1 + \lambda_2) \right) \\
&= \sum_{t=1}^n \left[\lambda_1 \left(\frac{(g(X_t) - g(u_n))_+}{\mathbf{E}\left[(g(X_0) - g(u_n))_+\right]} \ - \ 1 \right) \ + \ \lambda_2 \left(\frac{\mathbb{1}\{X_t > u_n(1+\epsilon_n)\}}{\mathbf{P}[X_0 > u_n(1+\epsilon_n)]} \ - \ 1 \right) \right] \\
&+ \ \frac{n}{r_n/\xi(u_n)} \lambda_2 R \,,
\end{aligned}$$

where

$$R \ := \ \frac{r_n/\xi(u_n)}{n} \sum_{t=1}^n \left(\frac{\mathbb{1}\{X_t > u_n\}}{\mathbf{P}[X_0 > u_n]} \ - \ \frac{\mathbb{1}\{X_t > u_n(1+\epsilon_n)\}}{\mathbf{P}[X_0 > u_n(1+\epsilon_n)]} \right) \ = \ R_1 \ - \ R_2 \,,$$

and R_1, R_2 are defined as

$$\begin{aligned}
R_1 \ &:= \ \frac{\mathbf{P}[X_0 > u_n] - \mathbf{P}[X_0 > u_n(1+\epsilon_n)]}{\mathbf{P}[X_0 > u_n]} \frac{r_n/\xi(u_n)}{n} \\
&\quad \cdot \sum_{t=1}^n \left(\frac{\mathbb{1}\{X_t > u_n\} \ - \ \mathbb{1}\{X_t > u_n(1+\epsilon_n)\}}{\mathbf{E}[\mathbb{1}\{X_0 > u_n\} \ - \ \mathbb{1}\{X_0 > u_n(1+\epsilon_n)\}]} \ - \ 1 \right), \\
R_2 \ &:= \ \frac{\mathbf{P}[X_0 > u_n] \ - \ \mathbf{P}[X_0 > u_n(1+\epsilon_n)]}{\mathbf{P}[X_0 > u_n]} \frac{r_n/\xi(u_n)}{n} \sum_{t=1}^n \left(\frac{\mathbb{1}\{X_t > u_n(1+\epsilon_n)\}}{\mathbf{P}[X_0 > u_n(1+\epsilon_n)]} \ - \ 1 \right).
\end{aligned}$$

If we can show that $R_1, R_2 \xrightarrow{\mathbf{P}} 0$, then the convergence in Equation (82) follows from the convergence of

$$\frac{r_n/\xi(u_n)}{n\mathbf{P}[X_0 > u_n]} \sum_{t=1}^n (G_n(X_t) \ - \ \mathbf{E}[G_n(X_0)]) \,.$$

Step 2: Obviously $r_n = n^{1-(d+1/\alpha)}$ with

$$\alpha \ = \ \begin{cases} \nu \wedge 2 & \text{in Case (i), and} \\ 2 & \text{in Case (ii).} \end{cases}$$

We aim to combine Theorems 3.5 and Theorem 3.1 to show that

$$\begin{aligned}
&\frac{r_n/\xi(u_n)}{n\mathbf{P}[X_0 > u_n]} \sum_{t=1}^n (G_n(X_t) \ - \ \mathbf{E}[G_n(X_0)]) \\
&= n^{1-(d+1/\alpha)} \cdot \frac{1}{f_{X_0}(u_n)} \cdot \frac{1}{n} \sum_{t=1}^n (G_n(X_t) \ - \ \mathbf{E}[G_n(X_0)])
\end{aligned} \tag{83}$$

converges in distribution to the limit announced in (82). To apply this result it is sufficient to show that

$$G'_{\infty,n}(0) \ = \ \begin{cases} f_{X_0}(u_n) \left(\lambda_1 \dfrac{\nu}{1+\nu} \ + \ \lambda_2 \ + \ o(1) \right) & \text{in Case (i),} \\ f_{X_0}(u_n)(\lambda_1 \ + \ \lambda_2 \ + \ o(1)) & \text{in Case (ii).} \end{cases} \tag{84}$$

Indeed, from this, Theorem 3.5 will imply that in Case (i),

$$\begin{aligned}
&\left\| n^{-d-1/\alpha} \frac{1}{f_{X_0}(u_n)} \sum_{t=1}^{n} (G_n(X_t) \ - \ \mathbf{E}[G_n(X_0)]) \ - \ \left(\lambda_1 \frac{\nu}{1+\nu} \ + \ \lambda_2 \ + \ o(1) \right) n^{-d-1/\alpha} \sum_{t=1}^{n} X_t \right\|_{L^{r_0}(\mathbf{P})} \\
&\lesssim \ \frac{n^{\kappa_0+\delta/2-d-1/\alpha}}{f_{X_0}(u_n)} \ \to \ 0 \,,
\end{aligned}$$

as $n \to \infty$. This conclusion follows from $f_{X_0} \in \mathrm{RV}_{-\nu-1}$ and $u_n = \mathrm{o}(n^{(d+1/\alpha-\kappa_0-\delta)/(\nu+1)})$. In Case (ii), we similarly obtain

$$\begin{aligned}
&\left\| n^{-d-1/2} \frac{1}{f_{X_0}(u_n)} \sum_{t=1}^{n} (G_n(X_t) \ - \ \mathbf{E}[G_n(X_0)]) \ - \ (\lambda_1 \ + \ \lambda_2 \ + \ o(1))\, n^{-d-1/2} \sum_{t=1}^{n} X_t \right\|_{L^{r_0}(\mathbf{P})} \\
&\lesssim \ \frac{n^{\kappa_0+\delta/2-d-1/2}}{f_{X_0}(u_n)} \to 0 \,,
\end{aligned}$$

as $n \to \infty$, due to (58). (83) and the announced result.

We now provide details on both steps.

Details Step 1: Analysis of R_1

We apply Theorem 3.5 using the function

$$G_n^{R_1}(x) \ = \ \mathbb{1}\{x > u_n\} \ - \ \mathbb{1}\{x > u_n(1+\epsilon_n)\} \,,$$

which satisfies the bound

$$|G_n^{R_1}(x)| \leq \mathbb{1}\{x > u_n(1-(-\epsilon_n)_+)\} \,.$$

This ensures that Assumption 2 is met, with the threshold u_n replaced by $u_n(1-(-\epsilon_n)_+)$. We then decompose R_1 as

$$R_1 \ = \ R_{1,1} \ + \ R_{1,2} \,,$$

where

$$\begin{aligned}
R_{1,1} \ &:= \ \frac{1}{\mathbf{P}[X_0 > u_n]} \frac{r_n/\xi(u_n)}{n} \cdot \sum_{t=1}^{n} \left(G_n^{R_1}(X_t) \ - \ \mathbf{E}[G_n^{R_1}(X_0)] \ - \ (G_{n,\infty}^{R_1})'(0) X_t \right) \,, \\
R_{1,2} \ &:= \ \frac{1}{\mathbf{P}[X_0 > u_n]} \frac{r_n/\xi(u_n)}{n} \cdot (G_{n,\infty}^{R_1})'(0) \sum_{t=1}^{n} X_t \,.
\end{aligned}$$

First we analyze $R_{1,1}$. Note that by construction

$$\frac{1}{\mathbf{P}[X_0 > u_n]} \frac{r_n/\xi(u_n)}{n} = \frac{n^{-d-1/\alpha}}{f_{X_0}(u_n)} \,.$$

Applying Theorem 3.5, we find that

$$\begin{aligned}
&\|R_{1,1}\|_{L^{r_0}(\mathbf{P})} \\
&= \ \frac{1}{f_{X_0}(u_n)} \left\| n^{-d-1/\alpha} \sum_{t=1}^{n} \left(G_n^{R_1}(X_t) \ - \ \mathbf{E}[G_n^{R_1}(X_0)] \ - \ (G_{n,\infty}^{R_1})'(0) X_t \right) \right\|_{L^{r_0}(\mathbf{P})} \\
&\leq \ \frac{n^{\kappa_0+\delta/2-d-1/\alpha}}{f_{X_0}(u_n)} \ \to \ 0 \,,
\end{aligned}$$

due to the assumed growth rate on u_n, as was shown during the outline of Step 2, so that $R_{1,1} \xrightarrow{\mathbf{P}} 0$. Next we turn to $R_{1,2}$. Since $r_n/(nu_n) = n^{-d-1/\alpha}$, it follows from Theorem 3.1 that

$$\frac{r_n}{nu_n} \sum_{t=1}^{n} X_t \xrightarrow{\mathrm{d}} Z_\alpha \,.$$

The remaining factor is

$$\frac{u_n (G^{R_1}_{n,\infty})'(0)}{\mathbf{P}[X_0 > u_n]\xi(u_n)} \,,$$

which we will show tends to zero as $n \to \infty$. The stated convergence of $R_{1,2}$ then follows by Slutsky's lemma. We now use Lemma A.4(ii) in order to evaluate

$$\frac{u_n (G^{R_1}_{n,\infty})'(0)}{\mathbf{P}[X_0 > u_n]\xi(u_n)} \;=\; \frac{f_{X_0}(u_n(1+\epsilon_n)) - f_{X_0}(u_n)}{f_{X_0}(u_n)} \;=\; \frac{f_{X_0}(u_n(1+\epsilon_n))}{f_{X_0}(u_n)} \;-\; 1\,.$$

In Case (i), we assume $f_{X_0} \in \mathrm{RV}_{-\nu-1}$, so by local uniformity of regular variation

$$\frac{f_{X_0}(u_n(1+\epsilon_n))}{f_{X_0}(u_n)} \;=\; (1+\epsilon_n)^{-\nu-1} \;+\; o(1) \;\to\; 1\,,$$

as $n \to \infty$, since $\epsilon_n \to 0$. In Case (ii), we have

$$\frac{f_{X_0}(u_n(1+\epsilon_n))}{f_{X_0}(u_n)} \;=\; \frac{\omega(u_n(1+\epsilon_n))}{\omega(u_n)} \exp(u_n^\beta(1-(1+\epsilon_n)^\beta)/\beta) \;\to\; 1\,, \tag{85}$$

since ϵ_n decays polynomially in n while u_n grows only logarithmically, and ω is either constant or regularly varying at infinity. This shows $R_{1,2} \xrightarrow{\mathbf{P}} 0$, and thus $R_1 \xrightarrow{\mathbf{P}} 0$.

Details Step 1: Analysis of R_2

The analysis of R_2 is more straightforward, as the conditions of Theorem 3.5 are easily verified for the function

$$G^{R_2}_n(x) \;=\; \mathbb{1}\{x > u_n(1+\varepsilon_n)\}\,.$$

Recall first of all that we have just shown the convergence $f_{X_0}(u_n(1+\epsilon_n))/f_{X_0}(u_n) \to 1$. Moreover, in Case (i), since $f_{X_0} \in \mathrm{RV}_{-\nu-1}$, it follows that $x \mapsto \mathbf{P}[X_0 > x] \in \mathrm{RV}_{-\nu}$. Therefore,

$$\frac{\mathbf{P}[X_0 > u_n(1+\epsilon_n)] \;-\; \mathbf{P}[X_0 > u_n]}{\mathbf{P}[X_0 > u_n]} \;=\; (1+\epsilon_n)^{-\nu} \;-\; 1 \;+\; o(1) \;\to\; 0\,.$$

In Case (ii), we apply the mean value theorem: there exists $(\theta_n) \subset (0,1)$ such that

$$\frac{\mathbf{P}[X_0 > u_n(1+\epsilon_n)] \;-\; \mathbf{P}[X_0 > u_n]}{\mathbf{P}[X_0 > u_n]} \;=\; -u_n\epsilon_n \frac{f_{X_0}(u_n)}{\mathbf{P}[X_0 > u_n]} \frac{f_{X_0}(u_n(1+\theta_n\epsilon_n))}{f_{X_0}(u_n)} \;\to\; 0\,,$$

as $n \to \infty$, since $u_n f_{X_0}/\mathbf{P}[X_0 > u_n] \sim u_n^\beta$ with $\epsilon_n u_n^\beta \to 0$, and the remaining term converges to 1 due to (85). This entails $\mathbf{P}[X_0 > u_n(1+\epsilon_n)]/\mathbf{P}[X_0 > u_n] \to 1$. Then, applying Theorem 3.5 and Theorem 3.1, we obtain that

$$n^{1-(d+1/\alpha)} \cdot \frac{\mathbf{P}[X_0 > u_n(1+\varepsilon_n)]}{f_{X_0}(u_n(1+\varepsilon_n))} \cdot \frac{1}{n} \sum_{t=1}^{n} \left(\frac{\mathbb{1}\{X_t > u_n(1+\epsilon_n)\}}{\mathbf{P}[X_0 > u_n(1+\epsilon_n)]} \;-\; 1 \right)$$

converges in distribution to a non-degenerate random variable, and so does

$$
\begin{aligned}
& n^{1-(d+1/\alpha)} \cdot \frac{\mathbf{P}[X_0 > u_n]}{f_{X_0}(u_n)} \cdot \frac{1}{n} \sum_{t=1}^{n} \left(\frac{\mathbb{1}\{X_t > u_n(1+\epsilon_n)\}}{\mathbf{P}[X_0 > u_n(1+\epsilon_n)]} \; - \; 1 \right) \\
&= \frac{r_n/\xi(u_n)}{n} \sum_{t=1}^{n} \left(\frac{\mathbb{1}\{X_t > u_n(1+\epsilon_n)\}}{\mathbf{P}[X_0 > u_n(1+\epsilon_n)]} \; - \; 1 \right) \\
&= \frac{\mathbf{P}[X_0 > u_n]}{\mathbf{P}[X_0 > u_n(1+\epsilon_n)] \; - \; \mathbf{P}[X_0 > u_n]} R_2 .
\end{aligned}
$$

We conclude that in both settings it holds $R_2 \xrightarrow{\mathbf{P}} 0$.

We will now continue the analysis separated into the settings, where it suffices to check the validity of (84) in order to complete the proof.

Details Step 2

We begin by verifying Equation (84) in Setting (i). To simplify the notation, we write

$$
(\log(x) - \log(u_n)) \mathbb{1}\{x > u_n\} \; = \; \max\{\log(x/u_n), 0\} \; =: \; \log(x/u_n)_+ .
$$

It then follows that

$$
\begin{aligned}
G'_{\infty,n}(0) \; &= \; - \int_{\mathbb{R}} G_n(x) f'_{X_0}(x) \, \mathrm{d}x \\
&= \; -\lambda_1 \frac{\int_{u_n}^{\infty} \log\left(x/u_n\right) f'_{X_0}(x) \, \mathrm{d}x}{\mathbf{E}\left[\log\left(X_0/u_n\right)_+\right]} \mathbf{P}[X_0 > u_n] \; + \; \lambda_2 \cdot f_{X_0}(u_n) \, .
\end{aligned}
$$

Using integration by parts and the regular variation of $f_{X_0} \in \mathrm{RV}_{-\nu-1}$, we obtain

$$
\begin{aligned}
- \int_{u_n}^{\infty} \log\left(x/u_n\right) f'_{X_0}(x) \, \mathrm{d}x \; &= \; \left[-f_{X_0}(x) \log(x/u_n)\right]_{x=u_n}^{x\to\infty} \; + \; \int_{u_n}^{\infty} \frac{1}{x} f_{X_0}(x) \, \mathrm{d}x \\
&= \; \int_{u_n}^{\infty} \frac{1}{x} f_{X_0}(x) \, \mathrm{d}x \, .
\end{aligned}
$$

Changing variables, and using the uniform convergence of $f_{X_0}(u_n x)/f_{X_0}(u_n)$ to 1 in neighborhoods of infinity, this becomes:

$$
\begin{aligned}
- \int_{u_n}^{\infty} \log\left(x/u_n\right) f'_{X_0}(x) \, \mathrm{d}x \; &= \; f_{X_0}(u_n) \int_{1}^{\infty} \frac{1}{x} \frac{f_{X_0}(u_n x)}{f_{X_0}(u_n)} \, \mathrm{d}x \\
&\sim \; f_{X_0}(u_n) \int_{1}^{\infty} \frac{1}{x^{2+\nu}} \, \mathrm{d}x \\
&= \; f_{X_0}(u_n) \frac{1}{1+\nu} \, .
\end{aligned}
$$

Recalling (81), it follows that

$$
G'_{\infty,n}(0) \quad = \quad f_{X_0}(u_n) \left(\lambda_1 \frac{\nu}{1+\nu} \; + \; \lambda_2 \; + \; o(1) \right) ,
$$

which verifies Equation (84) in Setting (i).

To prove the same equation in Case (ii), first note that in the proof of Corollary 3.12 we showed

$$
\frac{\mathbf{E}[G_n(X_0)]}{G'_{\infty,n}(0)} = \frac{\int_{u_n}^{\infty} \log(x/u_n) f_{X_0}(x) \, \mathrm{d}x}{- \int_{u_n}^{\infty} \log(x/u_n) f'_{X_0}(x) \, \mathrm{d}x} \; \sim \; \frac{\mathbf{P}[X_0 > u_n]}{f_{X_0}(u_n)} \; \sim \; u_n^{1-\beta} \, .
$$

From this it follows immediately that

$$\begin{aligned} G'_{\infty,n}(0) &= u_n^{\beta-1}\mathbf{E}[G_n(X_0)] = u_n^{\beta-1}\mathbf{P}[X_0 > u_n](\lambda_1+\lambda_2+o(1)) \\ &= f_{X_0}(u_n)(\lambda_1+\lambda_2+o(1))\,, \end{aligned}$$

that is, (84). Having shown (84) in both cases, we conclude the proof. $\square$

Chapter 6

Central limit theory for serial tail dependence estimators in heavy-tailed long memory linear time series

Central limit theory for serial tail dependence estimators in heavy-tailed long memory linear time series

Ioan Scheffel[a], Marco Oesting[a,b], Gilles Stupfler[c]

[a] Institute for Stochastics and Applications, University of Stuttgart, D-70563 Stuttgart, Germany

[b] Stuttgart Center for Simulation Science (SC SimTech), University of Stuttgart, D-70569 Stuttgart, Germany

[c] Univ Angers, CNRS, LAREMA, SFR MATHSTIC, F-49000 Angers, France

Abstract

We prove multiple central limit theorems for serial tail dependence estimators in heavy-tailed long memory linear time series. The main theoretical tools are two novel multivariate reduction principles for partial sums of heavy-tailed long memory linear time series, subordinated over sliding windows and above a threshold growing with sample size. This requires addressing several substantial difficulties, including handling a nonlinear, sample-size dependent, and multivariate subordination mechanism, the dependence between several overlapping linear processes, and the lack of higher-order moments of the marginal distribution. Despite these obstacles, our assumptions are mild and, in particular, the innovation process is allowed to have infinite variance. A key feature of our theory is that our second reduction principle holds uniformly in the threshold, allowing central limit theory for empirical extremograms with sample quantiles as thresholds. This question has received little attention in the literature on serial extremal dependence estimation even though the version of empirical extremograms with random thresholds is ubiquitous in practice. We compare our results in several respects with those that may be obtained under short-range dependence, thereby discovering markedly different convergence rates and limit laws in our long memory setting.

MSC 2020 subject classifications: 60F05, 60F17, 60G70

Keywords: Central limit theory, Extremogram, Heavy tails, Peaks-over-Threshold model, Stable distribution, Uniform reduction principle

1. Introduction

Traditionally, dependence in time series is characterized by the autocorrelation function (ACF). Since the ACF emphasizes dependence at central levels, it does not accurately capture dependence at extreme levels. Measuring and estimating tail dependence is therefore a separate task. The analog of the ACF in extreme value analysis is the so-called extremogram, proposed by Davis and Mikosch (2009). Given a stationary time series $(X_t)_{t\in\mathbb{Z}}$ whose marginal distribution has upper endpoint $+\infty$, the extremogram is defined as the limiting quantity

$$\gamma_{A,B}(h) \;:=\; \lim_{u\to\infty} \frac{\mathbf{P}\left[[X_\ell]_\ell \in u\cdot A\,, [X_{\ell+h}]_\ell \in u\cdot B\right]}{\mathbf{P}[X_0 > u]}\,, \qquad h\in\mathbb{Z}\,, A,B\subset\mathbb{R}^T\,, T\in\mathbb{N}\,,$$

provided that the limit exists, where throughout $[X_\ell]_\ell := [X_0,\ldots,X_{T-1}]$. The simplest example is for $A=B=(1,\infty)$, when the extremogram coincides with the tail-dependence coefficient

$$\chi(h) \;:=\; \lim_{u\to\infty} \mathbf{P}\left[X_h > u \mid X_0 > u\right]\,, \qquad h\in\mathbb{Z}\,.$$

With this quantity, one can answer the question whether an extreme value at time 0 makes an extreme value at time h more likely. Even though the probability of rare events vanishes, that is, $\mathbf{P}[X_0>u]\to 0$ as $u\to\infty$, the tail dependence coefficient $\chi(h)$ can be positive. We then call the stationary time series asymptotically dependent; see Chapter 9.5 in Beirlant et al. (2006) for a detailed discussion of modeling asymptotic dependence using tail dependence coefficients. In general, the extremogram is estimated from observations of $X_1,\ldots,X_n$ by the natural estimator

$$\widehat{\gamma}_{A,B}(h) := \frac{\frac{1}{n}\sum_{t=1-(0\wedge h)}^{n-T+1-(0\vee h)} \mathbb{1}\left\{[X_{t+\ell}]_\ell \in u_n\cdot A\,, [X_{t+\ell+h}]_\ell \in u_n\cdot B\right\}}{\frac{1}{n}\sum_{t=1}^{n}\mathbb{1}\{X_t>u_n\}}\,, \qquad h\in\mathbb{Z}\,, A,B\subset\mathbb{R}^T\,, T\in\mathbb{N}\,,$$

where the threshold sequence (u_n) is chosen such that $u_n\to\infty$ and $n\mathbf{P}[X_0>u_n]\to\infty$. We call this the empirical extremogram. The asymptotic analysis of the empirical extremogram requires dedicated techniques that differ from the usual ACF analysis; see Brockwell and Davis (1991) for a discussion of asymptotic properties of the empirical ACF in standard time series settings. The asymptotic behavior of the empirical extremogram has been studied in Davis and Mikosch (2009) for sufficiently strong α-mixing regularly varying time series satisfying an anti-clustering condition; see Basrak and Segers (2009) for a reference on regularly varying time series. The question of handling tail dependence estimation when mixing conditions of this type do not hold has remained much less explored. This is particularly true for long memory settings.

There is evidence, for example, that in hydrological time series, the assumptions mentioned above (implying short memory) are violated. In contrast, fractionally differenced ARIMA models have been used for hydrological time series (Montanari et al., 1997), long memory in river flows has been linked to prolonged droughts and temporal clustering of extreme floods (Mudelsee, 2007), and flood peak distributions often exhibit heavy-tail behavior (Merz et al., 2022). There is still debate over whether long memory is present in financial data; see for example Rossi and Santucci de Magistris (2013) or Cont (2005).

In this work, we study tail dependence estimators for long memory linear time series with possibly infinite variance. More precisely, the main contribution of our article is to derive central limit theorems for tail dependence estimators in long memory heavy-tailed linear time series. For this, the available theory does not provide the required tools, since it is essentially based on one-dimensional reduction principles (Scheffel et al., 2026) and does not cover the multivariate nature of dependence estimators. The empirical process theory of Kulik and Soulier (2011) is developed exclusively for stochastic volatility models and is therefore not applicable for the current class of processes. We close this gap by developing multivariate reduction principles for threshold-dependent indicator functions. This is not a routine extension of Scheffel et al. (2026), because the multivariate setting creates dependence between several overlapping linear processes, and the bounds needed to handle this dependence have to be developed separately. Classical tools used to control empirical ACFs, as in Hosking (1996), cannot be used either, because the extreme value context involves estimators whose summands depend on the sample size n. Accounting for long-range dependence and a subordination mechanism depending on sample size in a nonlinear way makes the arguments substantially different from those of Betken et al., in which short-range dependence is considered and the subordination mechanism varies with n through scaling constants.

We prove a general reduction principle that allows us to derive central limit theory for a large class of estimators that contains in particular all extremogram estimators considered in Davis and Mikosch (2009). We then restrict the class of admissible sets to obtain a uniform version of this reduction principle that allows us to consider extremogram estimators with random thresholds, that is,

$$\widetilde{\gamma}_{A,B}(h) \ := \ \frac{1}{k} \sum_{t=1-(0\wedge h)}^{n-T+1-(0\vee h)} \mathbb{1}\left\{[X_{t+\ell}]_\ell \in X_{n-k:n} \cdot A \, , [X_{t+\ell+h}]_\ell \in X_{n-k:n} \cdot B\right\} \, , \quad h \in \mathbb{Z} \, , A, B \subset \mathbb{R}^T \, , T \in \mathbb{N} \, ,$$

where $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ are the order statistics of the sample $X_1, \ldots, X_n$. This feature of our approach is uncommon in theoretical investigations: for example, the estimation theory in Davis and Mikosch (2009) covers only deterministic thresholds. However, the version with random thresholds $\widetilde{\gamma}_{A,B}(h)$ of the empirical extremogram is used in practice (Davis and Mikosch, 2009, Section 3.4). As far as we are aware, only Drees establishes the asymptotic behavior of the empirical extremogram with random thresholds in a weakly dependent setting and under several technical assumptions. See Corollary 3.6 therein, in which it appears that the weak limits of $\widehat{\gamma}_{A,B}(h)$ and $\widetilde{\gamma}_{A,B}(h)$ are identical when short-range dependence is present. In general, replacing deterministic with random thresholds in asymptotic theory is known to be difficult, especially in the presence of serial dependence, see Drees and Knežević (2020).

We also show in this work that, by contrast with the short-range dependence setting, the weak limits of $\widehat{\gamma}_{A,B}(h)$ and $\widetilde{\gamma}_{A,B}(h)$ are different for A and B in a wide class of sets relevant in applications. In particular, we show that the estimators

$$\frac{1}{k} \sum_{t=1}^{n-T+1} \mathbb{1}\left\{X_t > X_{n-k:n}, X_{t+1} > X_{n-k:n}, \ldots, X_{t+T-1} > X_{n-k:n}\right\}$$

and

$$\frac{1}{k} \sum_{t=1}^{n-T+1} \mathbb{1}\left\{X_t > X_{n-k:n}, X_{t+T-1} > X_{n-k:n}\right\},$$

of the tail-dependence coefficients

$$\lim_{u\to\infty} \mathbf{P}\left[X_0 > u, \ldots, X_{T-1} > u \,|\, X_0 > u\right]$$

and

$$\lim_{u\to\infty} \mathbf{P}\left[X_{T-1} > u \,|\, X_0 > u\right] \ = \ \chi(T-1),$$

respectively, have a degenerate asymptotic distribution which is a hitherto unknown phenomenon within the long memory setting.

This paper is organized as follows. In Section 2 we introduce our model and main assumptions. In Section 3 we derive reduction principles for tail dependence estimators that we use to derive the central limit theory of Section 4. In Section 5 we show that our theory for random thresholds is applicable in the class of rectangular sets by providing simplified assumptions and showing their validity. Then we investigate the cancellation in the asymptotic scale for rectangular sets. Since our theory is also valid in the univariate setting, we finally compare results for random thresholds with results that are readily derived from existing theory in the i.i.d. and weakly dependent setting.

2. Model and main assumptions

2.1. Notation

Throughout this article, $(\Omega, \mathcal{A}, \mathbf{P})$ denotes a probability space rich enough to support all random variables under consideration. The symbols $\mathbf{P}^*$ and $\mathbf{E}^*$ denote outer probability and expectation, respectively (Van Der Vaart and Wellner, 2023, Section 1.2). For $p \geq 1$, we write $L^p(\mathbf{P})$ for the space of random variables with finite p-th moment. For a generic, continuous random variable Z, we let F_Z (resp. f_Z, q_Z) denote its cumulative distribution function (resp. probability density function, quantile function). We write $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$ for $a, b \in \mathbb{R}$, and use the symbol $\sim$ to denote asymptotic equivalence of sequences and functions. We write $\lesssim$ to mean inequality up to a generic multiplicative constant $C > 0$ that can change from place to place. If the dependence of this constant on the surrounding variables is relevant to the argument, we will make this explicit, by saying, for example, that $C = C_n$ depends on n, or that C is independent of n. Throughout we will write $\mathbf{x} = [x_\ell]_\ell = [x_\ell]_{\ell \in \{0,\dots,T-1\}} \in \mathbb{R}^T$ where $T \geq 1$ is fixed. If the index set is different from $\{0, \dots, T-1\}$, we will make this explicit. For $p \geq 0$, the symbol $C^p(\mathbb{R}^T)$ denotes the vector space of real-valued functions on $\mathbb{R}^T$ with continuous partial derivatives up to order p. For a set $A \subset \mathbb{R}^T$ and $s \in \mathbb{R} \cup \{+\infty\}$ we write $s \cdot A = \{s \cdot x \mid x \in A\}$, and set by convention $\infty \cdot A = \emptyset$. Finally, we write $\mathbb{N} := \{1, 2, \dots\}$ and $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$.

2.2. Assumptions on the model

Let $(\varepsilon_t)_{t \in \mathbb{Z}}$ be a sequence of independent and identically distributed (i.i.d.) copies of a random variable ε. In this work, we focus on causal linear (or MA(∞)) time series $X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$, where $(a_j)_{j \in \mathbb{Z}}$ is a sequence of real-valued coefficients such that $a_0 = 1$ and $a_{-j} = 0$ for all $j \geq 1$. Our regularity assumptions focus on the distribution of the innovation sequence, on the rate of decay of the coefficients a_j which should be slow enough to yield a divergent series, yet fast enough to accommodate tail heaviness of the innovations, and on the tail of its marginal distribution. We consider the case of heavy-tailed time series whose innovations have a finite first moment, but possibly infinite second moment.

Assumption 1 (Regularity of the innovations)**.** Suppose that $F_\varepsilon \in C^2(\mathbb{R})$, and that ε has a symmetric distribution, and that there is a finite positive constant L and $\nu \in (1, \infty) \setminus \{2\}$ such that $\lim_{x\to\infty} x^\nu \mathbf{P}[\varepsilon > x] = \lim_{x\to\infty} x^\nu \mathbf{P}[\varepsilon < -x] = L/2$, and suppose that the probability density function $f_\varepsilon \in C^1(\mathbb{R})$ satisfies:

$$
\begin{aligned}
f_\varepsilon(x) \vee |f'_\varepsilon(x)| &\lesssim \frac{1}{(1+|x|)^\alpha} \qquad \text{for } x \in \mathbb{R}\,, \\
|f'_\varepsilon(x) - f'_\varepsilon(y)| &\lesssim |x-y| \cdot \frac{1}{(1+|x|)^\alpha} \qquad \text{for } x, y \in \mathbb{R}\,,\ |x-y| < 1\,,
\end{aligned}
$$

where $\alpha := 2 \wedge \nu$, and $\lesssim$ denotes inequality up to a multiplicative constant depending only on the distribution of ε.

Assumption 2 (Long memory)**.** $(n^{1-d} a_n)$ converges to a finite, nonzero limit c_a for some $d \in (0, 1 - 1/\alpha)$.

Assumption 3 (Regular variation of the marginal density)**.** The density f_{X_0} is regularly varying at $+\infty$ with index $-1-\nu$.

Remark 2.1 (Assumptions 1, 2 and 3)**.** These assumptions are already used in Scheffel et al. (2026). They are reasonably weak, given the difficulty of the heavy-tailed setting. The edge case $\nu = 2$ is excluded to streamline the analysis.

Remark 2.2 (Link between Assumption 3 and unimodality)**.** If the distribution of ε is unimodal, meaning that its distribution function is convex on $(-\infty, 0)$ and concave on $(0, \infty)$ (Dharmadhikari and Joag-Dev, 1988, Definition 1.1), then by Theorem 1.1 in Dharmadhikari and Joag-Dev (1988) it holds that any convergent moving average also has a unimodal distribution function. On the level of densities, this means monotonicity on $(-\infty, 0)$ and $(0, \infty)$. Then, from the regular variation of the moving average (Kulik and Soulier, 2020, (15.3.5)) and the monotone density theorem (De Haan and Ferreira, 2006, Proposition B.1.9.11), we get that the density of the moving average is also regularly varying with index $-1-\nu$. In particular, Assumption 3 is satisfied under Assumption 1 and unimodality of the innovations.

2.3. Assumptions on the multivariate subordination mechanism

We want to derive central limit theory for partial sums of long memory linear time series subordinated over sliding windows, that is, for

$$\sum_{t=1}^{n-T+1} \left(G_n([X_{t+\ell}]_\ell) - \mathbf{E}[G_n([X_\ell]_\ell)]\right)$$

where $(G_n : \mathbb{R}^T \to \mathbb{R})$ is a sequence of subordination mechanisms. The following definition specifies the class of such mechanisms considered in this work.

Definition 2.3 (Subordination mechanism)**.** Let $(u_n) \subset [1, \infty)$ be a threshold sequence with $u_n \to \infty$ as $n \to \infty$, let $A \subset \mathbb{R}^T$ be a Borel set, and let $s_1, s_2 \in (0, \infty]$ satisfy $s_1 \le s_2$. We define

$$G_n(\mathbf{z}, s_1, s_2, A) \;:=\; \mathbb{1}\left\{\mathbf{z} \;\in\; u_n \cdot ((s_1 \cdot A) \setminus (s_2 \cdot A))\right\}, \qquad \mathbf{z} \in \mathbb{R}^T.$$

Putting $s_1 = 1$ and $s_2 = \infty$ reflects extremogram-type tail-dependence estimators as in Davis and Mikosch (2009), where we require the set $A \subset \mathbb{R}^T$ to be bounded away from 0. Apart from that, we are interested in the asymptotic behavior of the (rescaled) empirical extremogram

$$\sum_{t=1}^{n-T+1} \mathbb{1}\{[X_{t+\ell}]_\ell \in X_{n-k:n} \cdot A\}.$$

Write

$$\sum_{t=1}^{n-T+1} \mathbb{1}\{[X_{t+\ell}]_\ell \in X_{n-k:n} \cdot A\} = \sum_{t=1}^{n-T+1} \mathbb{1}\{[X_{t+\ell}]_\ell \in u_n \cdot (X_{n-k:n}/u_n) \cdot A\},$$

where the deterministic threshold u_n is replaced by the sample quantile at level $1-k/n$. This suggests, under the assumption that $X_{n-k:n}/u_n \to_{\mathbf{P}} 1$ as $n \to \infty$, to investigate the asymptotic behavior of

$$\sum_{t=1}^{n-T+1} \mathbb{1}\{[X_{t+\ell}]_\ell \in u_n \cdot s \cdot A\}$$

uniformly in s in a neighborhood of 1. Our argument relies on a chaining procedure. We first analyze the pointwise asymptotic behavior of the relevant quantity and then show that for neighboring values s and s' around 1, the difference

$$\sum_{t=1}^{n-T+1} \mathbb{1}\{[X_{t+\ell}]_\ell \in u_n \cdot s \cdot A\} \; - \; \sum_{t=1}^{n-T+1} \mathbb{1}\{[X_{t+\ell}]_\ell \in u_n \cdot s' \cdot A\}$$

converges to 0; see Section 3.2 for further details. For sets A satisfying $s_2 \cdot A \subset s_1 \cdot A$ if $s_1 < s_2$, this takes the form

$$\sum_{t=1}^{n-T+1} G_n([X_{t+\ell}]_\ell, s_1, s_2, A)$$

which justifies Definition 2.3. The next definition fixes the properties of A that are needed for this extended analysis. Apart from the nesting property, we also need that differences of the form $(s_1 \cdot A) \setminus (s_2 \cdot A)$ are controlled by marginal threshold exceedances.

Definition 2.4 (Class of sets for the uniform reduction principle)**.** Define $\mathcal{C}_T$ to be the family of all Borel sets $\emptyset \neq A \subset \mathbb{R}^T$ bounded away from 0 such that there exist $\{\mathfrak{a}_j\}_{j \in \{1,\dots,T\}} \subset (0,\infty)$ such that for all $s_1, s_2 \in (0,\infty)$ satisfying $s_1 < s_2$, it holds $s_2 \cdot A \subset s_1 \cdot A$ and

$$(s_1 \cdot A) \setminus (s_2 \cdot A) \;\subset\; \left(\bigcup_{j=1}^{T} \{|x_{j-1}| \in \mathfrak{a}_j \cdot [s_1, s_2]\} \right) .$$

If $A \in \mathcal{C}_T$ with a particular $\mathfrak{a} = \{\mathfrak{a}_j\}$, then we write $A \in \mathcal{C}_T(\mathfrak{a})$.

Example 2.5 (Examples, counterexamples and properties of sets in $\mathcal{C}_T$)**.**

(i) [Cartesian product] Let $n, m \in \mathbb{N}$. If $A \in \mathcal{C}_n(\mathfrak{a})$ and $B \in \mathcal{C}_m(\mathfrak{b})$, then $A \times B \in \mathcal{C}_{n+m}(\mathfrak{c})$ with $\mathfrak{c}_j = \mathfrak{a}_j$ for $j \in \{1, \dots, n\}$ and $\mathfrak{c}_j = \mathfrak{b}_{j-n}$ for $j \in \{n+1, \dots, n+m\}$.

(ii) [Unboundedness]. Since A is bounded away from 0, there exists $\mathbf{x} \in A$ such that

$$\max_{j \in \{1,\dots,T\}} |x_{j-1}| \;>\; \delta \qquad \text{for some } \delta > 0 .$$

Then, for $R > \delta$,

$$\frac{R}{\delta} \cdot \mathbf{x} \;\in\; \frac{R}{\delta} \cdot A \;\subset\; A \qquad \text{and} \qquad \max_{j \in \{1,\dots,T\}} \left| \frac{R}{\delta} x_{j-1} \right| \;>\; R .$$

Therefore, A is unbounded.

(iii) [Rectangular sets with positive active coordinates] Let $A = \times_{j=1}^{T} (\mathfrak{a}_j, \infty)$ with $\mathfrak{a}_j \in (0,\infty) \cup \{-\infty\}$, and $\emptyset \neq J_A = \{j \mid \mathfrak{a}_j > 0\}$. It holds for $s_1, s_2 > 0$ satisfying $s_1 < s_2$

$$s_2 \cdot A \;=\; \bigcap_{j \in J_A} \{x_{j-1} \in (s_2 \cdot \mathfrak{a}_j, \infty)\} \;\subset\; \bigcap_{j \in J_A} \{x_{j-1} \in (s_1 \cdot \mathfrak{a}_j, \infty)\} \;=\; s_1 \cdot A ,$$

as well as

$$\begin{aligned} (s_1 \cdot A) \setminus (s_2 \cdot A) \;&=\; \left(\bigcap_{j \in J_A} \{x_{j-1} \in (s_1 \cdot \mathfrak{a}_j, \infty)\} \right) \cap \left(\bigcup_{j \in J_A} \{x_{j-1} \in (-\infty, s_2 \cdot \mathfrak{a}_j]\} \right) \\ &\subset\; \bigcup_{j \in J_A} \{|x_{j-1}| \in \mathfrak{a}_j \cdot [s_1, s_2]\} \end{aligned}$$

so that $A \in \mathcal{C}_T$ with arbitrary $\mathfrak{a}_j \in (0,\infty)$ for $j \notin J_A$.

(iv) [Complements of squares centered at the origin] Let $\|\cdot\|$ be the supremum norm on $\mathbb{R}^T$, $R > 0$, and $A = \{\|x\| > R\}$. Observe that $s \cdot A = \{\|x\| > sR\}$ for any $s > 0$. Pick $s_1, s_2 > 0$ with $s_1 < s_2$. Then $s_2 \cdot A \subseteq s_1 \cdot A$ and $(s_1 \cdot A) \setminus (s_2 \cdot A) = \{\|x\| \in (s_1 R, s_2 R]\}$. We conclude that

$$\begin{aligned}(s_1 \cdot A) \setminus (s_2 \cdot A) &= \left(\bigcap_{j=1}^{T} \{|x_{j-1}| \in [0, s_2 R]\}\right) \cap \left(\bigcup_{j=1}^{T} \{|x_{j-1}| \in (s_1 R, \infty)\}\right) \\ &\subset \bigcup_{j=1}^{T} \{|x_{j-1}| \in R \cdot [s_1, s_2]\}\end{aligned}$$

so that $A \in \mathcal{C}_T$ with $\mathfrak{a}_j = R \in (0, \infty)$ for $j \in \{1, \dots, T\}$.

(v) [Stability under set operations with scale invariant sets] Let B be a scale invariant set, that is, for $s > 0$, $s \cdot B = B$, and let $A \in \mathcal{C}_T(\mathfrak{a})$. Note that $s \cdot B^c = (s \cdot B)^c = B^c$ for $s > 0$, that is, B^c is also scale invariant. Therefore,

$$\begin{aligned} s_2 \cdot (A \cap B) &= (s_2 \cdot A) \cap B \subset (s_1 \cdot A) \cap B = s_1 \cdot (A \cap B), \\ s_2 \cdot (A \cup B) &= (s_2 \cdot A) \cup B \subset (s_1 \cdot A) \cup B = s_1 \cdot (A \cup B), \end{aligned}$$

and

$$\begin{aligned} (s_1 \cdot (A \cap B)) \setminus (s_2 \cdot (A \cap B)) &= ((s_1 \cdot A) \setminus (s_2 \cdot A)) \cap B \subset (s_1 \cdot A) \setminus (s_2 \cdot A), \\ (s_1 \cdot (A \cup B)) \setminus (s_2 \cdot (A \cup B)) &= ((s_1 \cdot A) \setminus (s_2 \cdot A)) \cap B^c \subset (s_1 \cdot A) \setminus (s_2 \cdot A). \end{aligned}$$

If, in addition, $A \cap B \neq \emptyset$, then $A \cap B \in \mathcal{C}_T(\mathfrak{a})$, and if $A \cup B$ is bounded away from 0, then $A \cup B \in \mathcal{C}_T(\mathfrak{a})$. Generally, $\mathcal{C}_T$ is not closed under these set operations.

3. Multivariate Reduction Principle

3.1. A general pointwise reduction principle

We give a heuristic overview of how the multivariate reduction principle works. For $k \in \mathbb{Z}$, let $\mathcal{F}_k = \sigma(\varepsilon_k, \varepsilon_{k-1}, \dots)$ denote the past σ-algebra generated by the sequence (ε_t) up to index k. For $k \geq 0$, let $X_{t,k} := \sum_{j=0}^{k} a_j \varepsilon_{t-j}$ be the k-truncated time series, let $A \subset \mathbb{R}^T$ be bounded away from 0, $s_1, s_2 \in (0, \infty]$ satisfying $s_1 < s_2$, and write

$$\begin{aligned} G_{k,n}(\mathbf{y}) &= G_{k,n}(\mathbf{y}, s_1, s_2, A) = \mathbf{E}[G_n([X_{\ell,k+\ell}]_\ell + \mathbf{y}, s_1, s_2, A)] \\ &= \mathbf{P}[[X_{\ell,k+\ell}]_\ell + \mathbf{y} \in u_n \cdot (s_1 \cdot A) \setminus (s_2 \cdot A)], \end{aligned}$$

and, by Lemma B.3 in the supplementary material,

$$\nabla G_{k,n}(\mathbf{z}) = \nabla_{\mathbf{y}} G_{k,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=\mathbf{z}} = -\int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n \cdot (s_1 \cdot A) \setminus (s_2 \cdot A)\} \nabla f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x} - \mathbf{z})\, \mathrm{d}\mathbf{x}, \quad (1)$$

where we set $G_{\infty,n}(\mathbf{z}) := \mathbf{E}[G_n([X_{\ell,\infty}]_\ell + \mathbf{z})] = \mathbf{E}[G_n([X_\ell]_\ell + \mathbf{z})]$ for $k = \infty$ in the natural way. The reduction principle consists of showing that

$$\begin{aligned}
&\sum_{t=1}^{n-T+1} (G_n([X_{t+\ell}]_\ell) \; - \; \mathbf{E}[G_n([X_\ell]_\ell)]) \\
&= \sum_{t=1}^{n-T+1} \sum_{k=0}^{\infty} \left(\mathbf{E}\left[G_n([X_{t+\ell}]_\ell) \mid \mathcal{F}_{t-k}\right] \; - \; \mathbf{E}\left[G_n([X_{t+\ell}]_\ell) \mid \mathcal{F}_{t-(k+1)}\right]\right) \\
&= \sum_{t=1}^{n-T+1} \sum_{k=0}^{\infty} (G_{k-1,n}([X_{t+\ell} - X_{t+\ell,k+\ell-1}]_\ell) \; - \; G_{k,n}([X_{t+\ell} - X_{t+\ell,k+\ell}]_\ell)) \\
&\approx \sum_{t=1}^{n-T+1} \sum_{k=0}^{\infty} (G_{k,n}([X_{t+\ell} - X_{t+\ell,k+\ell-1}]_\ell) \; - \; G_{k,n}([X_{t+\ell} - X_{t+\ell,k+\ell}]_\ell)) \qquad (2) \\
&\approx \sum_{t=1}^{n-T+1} \sum_{k=0}^{\infty} \varepsilon_{t-k} \cdot [a_{k+\ell}]_\ell^\top \cdot \nabla G_{k,n}([X_{t+\ell} - X_{t+\ell,k+\ell}]_\ell) \\
&= \sum_{t=1}^{n-T+1} [X_{t+\ell}]_\ell^\top \cdot \nabla G_{\infty,n}(\mathbf{0}_T) \\
&\quad + \sum_{t=1}^{n-T+1} \sum_{k=0}^{\infty} \varepsilon_{t-k} \cdot [a_{k+\ell}]_\ell^\top \left(\nabla G_{k,n}([X_{t+\ell} - X_{t+\ell,k+\ell}]_\ell) \; - \; \nabla G_{\infty,n}(\mathbf{0}_T)\right) .
\end{aligned}$$

The first term is the leading term that will determine the asymptotic profile, described in the next theorem.

Theorem 3.1 (Central limit theorem for multivariate partial sums)**.** *Let Assumptions 1 and 2 hold. Then*

$$n^{-d-1/\alpha} \sum_{t=1}^{n-T+1} [X_{t+\ell}]_\ell \;\to_d\; Z_\alpha \cdot \mathbf{1}_T \qquad \text{as } n \to \infty ,$$

where

$$Z_\alpha \quad \text{has distribution} \quad \begin{cases} \mathcal{N}(0, \sigma^2) & \text{if } \alpha = 2 \text{ with } \mathbf{E}[\varepsilon^2] < \infty, \\ S\alpha S(\eta) & \text{if } \alpha \in (1, 2) , \end{cases}$$

with variance

$$\sigma^2 := c_a^2 \mathbf{E}[\varepsilon^2] \frac{\Gamma(1-2d)\Gamma(d)}{d(2d+1)\Gamma(1-d)} \quad \text{when } \mathbf{E}[\varepsilon^2] < \infty,$$

or scale

$$\eta := \frac{c_a}{d} \cdot \left(L \cdot \frac{\Gamma(2-\alpha)\cos(\pi\alpha/2)}{1-\alpha} \right)^{1/\alpha} \left(\int_{-\infty}^{1} \left((1-v)_+^d - (-v)_+^d \right)^\alpha \mathrm{d}v \right)^{1/\alpha} .$$

Here, $S\alpha S(\eta)$ denotes the symmetric α stable distribution with characteristic function $x \mapsto \exp(-\eta^\alpha |x|^\alpha)$.

Having determined the asymptotic profile of the linear approximation, we make the reduction principle in (2) rigorous.

Theorem 3.2. *Let Assumptions 1 and 2 hold, and let either of the following conditions hold:*

(i) A is bounded away from 0, $s_1 \in (\underline{s}, \infty)$ for some $\underline{s} > 0$, and $s_2 = \infty$

(ii) $A \in \mathcal{C}_T$ and, for some $0 < \underline{s} < 1 < \overline{s} < \infty$, $s_1, s_2 \in (\underline{s}, \overline{s}]$ satisfy $s_1 < s_2$

Then for all $\delta > 0$ *and for* $n \in \mathbb{N}$ *,*

$$\mathbf{E}\left[\left|\sum_{t=1}^{n-T+1} G_n([X_{t+\ell}]_\ell, s_1, s_2, A) \;-\; \mathbf{E}[G_n([X_{t+\ell}]_\ell, s_1, s_2, A)] \;-\; \left(\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=\mathbf{0}_T}\right)^\top \cdot [X_{t+\ell}]_\ell\right|^{r_0}\right]$$
$$\lesssim \; ((s_2 - s_1) \wedge 1) \cdot n^{r_0(\kappa_0 + \delta/2)} ,$$

where

$$r_0 \;:=\; \begin{cases} \alpha(1-d) \;\; \textit{if} \;\; \dfrac{1}{(1-d)(1-2d)} < \alpha , \\ \dfrac{1}{2}\left(\alpha + \dfrac{1}{1-d}\right) \;\; \textit{otherwise,} \end{cases}$$

and

$$\kappa_0 \;=\; \begin{cases} \dfrac{1}{\alpha(1-d)} \;\; \textit{if} \;\; \dfrac{1}{(1-d)(1-2d)} < \alpha , \\ \dfrac{2(1-d) + (1 - \alpha(1-d)(1-2d))}{\alpha(1-d)+1} = d + (1-d)\dfrac{3 - \alpha(1-d)}{\alpha(1-d)+1} \;\; \textit{otherwise.} \end{cases} \tag{3}$$

As explained in Remark 3.4 in Scheffel et al. (2026), $\kappa_0 - d - 1/\alpha < 0$, so that there is a choice of $u_n \to \infty$ such that $u_n = o(n^{(d+1/(\nu\wedge 2)-\kappa_0-\delta)/(\nu+1)})$ for $\delta > 0$ small enough. Therefore,

$$\left(n^{d+1/\alpha} f_{X_0}(u_n)\right)^{-1} \cdot n^{\kappa_0+\delta/2} \;=\; \frac{n^{\kappa_0+\delta-d-1/\alpha}}{f_{X_0}(u_n)} \cdot n^{-\delta/2} \;\to\; 0 . \tag{4}$$

By Theorem 3.2 applied with $s_1 = s$ and $s_2 = \infty$ and a set A that is bounded away from 0, this proves the following result.

Corollary 3.3. *Let Assumptions 1, 2, and 3 hold, and assume that the threshold sequence* $u_n \to \infty$ *is such that* $u_n = o(n^{(d+1/(\nu\wedge 2)-\kappa_0-\delta)/(\nu+1)})$ *for some* $\delta > 0$ *. Then, for all sets* A *bounded away from* 0 *, and all* $s \in (0, \infty)$

$$\left|\frac{n^{-d-1/\alpha}}{f_{X_0}(u_n)} \sum_{t=1}^{n-T+1} \left(\mathbb{1}\{[X_{t+\ell}]_\ell \in u_n \cdot s \cdot A\} \;-\; \mathbf{P}\left[[X_{t+\ell}]_\ell \in u_n \cdot s \cdot A\right] \right.\right.$$
$$\left.\left. -\; \left(\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s, \infty, A)\,|_{\mathbf{y}=\mathbf{0}_T}\right)^\top [X_{t+\ell}]_\ell\right)\right|$$

vanishes in probability as $n \to \infty$ *.*

3.2. Uniform reduction principle

As explained in Section 2.3, we aim for a reduction principle that is uniform in s and therefore allows u_n to be replaced by a random threshold. Before we state the result, we give an overview of how it is derived by combining Theorem 3.2 under Condition (ii) therein, and a chaining technique as in (Koul and Surgailis, 2001, Proof of Theorem 2.1). We use, for $k \in \mathbb{N}_0$ and $j \in \{0, \ldots, 2^k\}$, a suitable partition $(\pi_{j,k})$ satisfying

$$\underline{s} \;=\; \pi_{2^k,k} \;<\; \pi_{2^k-1,k} \;<\; \cdots \;<\; \pi_{1,k} \;<\; \pi_{0,k} \;=\; \overline{s} ,$$

and define, for $k \in \mathbb{N}_0$ and $s \in (\underline{s}, \overline{s}]$, the unique index $j_k^s \in \{0, \ldots, 2^k - 1\}$ satisfying

$$\pi_{j_k^s+1,k} \;<\; s \;\leq\; \pi_{j_k^s,k} .$$

Choosing $K = K_n$, this gives a chain

$$\pi_{j_K^s+1,K} \;<\; s \;\leq\; \pi_{j_K^s,K} \;\leq\; \cdots \;\leq\; \pi_{j_1^s,1} \;\leq\; \pi_{j_0^s,0} \;=\; \overline{s}\,,$$

linking $\overline{s}$ to s, where we tighten the chain by choosing $K_n \to \infty$ at a suitable rate. Write

$$\mathcal{S}_n(s_1,s_2,A) \;=\; \sum_{t=1}^{n-T+1} \Bigg(G_n([X_{t+\ell}]_\ell, s_1, s_2, A) \;-\; \mathbf{E}[G_n([X_{t+\ell}]_\ell, s_1, s_2, A)] \\ -\; \left(\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=\mathbf{0}_T} \right)^{\top} \cdot [X_{t+\ell}]_\ell \Bigg)\,,$$

and decompose

$$\begin{aligned} \mathcal{S}_n(s,\infty,A) \;&=\; \mathcal{S}_n(s,\overline{s},A) \;+\; \mathcal{S}_n(\overline{s},\infty,A) \\ &=\; \sum_{k=1}^{K} \mathcal{S}_n(\pi_{j_k^s,k}, \pi_{j_{k-1}^s,k-1}, A) \;+\; \mathcal{S}_n(s, \pi_{j_K^s,K}, A) \;+\; \mathcal{S}_n(\overline{s},\infty,A)\,, \end{aligned}$$

so that

$$\begin{aligned} &\sup_{s\in(\underline{s},\overline{s})} |\mathcal{S}_n(s,\infty,A)| \\ &\leq\; \sum_{k=1}^{K} \sup_{s\in(\underline{s},\overline{s})} \left|\mathcal{S}_n(\pi_{j_k^s,k}, \pi_{j_{k-1}^s,k-1})\right| \;+\; \sup_{s\in(\underline{s},\overline{s})} \left|\mathcal{S}_n(s, \pi_{j_K^s,K})\right| \;+\; |\mathcal{S}_n(\overline{s},\infty,A)| \\ &\approx\; \sum_{k=1}^{K} \sum_{0\leq i\leq 2^k-1} |\mathcal{S}_n(\pi_{i+1,k}, \pi_{i,k}, A)| \;+\; |\mathcal{S}_n(\overline{s},\infty,A)|\,. \end{aligned}$$

By Theorem 3.2,

$$\begin{aligned} \sum_{k=1}^{K} \sum_{0\leq i\leq 2^k-1} \mathbf{E}\left[|\mathcal{S}_n(\pi_{i+1,k}, \pi_{i,k}, A)|^{r_0}\right] \;&\lesssim\; n^{r_0(\kappa_0+\delta/2)} \sum_{k=1}^{K} \sum_{0\leq i\leq 2^k-1} (\pi_{i,k} \;-\; \pi_{i+1,k}) \\ &=\; n^{r_0(\kappa_0+\delta/2)} \cdot K_n \cdot (\overline{s} - \underline{s})\,, \end{aligned}$$

and $\mathbf{E}[|\mathcal{S}_n(\overline{s},\infty,A)|^{r_0}] \lesssim n^{r_0(\kappa_0+\delta/2)}$. Choosing $K_n \to \infty$ appropriately, it follows from (4) that

$$\left(\frac{n^{-d-1/\alpha}}{f_{X_0}(u_n)} \cdot n^{\kappa_0+\delta/2} \right)^{r_0} \cdot K_n \cdot (\overline{s} - \underline{s}) \;\to\; 0\,,$$

so that

$$\frac{n^{-d-1/\alpha}}{f_{X_0}(u_n)} \sup_{s\in(\underline{s},\overline{s})} |\mathcal{S}_n(s,\infty,A)| \;\to\; 0\,,$$

in outer probability. This leads to the following uniform reduction principle.

Theorem 3.4. *Let Assumptions 1, 2, and 3 hold and assume that the threshold sequence $u_n \to \infty$ is such that $u_n = o(n^{(d+1/(\nu\wedge 2)-\kappa_0-\delta)/(\nu+1)})$ for some $\delta > 0$. Then, for all $A \in \mathcal{C}_T$, and all $\underline{s} \in (0,1)$ and $\overline{s} \in (1,\infty)$,*

$$\sup_{s\in(\underline{s},\overline{s})} \Bigg| \frac{n^{-d-1/\alpha}}{f_{X_0}(u_n)} \sum_{t=1}^{n-T+1} \Bigg(\mathbb{1}\{[X_{t+\ell}]_\ell \in u_n \cdot s \cdot A\} \;-\; \mathbf{P}\left[[X_{t+\ell}]_\ell \in u_n \cdot s \cdot A\right] \\ -\; \left(\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s, \infty, A)\Big|_{\mathbf{y}=\mathbf{0}_T} \right)^{\top} [X_{t+\ell}]_\ell \Bigg) \Bigg|$$

vanishes in outer probability as $n \to \infty$.

Remark 3.5 (Additional assumptions about sets in Theorem 3.4)**.** We already see in Corollary 3.6 in Drees that additional assumptions on the class of sets are needed to accomplish the transition from deterministic to random thresholds. We introduce these assumptions in a transparent way, showing in Example 2.5 that the restricted class $\mathcal{C}_T$ is sufficiently large from a statistical standpoint.

In the next section, we derive central limit theory from the results developed so far.

4. Central limit theory for extremogram-type tail-dependence estimators

We now apply our reduction principles to establish central limit theorems. We provide proofs of the results in Section G of the supplementary material.

4.1. Central limit theory for empirical tail measures

Let $B \subset \mathbb{R}$ be a Borel set bounded away from 0, whose boundary has Lebesgue measure 0. Then, by regular variation and symmetry of ε,

$$\frac{\mathbf{P}[\varepsilon \in u_n \cdot B]}{\mathbf{P}[|\varepsilon| > u_n]} \;\to\; \mu_\varepsilon[B] \;:=\; \int_B \frac{\nu}{2}|z|^{-1-\nu}\,\mathrm{d}z\,.$$

Let $A \subset \mathbb{R}^T$ be bounded away from 0 such that for all $j \geq -(T-1)$,

$$\mu_\varepsilon\left[\{x \in \mathbb{R} \mid x \cdot [a_j, \ldots, a_{j+T-1}] \in \partial A\}\right] \;=\; 0\,,$$

where ∂A is the boundary of the set A, and define the tail measure at lag $T-1$ by

$$\mu_T(A) \;=\; \frac{1}{\sum_{i=0}^{\infty}|a_i|^\nu} \sum_{j=-(T-1)}^{\infty} \mu_\varepsilon[\{x \in \mathbb{R} \mid x \cdot [a_j, \ldots, a_{j+T-1}] \in A\}]\,.$$

Since A is a continuity set of μ_T that is bounded away from 0, it holds for all $s > 0$ by Proposition 5.2.8 in Mikosch and Wintenberger (2024),

$$\frac{\mathbf{P}[[X_\ell]_\ell \in u_n \cdot s \cdot A]}{\mathbf{P}[|X_0| > u_n]} \;\to\; \mu_T(s \cdot A) \;=\; s^{-\nu} \cdot \mu_T(A) \qquad \text{as } n \to \infty\,.$$

This convergence holds locally uniformly in s, due to local uniformity of univariate regular variation. For $s > 0$ denote the empirical tail measure and its expectation as

$$\widehat{\mathbf{P}}_n(s, A) \;:=\; \frac{1}{n-T+1} \sum_{t=1}^{n-T+1} \frac{\mathbb{1}\{[X_{t+\ell}]_\ell \in u_n \cdot s \cdot A\}}{\mathbf{P}[|X_0| > u_n]} \qquad \text{and} \qquad \mathbf{P}_n(s, A) \;:=\; \frac{\mathbf{P}\left[[X_\ell]_\ell \in u_n \cdot s \cdot A\right]}{\mathbf{P}[|X_0| > u_n]}\,.$$

4.1.1. Result for deterministic thresholds

Setting $s = 1$, the reduction principle tells us that the asymptotic profile is determined by

$$\left(\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, 1, \infty, A)\,|_{\mathbf{y}=\mathbf{0}_T}\right)^\top \sum_{t=1}^{n-T+1} [X_{t+\ell}]_\ell\,,$$

where we see from Theorem 3.1 that the correct rate of convergence for the multivariate partial sum is $n^{-d-1/\alpha}$. Since $\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, 1, \infty, A)\Big|_{\mathbf{y}=\mathbf{0}_T}$, defined in (1), is an integral of an integrable function over an asymptotically vanishing domain, it generally converges to $\mathbf{0}_T$ as $n \to \infty$. Therefore, we need an assumption on its rate of decay to identify the limit. We verify this assumption in a non-trivial setting in Theorem 5.3.

Assumption 4 (Convergence of scaled gradient)**.** For the set A there exists $H_A(1) \in \mathbb{R}^T$ such that

$$\left| \frac{\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, 1, \infty, A)\Big|_{\mathbf{y}=\mathbf{0}_T}}{f_{X_0}(u_n)} \; - \; H_A(1) \right| \;\to\; 0 \qquad \text{as } n \to \infty \,.$$

Remark 4.1 (Assumption 4)**.** Later we allow s to vary around 1 in $\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s, \infty, A)\Big|_{\mathbf{y}=\mathbf{0}_T}$ which leads to a similar assumption but with a vector-valued function $H_A : s \mapsto H_A(s)$ instead of a constant vector $H_A(1)$.

Theorem 4.2 (Empirical tail measure with deterministic thresholds)**.** *Let Assumptions 1, 2, and 3 hold, let $A_0, \dots, A_h$, $h \in \mathbb{N}_0$, be bounded away from 0, and let $u_n \to \infty$ with $u_n = o(n^{(d+1/(\nu\wedge 2)-\kappa_0-\delta)/(\nu+1)})$ for some $\delta > 0$. If $A_0, \dots, A_h$ satisfy Assumption 4 with $H_{A_0}(1), \dots, H_{A_h}(1)$, then*

$$n^{1-d-1/\alpha} u_n \left[\widehat{\mathbf{P}}_n(1, A_i) \; - \; \mathbf{P}_n(1, A_i) \right]_{i=0,\dots,h} \;\to_d\; \frac{\nu}{2} Z_\alpha [H_{A_i}(1)^\top \mathbf{1}_T]_{i=0,\dots,h} \qquad \textit{as } n \to \infty \,.$$

Remark 4.3 (Increased speed in the extreme value setting)**.** Compared to standard central limit theory for long memory linear time series, the increased speed $n^{1-d-1/\alpha} u_n$ has already been observed in the univariate setting of Scheffel et al. (2026), see Remark 3.7 therein.

4.1.2. Heuristic for the transition to sample quantiles as thresholds

A consequence of the uniform reduction principle stated in Theorem 3.4 is that if $A_0, \dots, A_h \in \mathcal{C}_T$ satisfy a uniform version of Assumption 4 (see Assumption 7) and if $(Y_n^{(i)})$, for $i \in \{0, \dots, h\}$, are random sequences that satisfy $Y_n^{(i)}/u_n \to_{\mathbf{P}} 1$ as $n \to \infty$, we have

$$\begin{aligned} n^{1-d-1/\alpha} u_n & \left[\widehat{\mathbf{P}}_n(Y_n^{(i)}/u_n, A_i) \; - \; \mathbf{P}_n(s, A_i)\Big|_{s=Y_n^{(i)}/u_n} \right]_{i=0,\dots,h} \\ & \to_d \; \frac{\nu}{2} Z_\alpha [H_{A_i}(1)^\top \mathbf{1}_T]_{i=0,\dots,h} \qquad \text{as } n \to \infty \,; \end{aligned} \tag{5}$$

see Theorem G.3 in the supplementary material for a rigorous statement. From Lemma C.1 and Lemma D.1 in Scheffel et al. (2026) we know that $Y_n^{(i)} = Y_n = X_{n-k:n}$ is a valid choice when u_n is the marginal quantile $q_{X_0}(1 - k/n)$, under suitable assumptions on the growth of $k = k_n \to \infty$. To obtain a version of this convergence with centering $\mathbf{P}_n(1, A_i)$ instead of $\mathbf{P}_n(s, A_i)\Big|_{s=X_{n-k:n}/u_n}$, that is, with deterministic rather than random centering, we use the following Taylor approximation:

$$\begin{aligned} & \widehat{\mathbf{P}}_n(X_{n-k:n}/u_n, A_i) \; - \; \mathbf{P}_n(1, A_i) \\ & = \; \widehat{\mathbf{P}}_n(X_{n-k:n}/u_n, A_i) \; - \; \mathbf{P}_n(s, A_i)\Big|_{s=X_{n-k:n}/u_n} \\ & \qquad + \; \mathbf{P}_n(s, A_i)\Big|_{s=X_{n-k:n}/u_n} \; - \; \mathbf{P}_n(1, A_i) \\ & \approx \; \widehat{\mathbf{P}}_n(X_{n-k:n}/u_n, A_i) \; - \; \mathbf{P}_n(s, A_i)\Big|_{s=X_{n-k:n}/u_n} \\ & \qquad + \; \frac{\partial}{\partial s} \mathbf{P}_n(s, A_i)\Big|_{s=1} \cdot \left(\frac{X_{n-k:n}}{u_n} \; - \; 1 \right) . \end{aligned} \tag{6}$$

Following Lemma C.2 in Scheffel et al. (2026), we can analyze the joint convergence of these terms for $i \in \{0, \dots, h\}$ as follows: let $s_0, \dots, s_h, s^* \in \mathbb{R}$ and write

$$E_n(s_0, \dots, s_h) \; := \; \bigcap_{i=0}^{h} \left\{ n^{1-d-1/\alpha} u_n \left(\widehat{\mathbf{P}}_n(X_{n-k:n}/u_n, A_i) \; - \; \mathbf{P}_n(s, A_i)\Big|_{s=X_{n-k:n}/u_n} \right) \; \le \; s_i \right\} .$$

Then one can show that

$$\begin{aligned}&\mathbf{P}\left[E_n(s_0,\ldots,s_h)\,,\quad n^{1-d-1/\alpha}u_n\left(\frac{X_{n-k:n}}{u_n}\;-\;1\right)\;\leq\;s^*\right]\\&=\;\mathbf{P}\left[E_n(s_0,\ldots,s_h)\,,\quad \frac{n^{1-d-1/\alpha}u_n}{\nu/2}\left(\widehat{\mathbf{P}}_n(1,B)\;-\;\mathbf{P}_n(1,B)\right)\;\leq\;s^*\cdot(1+o(1))+o_{\mathbf{P}}(1)\right]\end{aligned} \tag{7}$$

for a suitable set B. This allows us to derive the joint limit using a suitable version of the convergence in (5) with $Y_n^{(i)} = X_{n-k:n}$ for $i \in \{0,\ldots,h\}$ and $Y_n^{(h+1)} = u_n$.

4.1.3. Assumptions

Having established the joint convergence, the approximation in (6) motivates us to study also $\frac{\partial}{\partial s}\mathbf{P}_n(s,A_i)\Big|_{s=1}$ which comes from the derivative of $G_{\infty,n}(\mathbf{y},s,\infty,A)$ with respect to the scaling variable s.

Assumption 5. $(x \mapsto x\cdot f_\varepsilon'(x)) \in L^1(\mathbb{R})$.

Remark 4.4 (Assumption 5)**.** This holds if f_ε' is regularly varying, since then the index of regular variation is $-2-\nu$ by Assumption 1. This is a very weak assumption, which in turn guarantees, by Lemma B.3(ii) in the supplementary material, the existence and continuity of $\frac{\partial}{\partial s}\mathbf{P}_n(s,A)$, together with the identity

$$\frac{\partial}{\partial s}\mathbf{P}_n(s,A)\;=\;\frac{1}{s}\left(T\frac{\mathbf{P}[[X_\ell]_\ell\in u_n\cdot s\cdot A]}{\mathbf{P}[|X_0|>u_n]}\;+\;\int_{\mathbb{R}^T}\mathbb{1}\{\mathbf{x}\in u_n\cdot s\cdot A\}\cdot\frac{\mathbf{x}^\top\cdot\nabla f_{[X_\ell]_\ell}(\mathbf{x})}{\mathbf{P}[|X_0|>u_n]}\,\mathrm{d}\mathbf{x}\right). \tag{8}$$

For rectangular sets as in Example 2.5(iii), existence and continuity of the derivative are readily obtained by partial integration without the need of Assumption 5. This, however, is not true for sets having a more complicated geometry like, for example, intersections with scale invariant sets that are covered in Example 2.5(v).

While Assumption 5 guarantees the existence of $\frac{\partial}{\partial s}\mathbf{P}_n(s,A)$ with identity (8), we also need assumptions that control its convergence as $n\to\infty$. The following assumption is on the convergence of the second term in (8), while the convergence of the first term in (8) follows immediately from multivariate regular variation.

Assumption 6 (Convergence of the second term in (8))**.** The set A is a continuity set of μ_T and there exists a compact neighborhood $1\in N_A\subset(0,\infty)$ and a function $p_A\colon N_A\to\mathbb{R}$ such that

$$\int_{\mathbb{R}^T}\mathbb{1}\{\mathbf{x}\in u_n\cdot s\cdot A\}\cdot\frac{\mathbf{x}^\top\nabla f_{[X_\ell]_\ell}(\mathbf{x})}{\mathbf{P}[|X_0|>u_n]}\,\mathrm{d}\mathbf{x}\;\to\;s^{-\nu}\left(\frac{\nu}{2}\cdot p_A(s)\;-\;T\mu_T(A)\right)$$

uniformly in $s\in N_A$ as $n\to\infty$.

Remark 4.5 (Convergence in (8))**.** This assumption guarantees the convergence

$$\frac{\partial}{\partial s}\mathbf{P}_n(s,A)\;\to\;s^{-1-\nu}\left(T\cdot\mu_T(A)\;+\;\frac{\nu}{2}\cdot p_A(s)\;-\;T\cdot\mu_T(A)\right)\;=\;\frac{\nu}{2}\cdot s^{-1-\nu}\cdot p_A(s)$$

uniformly in a neighborhood of 1 as $n\to\infty$. The factor $\nu/2$ appears in the joint limit derived from (7), so Assumption 6 is tailored to give a simple formula. Note that, by Lemma B.3(ii) in the supplementary material and local uniformity of regular variation, p_A is continuous on N_A.

As highlighted in Remark 4.1 and before (5), we now state a uniform version of Assumption 4.

Assumption 7 (Uniform version of Assumption 4)**.** For the set A there exists a compact neighborhood $1 \in N_A \subset (0,\infty)$ and a function $H_A \colon N_A \to \mathbb{R}^T$ such that

$$\sup_{s \in N_A} \left| \frac{\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s, \infty, A)\Big|_{\mathbf{y}=\mathbf{0}_T}}{f_{X_0}(s \cdot u_n)} \; - \; H_A(s) \right| \;\to\; 0 \qquad \text{as } n \to \infty\,.$$

Remark 4.6 (Continuity of the limit in Assumption 7)**.** The continuity of $(\mathbf{y}, s) \mapsto \nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s, \infty, A)$ obtained from Lemma B.3(ii) in the supplementary material, together with the uniform convergence in N_A guarantees that H_A is continuous on N_A.

Remark 4.7 (Simplification and validity of Assumptions 6 and 7 for rectangular sets)**.** In Proposition 5.2 we show for rectangular sets that Assumptions 6 and 7 are implied by a simpler assumption, which is then verified in Theorem 5.3 in a non-trivial setting.

4.1.4. Results for sample quantiles as thresholds

The next theorem is the analog of Theorem 4.2 with sample quantiles as thresholds. There we observe a speed increase by the factor u_n compared with the classical long memory rate $n^{-d-1/\alpha}$; see Remark 4.3. In the next result, we have the same increased speed, where we replace the factor u_n by $q_{X_0}(1-k/n)$, with $k = \lfloor n \cdot \mathbf{P}[X_0 > u_n] \rfloor$; note that by regular variation of X_0, $u_n / q_{X_0}(1-k/n) \to 1$ as $n \to \infty$.

Theorem 4.8 (Empirical tail measure with sample quantiles as thresholds)**.** *Let Assumptions 1, 2, and 3 hold, let $A_0, \dots, A_h$, $h \in \mathbb{N}_0$, be continuity sets of μ_T bounded away from 0, and let $u_n \to \infty$ with $u_n = o(n^{(d+1/(\nu\wedge 2)-\kappa_0-\delta)/(\nu+1)})$ for some $\delta > 0$. If Assumption 5 holds, and if $A_0, \dots, A_h$ are in $\mathcal{C}_T$ and satisfy Assumption 6 and 7 with functions $p_{A_0}, \dots, p_{A_h}$, and $H_{A_0}, \dots, H_{A_h}$, respectively, then, for $k = \lfloor n \cdot \mathbf{P}[X_0 > u_n] \rfloor$,*

$$\begin{aligned} &n^{1-d-1/\alpha} q_{X_0}\left(1 - \frac{k}{n}\right) \left[\widehat{\mathbf{P}}_n(X_{n-k:n}/u_n, A_i) \; - \; \mathbf{P}_n(1, A_i)\right]_{i=0,\dots,h} \\ &\to_d \; \frac{\nu}{2} Z_\alpha \left[(H_{A_i}(1))^\top \mathbf{1}_T \; + \; p_{A_i}(1)\right]_{i=0,\dots,h} \qquad \text{as } n \to \infty\,. \end{aligned}$$

Remark 4.9 (Different limiting behavior for random and deterministic thresholds)**.** It may happen that $H_A(1)^\top \mathbf{1}_T = -p_A(1)$ while $H_A(1)^\top \mathbf{1}_T \neq 0$. Then, in the setting of deterministic thresholds (Theorem 4.2), we get a non-degenerate limit, while sample quantiles as thresholds (Theorem 4.8) give a degenerate limit at the same rate. In Section 5.2, we show that this happens for some rectangular sets that are relevant in practice.

4.2. Central limit theory for the empirical extremogram

The final result is for extremogram estimators similar to those of Section 3.3 in Davis and Mikosch (2009).

Theorem 4.10 (Central limit theorems for empirical extremogram with deterministic and random thresholds)**.** *Let Assumptions 1, 2, and 3 hold, let $A_0, \dots, A_h$, $h \in \mathbb{N}_0$, be continuity sets of μ_T bounded away from 0, with $\mu_T(A_0) > 0$, and let $u_n \to \infty$ with $u_n = o(n^{(d+1/(\nu\wedge 2)-\kappa_0-\delta)/(\nu+1)})$ for some $\delta > 0$.*

(i) If $A_0, \dots, A_h$ *satisfy Assumption 4 with* $H_{A_0}(1), \dots, H_{A_h}(1)$ *,*

$$n^{1-d-1/\alpha} u_n \cdot \left[\frac{\widehat{\mathbf{P}}_n(1, A_i)}{\widehat{\mathbf{P}}_n(1, A_0)} - \frac{\mathbf{P}_n(1, A_i)}{\mathbf{P}_n(1, A_0)} \right]_{i=1,\dots,h}$$
$$\to_d \frac{\nu}{2} Z_\alpha \left[\frac{(\mu_T(A_0) \cdot H_{A_i}(1) - \mu_T(A_i) \cdot H_{A_0}(1))^\top \mathbf{1}_T}{(\mu_T(A_0))^2} \right]_{i=1,\dots,h} \qquad \text{as } n \to \infty .$$

(ii) If Assumption 5 holds, and if $A_0, \dots, A_h$ *are in* $\mathcal{C}_T$ *and satisfy Assumption 6 and 7 with functions* $p_{A_0}, \dots, p_{A_h}$ *and* $H_{A_0}, \dots, H_{A_h}$ *, respectively, then, for* $k = \lfloor n \cdot \mathbf{P}[X_0 > u_n] \rfloor$ *,*

$$n^{1-d-1/\alpha} q_{X_0} \left(1 - \frac{k}{n}\right) \left[\frac{\widehat{\mathbf{P}}_n(X_{n-k:n}/u_n, A_i)}{\widehat{\mathbf{P}}_n(X_{n-k:n}/u_n, A_0)} - \frac{\mathbf{P}_n(1, A_i)}{\mathbf{P}_n(1, A_0)} \right]_{i=1,\dots,h}$$
$$\to_d \frac{\nu/2}{(\mu_T(A_0))^2} Z_\alpha \Big[\mu_T(A_0) \Big(H_{A_i}(1)^\top \mathbf{1}_T + p_{A_i}(1) \Big)$$
$$- \mu_T(A_i) \Big(H_{A_0}(1)^\top \mathbf{1}_T + p_{A_0}(1) \Big) \Big]_{i=1,\dots,h} \qquad \text{as } n \to \infty .$$

Remark 4.11 (Comparison with existing literature)**.** Comparing the proofs of Corollary 3.4 in Davis and Mikosch (2009) and Theorem 4.10 in our work, we see that both feature the term $\mathbf{F} S_n$, where

$$\mathbf{F} = \Big[\mu_T(A_0) \cdot \mathbf{I}_h \quad [-\mu_T(A_i)]_{i=1,\dots,h} \Big]$$

(compare with $\mathbf{F}$ in Equation (3.18) in Davis and Mikosch (2009)), and

$$S_n = n^{1-d-1/\alpha} u_n \begin{bmatrix} \Big[\widehat{\mathbf{P}}_n(1, A_i) - \mathbf{P}_n(1, A_i) \Big]_{i=1,\dots,h} \\ \widehat{\mathbf{P}}_n(1, A_0) - \mathbf{P}_n(1, A_0) \end{bmatrix}$$

(compare with S_n in Equation (3.14) in Davis and Mikosch (2009)).

While Theorem 3.2 in Davis and Mikosch (2009) gives a non-degenerate multivariate Gaussian limit for their version of S_n , Theorem 4.2 in our work gives, for our version of S_n , a stable random vector with linearly dependent components. Therefore, in our long memory setting, the randomness in the limit comes from one random variable Z_α , rather than from a genuinely multivariate random vector. Since the theory of Davis and Mikosch (2009) is restricted to deterministic thresholds, we now compare Theorem 4.10(ii) (long memory—random thresholds) with Corollary 3.6 in Drees(short memory—random thresholds). In the short memory case, limits for random and deterministic thresholds are generally the same, while the limits in (i) and (ii) of Theorem 4.10 in our work can markedly differ; see Remark 4.9.

5. Theory for rectangular sets under random thresholds

Throughout this section we consider $A = \times_{j=1}^{T} (\mathfrak{a}_j, \infty)$ with $\{\mathfrak{a}_j\} \subset (0, \infty) \cup \{-\infty\}$ and $\emptyset \neq J_A = \{j \mid \mathfrak{a}_j > 0\} \subset \{1, \dots, T\}$. By Example 2.5(iii), $A \in \mathcal{C}_T$. Proofs of the results of this section are provided in Section H of the supplementary material.

5.1. Simplified assumptions

We show that our theory for random thresholds is applicable for rectangular sets A by providing simplified assumptions and showing their validity; see Remark 4.7.

Assumption 8. For the set A there exist $(p_{j,A})_{j\in J_A} \subset [0,1]$, such that

$$\mathbf{P}\left[[X_\ell]_\ell \in u\cdot A \mid X_{j-1} = u\cdot\mathfrak{a}_j\right] \;\to\; p_{j,A} \qquad \text{as } u\to\infty\,. \tag{9}$$

Remark 5.1 (Comparison with multivariate regular variation I)**.** Replacing the condition

$$X_{j-1} \;=\; u\cdot\mathfrak{a}_j \qquad \text{by} \qquad X_{j-1} \;>\; u\cdot\mathfrak{a}_j\,,$$

classical multivariate regular variation yields

$$\begin{aligned}\mathbf{P}\left[[X_\ell]_\ell \in u\cdot A \mid X_{j-1} > u\cdot\mathfrak{a}_j\right] \;&=\; \mathbf{P}\left[\forall i\in J_A,\ X_{i-1} > (u\cdot\mathfrak{a}_j)\,\frac{\mathfrak{a}_i}{\mathfrak{a}_j} \,\middle|\, X_{j-1} > u\cdot\mathfrak{a}_j\right]\\ \to\; 2\cdot\mu_T&\left[\bigtimes_{i=1}^{T}\left(\frac{\mathfrak{a}_i}{\mathfrak{a}_j},\infty\right)\right] \qquad \text{as } u\to\infty\,.\end{aligned}$$

The next result shows that Assumption 8 implies Assumptions 6 and 7, and is therefore sufficient for the application of Theorem 4.10.

Proposition 5.2 (Simplification of Assumptions 6 and 7)**.** *Let Assumptions 1, 2, 3, and 5 hold and let A be a continuity set of μ_T. Then Assumption 8 implies Assumptions 6 and 7, with*

$$p_A(s) \;=\; -\sum_{j\in J_A}\mathfrak{a}_j^{-\nu}p_{j,A} \qquad \textit{and} \qquad H_A(s) \;=\; \sum_{j\in J_A}\mathfrak{a}_j^{-(\nu+1)}\cdot p_{j,A}\cdot\mathbf{e}_j\,.$$

The next result verifies Assumption 8 in a non-trivial setting.

Theorem 5.3 (Validity of Assumption 8)**.** *Let $T=2$, $A=(1,\infty)\times(\mathfrak{a}_1,\infty)$, with $\mathfrak{a}_1\in[1,\infty)$. Assume that $a_0 = 1$, and that $(a_i)_{i\in\mathbb{N}}$ is positive and strictly decreasing. If Assumptions 1 and 2 hold, and the distribution of ε is unimodal, then Assumption 8 holds with*

$$p_{1,A} \;=\; 0 \qquad \textit{and} \qquad p_{2,A} \;=\; \frac{\sum_{i=1}^\infty a_i^\nu}{\sum_{i=0}^\infty a_i^\nu} \;=\; 1 \;-\; \frac{1}{\sum_{i=0}^\infty a_i^\nu}\,,$$

and Theorem 4.10 applies.

Remark 5.4 (Comparison with multivariate regular variation II)**.** We continue the comparison of Remark 5.1 under the assumptions of Theorem 5.3. Write $\mathfrak{a}_0 = 1$ and let $\sigma:\{0,1\}\to\{0,1\}$ be a permutation. Then

$$\begin{aligned}\mathbf{P}[X_{\sigma(1)} > \mathfrak{a}_{\sigma(1)}\cdot u \mid X_{\sigma(0)} > \mathfrak{a}_{\sigma(0)}\cdot u] \;&=\; \frac{\mathbf{P}[X_0 > u\,, X_1 > \mathfrak{a}_1\cdot u]}{\mathbf{P}[X_0 > \mathfrak{a}_{\sigma(0)}\cdot u]}\\ =\; 2\frac{\mathbf{P}[X_0>u]}{\mathbf{P}[X_0 > \mathfrak{a}_{\sigma(0)}\cdot u]}&\frac{\mathbf{P}[X_0 > u\,, X_1 > \mathfrak{a}_1\cdot u]}{\mathbf{P}[|X_0|>u]}\\ \to\; 2\cdot\mathfrak{a}_{\sigma(0)}^\nu\cdot\mu_2&[(1,\infty)\times(\mathfrak{a}_1,\infty)]\,,\end{aligned}$$

where the tail measure is given by

$$\mu_2[(1,\infty)\times(\mathfrak{a}_1,\infty)] \;=\; \frac{1}{\sum_{i=0}^\infty a_i^\nu}\sum_{j=-1}^{\infty}\int_{\mathbb{R}}\frac{\nu}{2}|z|^{-1-\nu}\mathbb{1}\{z\cdot a_j > 1, z\cdot a_{j+1} > \mathfrak{a}_1\}\,\mathrm{d}z \;=\; \frac{\mathfrak{a}_1^{-\nu}}{2}\frac{\sum_{i=1}^\infty a_i^\nu}{\sum_{i=0}^\infty a_i^\nu}\,,$$

so that

$$\lim_{u\to\infty}\mathbf{P}[X_1 > \mathfrak{a}_1\cdot u \mid X_0 > u] \;=\; \mathfrak{a}_1^{-\nu}\frac{\sum_{i=1}^\infty a_i^\nu}{\sum_{i=0}^\infty a_i^\nu} \quad \text{and} \quad \lim_{u\to\infty}\mathbf{P}[X_0 > u \mid X_1 > \mathfrak{a}_1\cdot u] \;=\; \frac{\sum_{i=1}^\infty a_i^\nu}{\sum_{i=0}^\infty a_i^\nu} \;=\; p_{2,A}\,.$$

5.2. Cancellation in the asymptotic scale

In Remark 4.9 we point out that the asymptotic scale in Theorem 4.8 may vanish. We now study this for rectangular sets. In the setting of Proposition 5.2, we have

$$H_A(1)^\top \mathbf{1}_T \;+\; p_A(1) \;=\; \sum_{j\in J_A} \mathfrak{a}_j^{-\nu}\left(\frac{1}{\mathfrak{a}_j} \;-\; 1\right) p_{j,A}\,.$$

This is the asymptotic scale in Theorem 4.8 multiplied by $2/\nu$. In the setting of Theorem 5.3, this equals

$$\mathfrak{a}_1^{-\nu}\left(\frac{1}{\mathfrak{a}_1} \;-\; 1\right)\left(1-\frac{1}{\sum_{i=0}^{\infty} a_i^{\nu}}\right).$$

Note that, if $\mathfrak{a}_j = 1$ for all $j \in J_A$, this vanishes and therefore the limit in Theorem 4.8 is degenerate. For $T = 1$, this is expected, since

$$\widehat{\mathbf{P}}_n(X_{n-k:n}/u_n, A) \;=\; \frac{1}{n}\frac{1}{\mathbf{P}[|X_0| > u_n]}\sum_{t=1}^{n} \mathbb{1}\{X_t > X_{n-k:n}\} \;=\; \frac{k}{n}\frac{1}{\mathbf{P}[|X_0| > u_n]}$$

is deterministic. For $T > 1$, this is remarkable and genuinely new, because

$$\widehat{\mathbf{P}}_n(X_{n-k:n}/u_n, A) \;=\; \frac{1}{n-T+1}\frac{1}{\mathbf{P}[|X_0| > u_n]}\sum_{t=1}^{n-T+1} \mathbb{1}\left\{\min_{j\in J_A} X_{t+j-1} > X_{n-k:n}\right\}$$

is generally random. Under the speed of Theorem 4.2, this randomness, however, is not sufficient to yield a non-degenerate limit in the setting of Theorem 4.8. It remains an open question what the correct rate of convergence is in this setting.

5.3. Univariate theory for random thresholds

Since our theory is also valid in the univariate case $T = 1$, we now compare our results with those in the i.i.d. and weakly dependent setting which are readily derived from existing theory. For $\mathfrak{a}_1 = \mathfrak{a} \neq 1$, it follows from Theorem 4.8 that

$$n^{1-d-1/\alpha} q_{X_0}\left(1-\frac{k}{n}\right)\left(\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\mathbb{1}\{X_t > \mathfrak{a}\cdot X_{n-k:n}\}}{\mathbf{P}[X_0 > u_n]} \;-\; \frac{\mathbf{P}[X_0 > \mathfrak{a}\cdot u_n]}{\mathbf{P}[X_0 > u_n]}\right)\right) \;\to_d\; \nu\mathfrak{a}^{-\nu}\left|\frac{1}{\mathfrak{a}}-1\right| Z_\alpha,$$

as $n \to \infty$. The next theorem provides a comparable result in the i.i.d. setting.

Theorem 5.5 (Univariate i.i.d. setting). *Let $X_1, \ldots, X_n$ be i.i.d. copies of a random variable X having a continuous distribution. Assume that there are $\nu > 0$, $\rho < 0$, and a function $\mathfrak{A}$ that is either positive or negative, such that, for all $x > 0$,*

$$\frac{\dfrac{\mathbf{P}[X > x\cdot u]}{\mathbf{P}[X > u]} \;-\; x^{-\nu}}{\mathfrak{A}(1/\mathbf{P}[X > u])} \;\to\; x^{-\nu}\frac{x^{\rho\nu}-1}{\rho/\nu} \qquad \text{as } u \to \infty\,.$$

Let $u_n \to \infty$ satisfy $n\mathbf{P}[X > u_n] \to \infty$ and $\mathfrak{A}(1/\mathbf{P}[X > u_n]) = O(1/\sqrt{n\mathbf{P}[X > u_n]})$, and write $k = \lfloor n\cdot \mathbf{P}[X > u_n]\rfloor$. Then, for any $\mathfrak{a} > 0$,

$$\sqrt{k}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\mathbb{1}\{X_i > \mathfrak{a}\cdot X_{n-k:n}\}}{\mathbf{P}[X > u_n]} \;-\; \frac{\mathbf{P}[X > \mathfrak{a}\cdot u_n]}{\mathbf{P}[X > u_n]}\right)\right) \;\to_d\; \left(\mathfrak{a}^{-\nu}\left|\frac{1}{\mathfrak{a}^{\nu}}-1\right|\right)^{1/2} Z\,,$$

as $n \to \infty$, where Z is a standard normal random variable.

Remark 5.6 (Weakly dependent time series)**.** The asymptotic variance in Theorem 5.5 comes from

$$\mathrm{Var}[W(\mathfrak{a}^{-\nu})] \;+\; \mathfrak{a}^{-2\nu}\mathrm{Var}[W(1)] \;-\; 2\mathfrak{a}^{-\nu}\mathrm{Cov}[W(\mathfrak{a}^{-\nu}),W(1)] \;=\; \mathfrak{a}^{-\nu}\left|\frac{1}{\mathfrak{a}^{\nu}} \;-\; 1\right| ,$$

where W is a Brownian motion. We compare this with the short memory setting. Let (X_t) be a weakly dependent linear time series as in Section 3.2 in Drees (2003) with non-negative, decreasing coefficients (a_i). Under stronger assumptions, it follows from Theorem 2.1 in Drees (2003), by an argument similar to that used in the proof of Theorem 5.5, that

$$\begin{aligned}\sqrt{k}&\left(\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\mathbb{1}\{X_t > \mathfrak{a}\cdot X_{n-k:n}\}}{\mathbf{P}[X_0 > u_n]} \;-\; \frac{\mathbf{P}[X_0 > \mathfrak{a}\cdot u_n]}{\mathbf{P}[X_0 > u_n]}\right)\right)\\ &\to_d \;\left(c(\mathfrak{a}^{-\nu},\mathfrak{a}^{-\nu}) \;+\; \mathfrak{a}^{-2\nu}\cdot c(1,1) \;-\; 2\cdot\mathfrak{a}^{-\nu}\cdot c(\mathfrak{a}^{-\nu},1)\right)^{1/2} Z\end{aligned}$$

as $n\to\infty$, where Z is a standard normal random variable, and where

$$c(x,y) \;=\; (x\wedge y) \;+\; \sum_{m=1}^{\infty}\left(c_m(x,y) \;+\; c_m(y,x)\right)$$

with

$$c_m(x,y) \;=\; \frac{1}{\sum_{i=0}^{\infty} a_i^{\nu}}\sum_{j=0}^{\infty}\left((x\cdot a_j^{\nu})\wedge(y\cdot a_{j+m}^{\nu})\right) , \qquad m\in\mathbb{N}.$$

Thus, the Brownian motion W is replaced by a Gaussian process with covariance function c. The term $x\wedge y$ yields the variance in the i.i.d. setting. The additional contribution resulting from weak dependence is

$$\begin{aligned}&\sum_{m=1}^{\infty} 2\cdot\left(c_m(\mathfrak{a}^{-\nu},\mathfrak{a}^{-\nu}) \;+\; \mathfrak{a}^{-2\nu}c_m(1,1) \;-\; \mathfrak{a}^{-\nu}\left(c_m(\mathfrak{a}^{-\nu},1) \;+\; c_m(1,\mathfrak{a}^{-\nu})\right)\right)\\ &= \frac{2}{\sum_{i=0}^{\infty} a_i^{\nu}}\sum_{m=1}^{\infty}\sum_{j=0}^{\infty}\left(\mathfrak{a}^{-\nu}a_{j+m}^{\nu} \;+\; \mathfrak{a}^{-2\nu}a_{j+m}^{\nu} \;-\; \mathfrak{a}^{-\nu}\left((\mathfrak{a}^{-\nu}a_j^{\nu})\wedge a_{j+m}^{\nu} \;+\; a_j^{\nu}\wedge(\mathfrak{a}^{-\nu}a_{j+m}^{\nu})\right)\right)\\ &= \mathfrak{a}^{-\nu}\frac{2}{\sum_{i=0}^{\infty} a_i^{\nu}}\sum_{m=1}^{\infty}\sum_{j=0}^{\infty}\left(a_{j+m}^{\nu} \;+\; \mathfrak{a}^{-\nu}a_{j+m}^{\nu} \;-\; \left((\mathfrak{a}^{-\nu}a_j^{\nu})\wedge a_{j+m}^{\nu} \;+\; a_j^{\nu}\wedge(\mathfrak{a}^{-\nu}a_{j+m}^{\nu})\right)\right) .\end{aligned}$$

Remark 5.7 (Comparison of univariate results)**.** While both the i.i.d. and weakly dependent settings yield the speed $\sqrt{k}$, our long memory setting gives the rate $n^{1-d-1/\alpha}q_{X_0}(1-k/n)$. A similar difference in speed has also been observed in the univariate setting for the Hill estimator (Scheffel et al., 2026, Section 1.2). While we see a Gaussian limit with modified variance under weak dependence, the long memory setting with $\nu\in(1,2)$ changes the class of limiting distributions to $S\alpha S$ where $\alpha=(\nu\wedge 2)\in(1,2)$. We also note that the dependence of the limit on $\mathfrak{a}$ generally differs among the three settings.

Acknowledgments

Financial support from the French CNRS within the project "Extreme value analysis of time series through de-randomization techniques", funded through the IEA program, is gratefully acknowledged. G. Stupfler acknowledges further financial support from the French *Agence Nationale de la Recherche* under the grants ANR-23-CE40-0009 (EXSTA project) and ANR-11-LABX-0020-01 (Centre Henri Lebesgue), as well as from the Chair Stress Test, RISK Management and Financial Steering of the Foundation Ecole Polytechnique.

References

H. W. Alt. *Linear Functional Analysis.* Springer, 2016.

B. Basrak and J. Segers. Regularly varying multivariate time series. *Stochastic Processes and their Applications*, 119(4):1055–1080, 2009.

J. Beirlant, Y. Goegebeur, J. Segers, and J. L. Teugels. *Statistics of Extremes: Theory and Applications.* John Wiley & Sons, 2006.

A. Betken, G. Micali, and J. Schmidt-Hieber. Central limit theorems for the outputs of fully convolutional neural networks with time series input. arXiv preprint arXiv:2603.29612.

N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation.* Cambridge University Press, Cambridge, 1987.

P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods (second edition).* Springer, 1991.

R. Cont. Long range dependence in financial markets. In J. Lévy-Véhel and E. Lutton, editors, *Fractals in Engineering*, pages 159–179. Springer London, 2005.

R. A. Davis and T. Mikosch. The extremogram: A correlogram for extreme events. *Bernoulli*, 15(4): 977–1009, 2009.

L. De Haan and A. Ferreira. *Extreme Value Theory.* Springer, New York, NY, 2006.

S. W. Dharmadhikari and K. Joag-Dev. *Unimodality, Convexity, and Applications.* Academic Press, 1988.

H. Drees. Bootstrapping empirical processes of cluster functionals with application to extremograms. arXiv preprint arXiv:1511.00420.

H. Drees. Extreme quantile estimation for dependent data, with applications to finance. *Bernoulli*, 9(4): 617–657, 2003.

H. Drees and M. Knežević. Peak-over-threshold estimators for spectral tail processes: random vs deterministic thresholds. *Extremes*, 23(3):465–491, 2020.

J. R. M. Hosking. Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series. *Journal of Econometrics*, 73(1):261–284, 1996.

H. L. Koul and D. Surgailis. Asymptotics of empirical processes of long memory moving averages with infinite variance. *Stochastic Processes and their Applications*, 91(2):309–336, 2001.

R. Kulik and P. Soulier. The tail empirical process for long memory stochastic volatility sequences. *Stochastic Processes and their Applications*, 121(1):109–134, 2011.

R. Kulik and P. Soulier. *Heavy-Tailed Time Series.* Springer, New York, NY, 2020.

B. Merz, S. Basso, S. Fischer, D. Lun, G. Blöschl, R. Merz, B. Guse, A. Viglione, S. Vorogushyn, E. Macdonald, L. Wietzke, and A. Schumann. Understanding Heavy Tails of Flood Peak Distributions. *Water Resources Research*, 58(6):e2021WR030506, 2022.

T. Mikosch and O. Wintenberger. *Extreme Value Theory for Time Series: Models with Power-Law Tails*. Springer Nature Switzerland, Cham, 2024.

A. Montanari, R. Rosso, and M. S. Taqqu. Fractionally differenced ARIMA models applied to hydrologic time series: Identification, estimation, and simulation. *Water Resources Research*, 33(5):1035–1044, 1997.

M. Mudelsee. Long memory of rivers from spatial aggregation. *Water Resources Research*, 43(1), 2007.

J. W. Pratt. On Interchanging Limits and Integrals. *The Annals of Mathematical Statistics*, 31(1):74–77, 1960.

E. Rossi and P. Santucci de Magistris. Long memory and tail dependence in trading volume and volatility. *Journal of Empirical Finance*, 22:94–112, 2013.

I. Scheffel, M. Oesting, and G. Stupfler. Central limit theory for peaks-over-threshold partial sums of long memory linear time series. *Stochastic Processes and their Applications*, 197:104929, 2026.

A. W. Van Der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer International Publishing, 2023.

Supplementary Material to the article

Central limit theory for serial tail dependence estimators in heavy-tailed long memory linear time series

Ioan Scheffel [a] , Marco Oesting [a,b] , Gilles Stupfler [c]

[a] Institute for Stochastics and Applications, University of Stuttgart, D-70563 Stuttgart, Germany

[b] Stuttgart Center for Simulation Science (SC SimTech), University of Stuttgart, D-70569 Stuttgart, Germany

[c] Univ Angers, CNRS, LAREMA, SFR MATHSTIC, F-49000 Angers, France

Fix $T \geq 1$ and $0 < \underline{s} < 1 < \overline{s} < \infty$. The first lemma reduces the multivariate problem to a series of univariate problems, which motivates us to distinguish it from the rest of the supplementary material.

Lemma. *Let either of the following conditions hold:*

(i) A *is bounded away from 0,* $s_1 \in (\underline{s}, \infty)$ *and* $s_2 = \infty$

(ii) $A \in \mathcal{C}_T$ *and* $s_1, s_2 \in (\underline{s}, \overline{s}]$ *satisfy* $s_1 < s_2$

Then there exists $\{\mathfrak{a}_j\}_{j \in \{1,\dots,T\}} \subset (0,\infty)$ *independent of* s_1 *and* s_2*, such that*

$$\mathbb{1}\{\mathbf{z} \in u_n \cdot (s_1 \cdot A) \setminus (s_2 \cdot A)\} \;\leq\; \sum_{j=1}^{T} \mathbb{1}\{|z_{j-1}| \in u_n \cdot \mathfrak{a}_j \cdot [s_1, s_2]\}\,, \qquad \mathbf{z} \in \mathbb{R}^T\,. \tag{10}$$

Proof. Under condition (ii), the statement follows immediately from the definition of $\mathcal{C}_T$. It remains to show the statement under condition (i).

Proof under condition (i): A set A is bounded away from 0 if and only if there exists $\{\mathfrak{a}_j\}_{j\in\{1,\dots,T\}} \subset (0,\infty)$ such that

$$A \;\subset\; \bigcup_{j=1}^{T} \{|x_{j-1}| \geq \mathfrak{a}_j\}\,.$$

For $s_1 \in (\underline{s}, \infty)$ and $s_2 = \infty$, it holds

$$(s_1 \cdot A) \setminus (s_2 \cdot A) \;=\; s_1 \cdot A \;\subset\; s_1 \cdot \bigcup_{j=1}^{T}\{|x_{j-1}| \geq \mathfrak{a}_j\} \;=\; \bigcup_{j=1}^{T}\{|x_{j-1}| \in \mathfrak{a}_j \cdot [s_1, \infty)\}\,.$$

Taking indicators of these sets proves the claimed inequality. □

Assumption. Throughout the supplementary material, we assume that the conditions of the lemma leading to (10) are satisfied with $\{\mathfrak{a}_j\}$, s_1 and s_2 fixed.

A. Extensions of results from Scheffel et al. (2026)

Lemma A.1 (Lipschitz bound for f_ε)**.** *Let Assumption 1 hold. Then, for all* $x, y \in \mathbb{R}$ *satisfying* $|x-y| < 1$,

$$|f_\varepsilon(x) - f_\varepsilon(y)| \;\lesssim\; |x-y| \cdot g_{\alpha-1}(x)\,,$$

where $\lesssim$ *means inequality up to a multiplicative constant that depends only on the distribution of* ε *.*

Proof. Applying the mean value theorem together with the triangle inequality and Assumption 1 it follows that

$$\begin{aligned} |f_\varepsilon(x) - f_\varepsilon(y)| &\le \int_{0\wedge(y-x)}^{0\vee(y-x)} |f'_\varepsilon(x+s)| \, \mathrm{d}s \lesssim \int_{0\wedge(y-x)}^{0\vee(y-x)} g_{\alpha-1}(x+s)\, \mathrm{d}s \\ &\lesssim g_{\alpha-1}(x) \cdot \int_{0\wedge(y-x)}^{0\vee(y-x)} (1 \vee |s|)^\alpha \, \mathrm{d}s \,, \end{aligned}$$

where the last step follows from Lemma A.2. To conclude the proof, note that $|y-x| < 1$ by assumption, so that

$$\int_{0\wedge(y-x)}^{0\vee(y-x)} (1 \vee |s|)^\alpha \, \mathrm{d}s \;=\; |y-x| \,.$$

This proves the asserted bound for $|f_\varepsilon(x) - f_\varepsilon(y)|$. □

For $\gamma > 0$ we define

$$g_\gamma(x) \;:=\; \frac{1}{(1+|x|)^{1+\gamma}} \qquad \text{for } x \in \mathbb{R}\,.$$

Lemma A.2 (Bounds for g_γ)**.** *Let $\gamma \in (0,\infty)$. Then the following holds for g_γ:*

(i) For all $y, z \in \mathbb{R}$, and all $\gamma \in (0,1)$,

$$g_\gamma(z+y) \;\lesssim\; g_\gamma(z) \cdot (1 \vee |y|)^{1+\gamma}\,,$$

where $\lesssim$ means inequality up to a multiplicative constant that depends only on γ.

(ii) Let $a, b \in \mathbb{R}$ with $a < b$. Then, for all $z \in \mathbb{R}$, and all $\gamma \in (0,1)$,

$$\int_a^b g_\gamma(z-w)\, \mathrm{d}w \;\lesssim\; g_\gamma(z) \cdot ((b-a) \vee 1)^{1+\gamma}\,,$$

where $\lesssim$ means inequality up to a multiplicative constant that depends only on γ.

(iii) For $c \neq 0$,

$$g_\gamma(z \cdot c) \;\le\; \left(1 \vee \frac{1}{|c|}\right)^{1+\gamma} g_\gamma(z) \;\lesssim\; g_\gamma(z)\,,$$

where $\lesssim$ means inequality up to a multiplicative constant that depends only on γ and c.

(iv) Let $\mathfrak{x}, \mathfrak{y}, \mathfrak{z} \in \mathbb{R}$ and $\gamma \in (0,1)$. Then

$$\int_{\mathfrak{b}\wedge\mathfrak{c}}^{\mathfrak{b}\vee\mathfrak{c}} (1 \wedge |s \cdot \mathfrak{x}|) \cdot g_\gamma\left(\mathfrak{z} + s \cdot \mathfrak{y}\right) \mathrm{d}s \;\lesssim\; \left(|\mathfrak{y}|^\gamma + |\mathfrak{x}|^\gamma\right) \cdot g_\gamma(\mathfrak{z}) \cdot \left(|\mathfrak{b}|^{1+\gamma} + |\mathfrak{c}|^{1+\gamma}\right)\,,$$

where $\lesssim$ means inequality up to a multiplicative constant that depends only on γ.

Proof. For a proof of (i) and (ii) see the proof of Lemma A.3 in Scheffel et al. (2026).

Proof of (iii): We distinguish $|c| \in (0,1)$ and $|c| \ge 1$. For $|c| \in (0,1)$ it holds $1/|c| > 1$, and therefore

$$\begin{aligned} g_\gamma(z \cdot c) &= \frac{1}{(1+|z\cdot c|)^{1+\gamma}} = \left(\frac{1}{|c|}\right)^{1+\gamma} \frac{1}{(1/|c| + |z|)^{1+\gamma}} \\ &\le \left(1 \;\vee\; \frac{1}{|c|}\right)^{1+\gamma} \frac{1}{(1+|z|)^{1+\gamma}} \;=\; \left(1 \;\vee\; \frac{1}{|c|}\right)^{1+\gamma} g_\gamma(z)\,. \end{aligned}$$

If $|c| \geq 1$, then $1/|c| \leq 1$, and $|z \cdot c| \geq |z|$, so that

$$g_\gamma(z \cdot c) \;=\; \frac{1}{(1+|z\cdot c|)^{1+\gamma}} \;\leq\; \frac{1}{(1+|z|)^{1+\gamma}} \;=\; \left(1 \,\vee\, \frac{1}{|c|}\right)^{1+\gamma} g_\gamma(z)\,.$$

Combining the two cases proves the asserted bound for $g_\gamma(z \cdot c)$.

Proof of (iv): If $\mathfrak{x} = 0$ there is nothing to prove. Therefore we assume $\mathfrak{x} \neq 0$.

Case $\mathfrak{y} = 0$*:* Since $\gamma \in (0,1)$,

$$(1 \wedge |s \cdot \mathfrak{x}|) \;\leq\; (1 \wedge |s \cdot \mathfrak{x}|)^\gamma \;\leq\; |s \cdot \mathfrak{x}|^\gamma\,, \tag{11}$$

so that

$$\begin{aligned}
&\int_{\mathfrak{b}\wedge\mathfrak{c}}^{\mathfrak{b}\vee\mathfrak{c}} (1 \wedge |s \cdot \mathfrak{x}|) \cdot g_\gamma\left(\mathfrak{z} + s \cdot \mathfrak{y}\right) \mathrm{d}s \;=\; g_\gamma(\mathfrak{z}) \int_{\mathfrak{b}\wedge\mathfrak{c}}^{\mathfrak{b}\vee\mathfrak{c}} (1 \wedge |s \cdot \mathfrak{x}|)\, \mathrm{d}s \;\leq\; g_\gamma(\mathfrak{z}) \int_{\mathfrak{b}\wedge\mathfrak{c}}^{\mathfrak{b}\vee\mathfrak{c}} |s \cdot \mathfrak{x}|^\gamma\, \mathrm{d}s \\
&\lesssim\; g_\gamma(\mathfrak{z}) \cdot |\mathfrak{x}|^\gamma \cdot \left(|\mathfrak{b}|^{1+\gamma} + |\mathfrak{c}|^{1+\gamma}\right) \;=\; \left(|\mathfrak{y}|^\gamma + |\mathfrak{x}|^\gamma\right) \cdot g_\gamma(\mathfrak{z}) \cdot \left(|\mathfrak{b}|^{1+\gamma} + |\mathfrak{c}|^{1+\gamma}\right)\,,
\end{aligned}$$

where the multiplicative constant in $\lesssim$ depends only on γ.

Case $\mathfrak{y} \neq 0$*:* We make the split

$$\begin{aligned}
&\int_{\mathfrak{b}\wedge\mathfrak{c}}^{\mathfrak{b}\vee\mathfrak{c}} (1 \wedge |s \cdot \mathfrak{x}|) \cdot g_\gamma\left(\mathfrak{z} + s \cdot \mathfrak{y}\right) \mathrm{d}s \\
&\leq\; \int_{\mathfrak{b}\wedge\mathfrak{c}}^{\mathfrak{b}\vee\mathfrak{c}} (1 \wedge |s \cdot \mathfrak{x}|) \cdot \mathbb{1}\{|s \cdot \mathfrak{y}| < 1\} \cdot g_\gamma\left(\mathfrak{z} + s \cdot \mathfrak{y}\right) \mathrm{d}s + \int_{\mathfrak{b}\wedge\mathfrak{c}}^{\mathfrak{b}\vee\mathfrak{c}} \mathbb{1}\{|s \cdot \mathfrak{y}| \geq 1\} \cdot g_\gamma\left(\mathfrak{z} + s \cdot \mathfrak{y}\right) \mathrm{d}s\,.
\end{aligned}$$

Analysis of $\int_{\mathfrak{b}\wedge\mathfrak{c}}^{\mathfrak{b}\vee\mathfrak{c}} (1 \wedge |s \cdot \mathfrak{x}|)\, \mathbb{1}\{|s \cdot \mathfrak{y}| < 1\} \cdot g_\gamma\left(\mathfrak{z} + s \cdot \mathfrak{y}\right) \mathrm{d}s$*:* By (i), $\gamma \in (0,1)$, and (11),

$$\begin{aligned}
&\mathbb{1}\{|s \cdot \mathfrak{y}| < 1\} \cdot (1 \wedge |s \cdot \mathfrak{x}|) \cdot g_\gamma\left(\mathfrak{z} + s \cdot \mathfrak{y}\right) \\
&\lesssim\; \mathbb{1}\{|s \cdot \mathfrak{y}| < 1\} \cdot (1 \wedge |s \cdot \mathfrak{x}|) \cdot (1 \vee |s \cdot \mathfrak{y}|)^{1+\gamma} \cdot g_\gamma(\mathfrak{z}) \\
&=\; \mathbb{1}\{|s \cdot \mathfrak{y}| < 1\} \cdot (1 \wedge |s \cdot \mathfrak{x}|) \cdot g_\gamma(\mathfrak{z}) \\
&\leq\; (1 \wedge |s \cdot \mathfrak{x}|) \cdot g_\gamma(\mathfrak{z}) \\
&\leq\; |s \cdot \mathfrak{x}|^\gamma \cdot g_\gamma(\mathfrak{z})\,,
\end{aligned}$$

so that

$$\begin{aligned}
&\int_{\mathfrak{b}\wedge\mathfrak{c}}^{\mathfrak{b}\vee\mathfrak{c}} \mathbb{1}\{|s \cdot \mathfrak{y}| < 1\} \cdot (1 \wedge |s \cdot \mathfrak{x}|) \cdot g_\gamma\left(\mathfrak{z} + s \cdot \mathfrak{y}\right) \mathrm{d}s \\
&\lesssim\; |\mathfrak{x}|^\gamma \cdot g_\gamma(\mathfrak{z}) \cdot \int_{\mathfrak{b}\wedge\mathfrak{c}}^{\mathfrak{b}\vee\mathfrak{c}} |s|^\gamma\, \mathrm{d}s \\
&\lesssim\; |\mathfrak{x}|^\gamma \cdot g_\gamma(\mathfrak{z}) \cdot \left(|\mathfrak{b}|^{1+\gamma} + |\mathfrak{c}|^{1+\gamma}\right) \\
&\leq\; \left(|\mathfrak{y}|^\gamma + |\mathfrak{x}|^\gamma\right) \cdot g_\gamma(\mathfrak{z}) \cdot \left(|\mathfrak{b}|^{1+\gamma} + |\mathfrak{c}|^{1+\gamma}\right)\,,
\end{aligned}$$

where the constant in $\lesssim$ depends only on γ.

Analysis of $\int_{\mathfrak{b}\wedge\mathfrak{c}}^{\mathfrak{b}\vee\mathfrak{c}} \mathbb{1}\{|s \cdot \mathfrak{y}| \geq 1\} \cdot g_\gamma\left(\mathfrak{z} + s \cdot \mathfrak{y}\right) \mathrm{d}s$*:* Note that $s \in (\mathfrak{b} \wedge \mathfrak{c}, \mathfrak{b} \vee \mathfrak{c})$ implies $|s| < |\mathfrak{b}| \vee |\mathfrak{c}|$. Moreover, we have

$$1 \;\geq\; |\mathfrak{y}| \cdot (|\mathfrak{b}| \vee |\mathfrak{c}|)\,, \qquad |\mathfrak{y}||\mathfrak{b}| \;\geq\; (1 \vee |\mathfrak{y}||\mathfrak{c}|) \qquad \text{or} \qquad |\mathfrak{y}||\mathfrak{c}| \;\geq\; (1 \vee |\mathfrak{y}||\mathfrak{b}|)\,.$$

We consider the three cases separately.
Case $1 \geq |\mathfrak{y}| \cdot (|\mathfrak{b}| \vee |\mathfrak{c}|)$ *:* Since $|s \cdot \mathfrak{y}| < |\mathfrak{y}| \cdot (|\mathfrak{b}| \vee |\mathfrak{c}|) \leq 1$ for $|s| < (|\mathfrak{b}| \vee |\mathfrak{c}|)$,

$$\mathbb{1}\{1 \geq |\mathfrak{y}|(|\mathfrak{b}| \vee |\mathfrak{c}|)\} \int_{\mathfrak{b}\wedge\mathfrak{c}}^{\mathfrak{b}\vee\mathfrak{c}} \mathbb{1}\{|s \cdot \mathfrak{y}| \geq 1\} \cdot g_\gamma(\mathfrak{z} + s \cdot \mathfrak{y}) \, \mathrm{d}s \;=\; 0$$

Case $|\mathfrak{y}||\mathfrak{b}| \geq (1 \vee |\mathfrak{y}||\mathfrak{c}|)$ *:* Since $\mathfrak{y} \neq 0$, by the change of variables $s \mapsto s/\mathfrak{y}$,

$$\begin{aligned}
&\mathbb{1}\{|\mathfrak{y}||\mathfrak{b}| \geq (1 \vee |\mathfrak{y}||\mathfrak{c}|)\} \int_{\mathfrak{b}\wedge\mathfrak{c}}^{\mathfrak{b}\vee\mathfrak{c}} \mathbb{1}\{|s \cdot \mathfrak{y}| \geq 1\} \cdot g_\gamma(\mathfrak{z} + s \cdot \mathfrak{y}) \, \mathrm{d}s \\
&\leq\; \mathbb{1}\{|\mathfrak{y}||\mathfrak{b}| \geq 1\} \cdot \frac{1}{|\mathfrak{y}|} \cdot \int_{|s| \leq |\mathfrak{y}||\mathfrak{b}|} g_\gamma(\mathfrak{z} + s) \, \mathrm{d}s \\
&\lesssim\; \mathbb{1}\{|\mathfrak{y}||\mathfrak{b}| \geq 1\} \cdot g_\gamma(\mathfrak{z}) \cdot \frac{1}{|\mathfrak{y}|} \cdot (1 \vee |\mathfrak{b} \cdot \mathfrak{y}|)^{1+\gamma} \\
&\leq\; g_\gamma(\mathfrak{z}) \cdot |\mathfrak{b}|^{1+\gamma} \cdot |\mathfrak{y}|^\gamma \\
&\leq\; (|\mathfrak{y}|^\gamma + |\mathfrak{x}|^\gamma) \cdot g_\gamma(\mathfrak{z}) \cdot \left(|\mathfrak{b}|^{1+\gamma} + |\mathfrak{c}|^{1+\gamma}\right) ,
\end{aligned}$$

where the third to last inequality follows from (ii), and therefore the multiplicative constant in $\lesssim$ depends only on γ.
Case $|\mathfrak{y}||\mathfrak{c}| \geq (1 \vee |\mathfrak{y}||\mathfrak{b}|)$ *:* The same argument yields

$$\begin{aligned}
&\mathbb{1}\{|\mathfrak{c}||\mathfrak{y}| \geq (1 \vee |\mathfrak{b}||\mathfrak{y}|)\} \int_{\mathfrak{b}\wedge\mathfrak{c}}^{\mathfrak{b}\vee\mathfrak{c}} \mathbb{1}\{|s \cdot \mathfrak{y}| \geq 1\} \cdot g_\gamma(\mathfrak{z} + s \cdot \mathfrak{y}) \, \mathrm{d}s \\
&\lesssim\; g_\gamma(\mathfrak{z}) \cdot |\mathfrak{c}|^{1+\gamma} \cdot |\mathfrak{y}|^\gamma \\
&\leq\; (|\mathfrak{y}|^\gamma + |\mathfrak{x}|^\gamma) \cdot g_\gamma(\mathfrak{z}) \cdot \left(|\mathfrak{b}|^{1+\gamma} + |\mathfrak{c}|^{1+\gamma}\right) ,
\end{aligned}$$

where the constant in $\lesssim$ depends only on γ. □

Lemma A.3 (Martingale difference inequality). *Let $p \in [1,2]$, and let $(Y_{m,i}\,, i = 1, \ldots, m\,, m \in \mathbb{N}) \subset L^p(\mathbf{P})$ be an array of random variables satisfying*

$$\mathbf{E}\left[Y_{m,\ell+1} \,\middle|\, \sum_{i=1}^{\ell} Y_{m,i} \right] \;=\; 0 \qquad \text{for all } 1 \leq \ell \leq m-1 \text{ and all } m \in \mathbb{N}\,. \tag{12}$$

(i) It holds that

$$\mathbf{E}\left|\sum_{i=1}^{m} Y_{m,i}\right|^p \;\leq\; 2\sum_{i=1}^{m} \mathbf{E}\,|Y_{m,i}|^p \qquad \text{for all } m \in \mathbb{N}\,. \tag{13}$$

(ii) If there exists a sequence $(Y_i)_{i\in\mathbb{N}} \subset L^p(\mathbf{P})$ such that

$$\lim_{m\to\infty} \sum_{i=1}^{m} Y_{m,i} = \sum_{i=1}^{\infty} Y_i \quad \mathbf{P}\text{-almost surely, and} \qquad \liminf_{m\to\infty} \sum_{i=1}^{m} \mathbf{E}|Y_{m,i}|^p \leq \sum_{i=1}^{\infty} \mathbf{E}|Y_i|^p\,, \tag{14}$$

then

$$\mathbf{E}\left|\sum_{i=1}^{\infty} Y_i\right|^p \;\leq\; 2\sum_{i=1}^{\infty} \mathbf{E}\,|Y_i|^p\,. \tag{15}$$

Proof. See the proof of Lemma A.1 in Scheffel et al. (2026). □

Lemma A.4 (Moments of the marginal distribution)**.** *Let Assumptions 1 and 2 hold. Then* $X_0 \in L^r(\mathbf{P})$ *for all* $r \in [1, \alpha)$ *.*

Proof. Since $L^q(\mathbf{P}) \subset L^p(\mathbf{P})$ for $1 \le p < q$, it suffices to show the statement for $r \in (1/(1-d), \alpha)$. To this end, we apply Lemma A.3 to the array $(Y_{m,i}) = (a_i\varepsilon_{-i} : i = 0, \ldots, m, m \in \mathbb{N})$ and the sequence $(Y_i) = (a_i\varepsilon_{-i})_{i\ge 0}$. Since $r < \alpha$ by assumption, $(Y_{m,i}), (Y_i) \subset L^r(\mathbf{P})$. Since the innovations are independent and centered, Condition (12) is satisfied. The other conditions of Lemma A.3 are clearly met, so that

$$\mathbf{E}[|X_0|^r] \;=\; \mathbf{E}\left[\left|\sum_{i=0}^{\infty} Y_i\right|^r\right] \;\lesssim\; \sum_{i=0}^{\infty}\mathbf{E}[|Y_i|^r] \;=\; \mathbf{E}[|\varepsilon|^r]\sum_{i=0}^{\infty}|a_i|^r \;\lesssim\; \sum_{i=0}^{\infty} i^{-(1-d)r}\,.$$

The latter expression is finite by the assumption $1/(1-d) < r$, so that $(1-d)r > 1$. This shows that $X_0 \in L^r(\mathbf{P})$. □

Lemma A.5 (Decay rate of the moment of the remainder)**.** *Let Assumptions 1 and 2 hold. Then, for all* $k \ge 1$ *and* $r \in (1/(1-d), \alpha)$ *,*

$$\mathbf{E}\left[\left|\sum_{i=k+1}^{\infty} a_i\varepsilon_{-i}\right|^r\right] \;\lesssim\; k^{1-(1-d)r}\,,$$

where $\lesssim$ *means inequality up to a multiplicative constant that depends on the coefficients* (a_i) *, the distribution of* ε *, and* r *, and is independent of* k *.*

Proof. We define the array and limiting sequence

$$(Y_{m,i}) \;=\; (a_{k+i}\varepsilon_{-(k+i)}\,, i = 1, \ldots, m\,, m \in \mathbb{N}) \qquad \text{and} \qquad (Y_i) \;=\; (a_{k+i}\varepsilon_{-(k+i)})_{i\in\mathbb{N}}\,.$$

Note that (12) and (14) are satisfied for $(Y_{m,i})$ and (Y_i). Then we may write

$$\sum_{i=k+1}^{\infty} a_i\varepsilon_{-i} \;=\; \sum_{i=1}^{\infty} a_{k+i}\varepsilon_{-(k+i)} \;=\; \sum_{i=1}^{\infty} Y_i\,.$$

The ε_j are independent, centered and belong to $L^r(\mathbf{P})$, so applying Inequality (15) in Lemma A.3(ii), we obtain

$$\mathbf{E}\left[\left|\sum_{i=k+1}^{\infty} a_i\varepsilon_{-i}\right|^r\right] \;\lesssim\; \sum_{i=k+1}^{\infty}|a_i|^r\mathbf{E}|\varepsilon|^r \;\lesssim\; \sum_{i=k+1}^{\infty} i^{-(1-d)r} \;\lesssim\; k^{1-(1-d)r} \;\le\; 1\,,$$

where we used $a_i i^{1-d} \sim c_a$ and $(1-d)r > 1$. The multiplicative constant in $\lesssim$ depends on the coefficients (a_i), the distribution of ε, and r, and is independent of k. This proves the asserted upper bound. □

B. Multivariate techniques

We define, for $k \ge 0$, the k-truncated time series $(X_{t,k})$ by $X_{t,k} := \sum_{i=0}^{k} a_i\varepsilon_{t-i}$ and make the convention $X_{t,\infty} = X_t$ so that $X_{t,k} = X_{t,\infty} - \sum_{i=k+1}^{\infty} a_i\varepsilon_{t-i}$. For $i \in \{1, \ldots, T\}$, let $\mathbf{e}_i \in \mathbb{R}^T$ be the i-th unit vector of the canonical basis, and define the coefficient matrix $\mathbf{A}_T \in \mathbb{R}^{T\times T}$ by

$$\mathbf{A}_T \;:=\; \begin{bmatrix} a_0 & 0 & \cdots & 0 \\ a_1 & a_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{T-1} & \ldots & a_1 & a_0 \end{bmatrix} \;=\; \begin{bmatrix} 1 & 0 & \cdots & 0 \\ a_1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{T-1} & \ldots & a_1 & 1 \end{bmatrix}. \tag{16}$$

B.1. Density of the joint distribution and its gradient

Lemma B.1. *Let Assumptions 1 and 2 hold. Then, for all* $k \in \mathbb{N}_0 \cup \{\infty\}$ *, and all* $\mathbf{x} \in \mathbb{R}^T$ *, the following holds for the density of* $[X_{\ell,k+\ell}]_\ell$ *and its gradient:*

(i)

$$\begin{aligned} f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x}) &= \int_{\mathbb{R}^T} \prod_{i=1}^{T} f_\varepsilon(\mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot (\mathbf{x} - \widetilde{\mathbf{x}}))\, \mathrm{d}\mathbf{P}_{[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell}(\widetilde{\mathbf{x}}) \\ &= \mathbf{E}\left[\prod_{i=1}^{T} f_\varepsilon(\mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot [x_\ell - (X_{\ell,k+\ell} - X_{\ell,\ell})]_\ell)\right]. \end{aligned}$$

(ii)

$$\begin{aligned} &\nabla f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x}) \\ &= \left(\mathbf{A}_T^{-1}\right)^\top \left[\int_{\mathbb{R}^T} f_\varepsilon'(\mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot (\mathbf{x} - \widetilde{\mathbf{x}})) \prod_{j\neq i} f_\varepsilon(\mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot (\mathbf{x} - \widetilde{\mathbf{x}}))\, \mathrm{d}\mathbf{P}_{[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell}(\widetilde{\mathbf{x}})\right]_{i\in\{1,\ldots,T\}} \\ &= \left(\mathbf{A}_T^{-1}\right)^\top \\ &\quad \times \left[\mathbf{E}\left[f_\varepsilon'(\mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot (\mathbf{x} - [X_{\ell,k+\ell} - X_{\ell,\ell}]_\ell)) \prod_{j\neq i} f_\varepsilon(\mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot (\mathbf{x} - [X_{\ell,k+\ell} - X_{\ell,\ell}]_\ell))\right]\right]_{i\in\{1,\ldots,T\}} \end{aligned}$$

(iii) For fixed $\mathbf{z} \in \mathbb{R}^T$ *, it holds* $\mathbf{z}^\top \cdot \nabla f_{[X_{\ell,k+\ell}]_\ell} \in L^1(\mathbb{R}^T) \cap C^0(\mathbb{R}^T)$.

(iv) If Assumption 5 holds, then

$$\left(\mathbf{x} \;\mapsto\; x_{i-1} \frac{\partial}{\partial x_{i-1}} f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x})\right) \;\in\; L^1(\mathbb{R}^T) \cap C^0(\mathbb{R}^T) \qquad \textit{for all } i \in \{1, \ldots, T\}.$$

In particular, $(\mathbf{x} \mapsto \mathbf{x}^\top \cdot \nabla f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x})) \in L^1(\mathbb{R}^T) \cap C^0(\mathbb{R}^T)$.

Proof of (i):

Case $k < \infty$ **:** By the stationarity of the (truncated) time series it holds

$$f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x}) \;=\; f_{[X_{k+\ell,k+\ell}]_\ell}(\mathbf{x}) \;=\; \int_{\mathbb{R}^k} f_{[X_{\ell,\ell}]_{\ell\in\{0,\ldots,T-1+k\}}}\left(\begin{bmatrix}\mathbf{z}\\ \mathbf{x}\end{bmatrix}\right) \mathrm{d}\mathbf{z}\,.$$

We use that

$$[X_{\ell,\ell}]_{\ell\in\{0,\ldots,T-1+k\}} \;=\; \mathbf{A}_{T+k} \cdot [\varepsilon_\ell]_{\ell\in\{0,\ldots,T-1+k\}} \qquad \text{and} \qquad \det \mathbf{A}_{T+k} \;=\; 1\,,$$

to make a change of variables

$$f_{[X_{\ell,\ell}]_{\ell\in\{0,\ldots,T-1+k\}}}\left(\begin{bmatrix}\mathbf{z}\\ \mathbf{x}\end{bmatrix}\right) \;=\; f_{[\varepsilon_\ell]_{\ell\in\{0,\ldots,T-1+k\}}}\left(\mathbf{A}_{T+k}^{-1} \cdot \begin{bmatrix}\mathbf{z}\\ \mathbf{x}\end{bmatrix}\right).$$

Our goal is to simplify the expression $\mathbf{A}_{T+k}^{-1} \cdot \begin{bmatrix}\mathbf{z}\\ \mathbf{x}\end{bmatrix}$. Consider $\mathbf{A}_{T+k}$ as a block matrix, that is

$$\mathbf{A}_{T+k} \;=\; \begin{bmatrix}\mathbf{A}_k & \mathbf{0}_{k,T}\\ \mathbf{A}_{T,k} & \mathbf{A}_T\end{bmatrix} \quad \text{where} \quad \mathbf{A}_{T,k} \;:=\; \begin{bmatrix} a_k & a_{k-1} & \cdots & a_1 \\ a_{k+1} & a_k & \cdots & a_2 \\ \vdots & \vdots & \vdots & \vdots \\ a_{T-1+k} & a_{T-2+k} & \cdots & a_T \end{bmatrix} \in \mathbb{R}^{T\times k}$$

and $\mathbf{0}_{k,T}$ is the zero matrix in $\mathbb{R}^{k\times T}$. With Schur's formula for inversion of block matrices we get

$$\mathbf{A}_{T+k}^{-1} = \begin{bmatrix} \mathbf{A}_k^{-1} & \mathbf{0}_{k,T} \\ -\mathbf{A}_T^{-1}\mathbf{A}_{T,k}\mathbf{A}_k^{-1} & \mathbf{A}_T^{-1} \end{bmatrix}.$$

For all $\mathbf{z} \in \mathbb{R}^k$ and $\mathbf{x} \in \mathbb{R}^T$ it then holds

$$\mathbf{A}_{T+k}^{-1} \begin{bmatrix} \mathbf{z} \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_k^{-1}\mathbf{z} \\ \mathbf{A}_T^{-1}\left(\mathbf{x} - \mathbf{A}_{T,k}\mathbf{A}_k^{-1}\mathbf{z}\right) \end{bmatrix}.$$

Making the change of variables $\mathbf{A}_k^{-1}\mathbf{z} \mapsto \mathbf{z}$ and recalling that $\det \mathbf{A}_k^{-1} = 1$ we get

$$\begin{aligned} &\int_{\mathbb{R}^k} f_{[\varepsilon_\ell]_{\ell\in\{0,\ldots,T-1+k\}}}\left(\mathbf{A}_{T+k}^{-1}\begin{bmatrix} \mathbf{z} \\ \mathbf{x} \end{bmatrix}\right) \mathrm{d}\mathbf{z} \\ &= \int_{\mathbb{R}^k} f_{[\varepsilon_\ell]_{\ell\in\{0,\ldots,T-1+k\}}}\left(\begin{bmatrix} \mathbf{z} \\ \mathbf{A}_T^{-1}(\mathbf{x} - \mathbf{A}_{T,k}\mathbf{z}) \end{bmatrix}\right) \mathrm{d}\mathbf{z}. \end{aligned}$$

Using that the innovations are i.i.d., this equals

$$\begin{aligned} &\int_{\mathbb{R}^k} f_{[\varepsilon_\ell]_{\ell\in\{0,\ldots,T-1+k\}}}\left(\begin{bmatrix} \mathbf{z} \\ \mathbf{A}_T^{-1}(\mathbf{x} - \mathbf{A}_{T,k}\mathbf{z}) \end{bmatrix}\right) \mathrm{d}\mathbf{z} = \mathbf{E}\left[\prod_{i=1}^T f_\varepsilon\left(\mathbf{e}_i^\top \mathbf{A}_T^{-1}\left(\mathbf{x} - \mathbf{A}_{T,k}\cdot[\varepsilon_\ell]_{\ell\in\{0,\ldots,k-1\}}\right)\right)\right] \\ &= \mathbf{E}\left[\prod_{i=1}^T f_\varepsilon\left(\mathbf{e}_i^\top \mathbf{A}_T^{-1}\left(\mathbf{x} - \mathbf{A}_{T,k}\cdot[\varepsilon_{-k+\ell}]_{\ell\in\{0,\ldots,k-1\}}\right)\right)\right]. \end{aligned}$$

We have the simplification

$$\begin{aligned} \mathbf{A}_{T,k}\cdot[\varepsilon_{-k+\ell}]_{\ell\in\{0,\ldots,k-1\}} &= \left[\sum_{i=0}^{k-1} a_{k+\ell-i}\varepsilon_{-k+i}\right]_{\ell\in\{0,\ldots,T-1\}} \\ &= \left[\sum_{i=1+\ell}^{k+\ell} a_i\varepsilon_{\ell-i}\right]_{\ell\in\{0,\ldots,T-1\}} \\ &= \left[\sum_{i=0}^{k+\ell} a_i\varepsilon_{\ell-i} - \sum_{i=0}^{\ell} a_i\varepsilon_{\ell-i}\right]_{\ell\in\{0,\ldots,T-1\}} \\ &= [X_{\ell,k+\ell} - X_{\ell,\ell}]_{\ell\in\{0,\ldots,T-1\}}, \end{aligned}$$

so that

$$\begin{aligned} &\mathbf{E}\left[\prod_{i=1}^T f_\varepsilon\left(\mathbf{e}_i^\top \mathbf{A}_T^{-1}\left(\mathbf{x} - \mathbf{A}_{T,k}\cdot[\varepsilon_{-k+\ell}]_{\ell\in\{0,\ldots,k-1\}}\right)\right)\right] \\ &= \mathbf{E}\left[\prod_{i=1}^T f_\varepsilon\left(\mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1}\cdot[x_\ell - (X_{\ell,k+\ell} - X_{\ell,\ell})]_\ell\right)\right]. \end{aligned}$$

This proves the claimed identity for $f_{[X_{\ell,k+\ell}]_\ell}$ if $k < \infty$.

Case $k = \infty$**:** Using the continuity and boundedness of f_ε and the weak convergence $[X_{\ell,k+\ell} - X_{\ell,\ell}]_\ell \to [X_\ell - X_{\ell,\ell}]_\ell$ as $k \to \infty$, we find that $f_{[X_{\ell,k+\ell}]_\ell}$ converges pointwise to the candidate density

$$\mathbf{x} \mapsto \mathbf{E}\left[\prod_{i=1}^T f_\varepsilon(\mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1}\cdot(\mathbf{x} - [X_\ell - X_{\ell,\ell}]_\ell))\right].$$

To show that this is $f_{[X_\ell]_\ell}$, that is, the density of $[X_\ell]_\ell$, we also need to show that, for any bounded and continuous function $h:\mathbb{R}^T\to\mathbb{R}$, it holds that

$$\mathbf{E}[h([X_{\ell,k+\ell}]_\ell)] \;=\; \int_{\mathbb{R}^T} h(\mathbf{x}) f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x})\,\mathrm{d}\mathbf{x} \;\to\; \int_{\mathbb{R}^T} h(\mathbf{x})\mathbf{E}\left[\prod_{i=1}^T f_\varepsilon(\mathbf{e}_i^\top\cdot\mathbf{A}_T^{-1}\cdot(\mathbf{x}-[X_\ell-X_{\ell,\ell}]_\ell))\right]\mathrm{d}\mathbf{x}\,,$$

as $k\to\infty$. To this end, we apply Pratt's version of dominated convergence (Pratt, 1960). First, note that, by $\det\mathbf{A}_T^{-1}=1/(\det\mathbf{A}_T)=1$ and the result for $k<\infty$, it holds

$$\begin{aligned}\int_{\mathbb{R}^T} h(\mathbf{x}) f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x})\,\mathrm{d}\mathbf{x} \;&=\; \int_{\mathbb{R}^T} h(\mathbf{A}_T\cdot\mathbf{x}) f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{A}_T\cdot\mathbf{x})\,\mathrm{d}\mathbf{x}\\ &=\; \int_{\mathbb{R}^T} h(\mathbf{A}_T\cdot\mathbf{x})\mathbf{E}\left[\prod_{i=1}^T f_\varepsilon(x_{i-1}-\mathbf{e}_i^\top\cdot\mathbf{A}_T^{-1}\cdot[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell)\right]\mathrm{d}\mathbf{x}\,.\end{aligned}$$

By Assumption 1 and the boundedness of h, it holds

$$\begin{aligned}&\left|h(\mathbf{x})\mathbf{E}\left[\prod_{i=1}^T f_\varepsilon(x_{i-1}-\mathbf{e}_i^\top\cdot\mathbf{A}_T^{-1}\cdot[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell)\right]\right|\\ &\lesssim\; \mathbf{E}\left[\prod_{i=1}^T g_{\alpha-1}(x_{i-1}-\mathbf{e}_i^\top\cdot\mathbf{A}_T^{-1}\cdot[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell)\right] \;=:\; \varphi_k(\mathbf{x})\,.\end{aligned}$$

Since $g_{\alpha-1}$ is bounded and continuous, it follows from the weak convergence of $[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell$ to $[X_\ell-X_{\ell,\ell}]_\ell$ as $k\to\infty$ that

$$\varphi_k(\mathbf{x}) \;\to\; \varphi_\infty(\mathbf{x}) \;:=\; \mathbf{E}\left[\prod_{i=1}^T g_{\alpha-1}(x_{i-1}-\mathbf{e}_i^\top\cdot\mathbf{A}_T^{-1}\cdot[X_\ell-X_{\ell,\ell}]_\ell)\right] \qquad \text{pointwise for all } \mathbf{x}\in\mathbb{R}^T.$$

Making the shift of variables $x_{i-1}-\mathbf{e}_i^\top\cdot\mathbf{A}_T^{-1}\cdot[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell)\mapsto x_{i-1}$ for each component $i\in\{1,\dots,T\}$, we get that

$$\int_{\mathbb{R}^T}\varphi_k(\mathbf{x})\,\mathrm{d}\mathbf{x} \;=\; \int_{\mathbb{R}^T}\varphi_\infty(\mathbf{x})\,\mathrm{d}\mathbf{x} \;=\; \prod_{i=1}^T\int_{\mathbb{R}} g_{\alpha-1}(x_{i-1})\,\mathrm{d}x_{i-1} \;=\; \|g_{\alpha-1}\|_{L^1(\mathbb{R})}^T \;<\; \infty\,,$$

so that applying Pratt's version of dominated convergence (Pratt, 1960) allows us to conclude the proof.

Proof of (ii): We show this for the l-th component of the gradient, that is $\partial/\partial x_{l-1}$. To this end, we use the identity

$$\prod_{i=1}^T a_i \;-\; \prod_{i=1}^T b_i \;=\; \sum_{i=1}^T(\prod_{j<i}a_j)(a_i-b_i)(\prod_{j>i}b_j)\,,$$

with

$$a_j \;:=\; f_\varepsilon(\mathbf{e}_j^\top\mathbf{A}_T^{-1}(\mathbf{x}-[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell+h\cdot\mathbf{e}_l)) \qquad\text{and}\qquad b_j \;:=\; f_\varepsilon(\mathbf{e}_j^\top\mathbf{A}_T^{-1}(\mathbf{x}-[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell))$$

for a fixed $h\neq 0$, and write, using (i),

$$\begin{aligned}&\frac{f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x}+h\cdot\mathbf{e}_l)-f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x})}{h}\\ &=\;\sum_{i=1}^T\mathbf{E}\Bigg[\frac{f_\varepsilon(\mathbf{e}_i^\top\mathbf{A}_T^{-1}(\mathbf{x}-[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell+h\cdot\mathbf{e}_l))-f_\varepsilon(\mathbf{e}_i^\top\mathbf{A}_T^{-1}(\mathbf{x}-[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell))}{h}\\ &\qquad\times\prod_{j<i}f_\varepsilon(\mathbf{e}_j^\top\mathbf{A}_T^{-1}(\mathbf{x}-[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell+h\cdot\mathbf{e}_l))\prod_{j>i}f_\varepsilon(\mathbf{e}_j^\top\mathbf{A}_T^{-1}(\mathbf{x}-[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell))\Bigg]\end{aligned}$$

$$= \sum_{i=1}^{T} \int_{\mathbb{R}^T} \mathrm{d}\mathbf{P}_{[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell}(\widetilde{\mathbf{x}}) \; \frac{f_\varepsilon(\mathbf{e}_i^\top \mathbf{A}_T^{-1}(\mathbf{x}-\widetilde{\mathbf{x}}+h\cdot \mathbf{e}_l)) - f_\varepsilon(\mathbf{e}_i^\top \mathbf{A}_T^{-1}(\mathbf{x}-\widetilde{\mathbf{x}}))}{h}$$
$$\times \prod_{j<i} f_\varepsilon(\mathbf{e}_j^\top \mathbf{A}_T^{-1}(\mathbf{x}-\widetilde{\mathbf{x}}+h\cdot \mathbf{e}_l)) \prod_{j>i} f_\varepsilon(\mathbf{e}_j^\top \mathbf{A}_T^{-1}(\mathbf{x}-\widetilde{\mathbf{x}}))\,.$$

We show that we can apply dominated convergence to move the limit as $h \to 0$ inside the integral. To this end, note that the terms in the product $j < i$ and $j > i$ are bounded by Assumption 1. Also note that

$$\left| \frac{f_\varepsilon(\mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot (\mathbf{x}-\widetilde{\mathbf{x}}+h\cdot \mathbf{e}_l)) - f_\varepsilon(\mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot (\mathbf{x}-\widetilde{\mathbf{x}}))}{h} \right|$$
$$\leq \frac{1}{|h|} \int_{0\wedge h}^{0\vee h} \left| f'_\varepsilon(\mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot (\mathbf{x}-\widetilde{\mathbf{x}}+s\cdot \mathbf{e}_l)) \right| \mathrm{d}s$$

and $\left| f'_\varepsilon(\mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot (\mathbf{x}-\widetilde{\mathbf{x}}+s\cdot \mathbf{e}_l)) \right| \lesssim (1 \vee (|s||\mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot \mathbf{e}_l|))^{1+\gamma}$, by Assumption 1 and Lemma A.2(i). Therefore,

$$\left| \frac{f_\varepsilon(\mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot (\mathbf{x}-\widetilde{\mathbf{x}}+h\cdot \mathbf{e}_l)) - f_\varepsilon(\mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot (\mathbf{x}-\widetilde{\mathbf{x}}))}{h} \right|$$
$$\leq \frac{1}{|h|} \int_{0\wedge h}^{0\vee h} (1 \vee |s||\mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot \mathbf{e}_l|)^{1+\gamma} \,\mathrm{d}s \;\leq\; (1 \vee |h||\mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot \mathbf{e}_l|)^{1+\gamma} \;\leq\; 2$$

for h small enough, whatever the values of $\mathbf{x}$ and $\widetilde{\mathbf{x}}$. In order to identify the limit as $h \to 0$ inside the integral, we note that

$$\lim_{h\to 0} \frac{f_\varepsilon(\mathbf{e}_i^\top \mathbf{A}_T^{-1}(\mathbf{x}-\widetilde{\mathbf{x}}+h\cdot \mathbf{e}_l)) - f_\varepsilon(\mathbf{e}_i^\top \mathbf{A}_T^{-1}(\mathbf{x}-\widetilde{\mathbf{x}}))}{h}$$
$$= \frac{\partial}{\partial x_{l-1}} f_\varepsilon(\mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot (\mathbf{x}-\widetilde{\mathbf{x}})) \;=\; \mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot \mathbf{e}_l \cdot f'_\varepsilon(\mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot (\mathbf{x}-\widetilde{\mathbf{x}}))\,.$$

Besides, by the continuity of f_ε

$$\lim_{h\to 0} \prod_{j<i} f_\varepsilon(\mathbf{e}_j^\top \mathbf{A}_T^{-1}(\mathbf{x}-\widetilde{\mathbf{x}}+h\cdot \mathbf{e}_l)) \prod_{j>i} f_\varepsilon(\mathbf{e}_j^\top \mathbf{A}_T^{-1}(\mathbf{x}-\widetilde{\mathbf{x}})) \;=\; \prod_{j\neq i} f_\varepsilon(\mathbf{e}_j^\top \mathbf{A}_T^{-1}(\mathbf{x}-\widetilde{\mathbf{x}}))\,.$$

It follows that

$$\lim_{h\to 0} \frac{f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x}+h\cdot \mathbf{e}_l) - f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x})}{h}$$
$$= \sum_{i=1}^{T} \mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot \mathbf{e}_l \int_{\mathbb{R}^T} f'_\varepsilon(\mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot (\mathbf{x}-\widetilde{\mathbf{x}})) \;\cdot\; \prod_{j\neq i} f_\varepsilon(\mathbf{e}_j^\top \mathbf{A}_T^{-1}(\mathbf{x}-\widetilde{\mathbf{x}}))\,\mathrm{d}\mathbf{P}_{[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell}(\widetilde{\mathbf{x}})\,.$$

Having identified the l-th component of the gradient, we now rewrite the whole expression and make it match the announced result. It holds

$$\begin{aligned}
\nabla f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x}) \;&=\; \sum_{l=1}^{T} \mathbf{e}_l \cdot \frac{\partial}{\partial x_{l-1}} f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x}) \\
&=\; \sum_{l=1}^{T} \mathbf{e}_l \cdot \lim_{h\to 0} \frac{f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x}+h\cdot \mathbf{e}_l) - f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x})}{h} \\
&=\; \sum_{l=1}^{T} \mathbf{e}_l \sum_{i=1}^{T} \mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot \mathbf{e}_l \\
&\qquad \times \int_{\mathbb{R}^T} f'_\varepsilon(\mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot (\mathbf{x}-\widetilde{\mathbf{x}})) \;\cdot\; \prod_{j\neq i} f_\varepsilon(\mathbf{e}_j^\top \mathbf{A}_T^{-1}(\mathbf{x}-\widetilde{\mathbf{x}}))\,\mathrm{d}\mathbf{P}_{[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell}(\widetilde{\mathbf{x}})\,.
\end{aligned}$$

Note that, for any sequence $[v_i]_{i\in\{1,\dots,T\}}$,

$$\begin{aligned}\sum_{l=1}^{T}\mathbf{e}_l\sum_{i=1}^{T}\mathbf{e}_i^\top\cdot\mathbf{A}_T^{-1}\cdot\mathbf{e}_l\cdot v_i &= \left(\sum_{l=1}^{T}\mathbf{e}_l\cdot\mathbf{e}_l^\top\right)\left(\sum_{i=1}^{T}\left(\mathbf{e}_i^\top\cdot\mathbf{A}_T^{-1}\right)^\top\cdot v_i\right)\\ &= \mathbf{I}_T\cdot\left(\mathbf{A}_T^{-1}\right)^\top\sum_{i=1}^{T}\mathbf{e}_i\cdot v_i\\ &= \left(\mathbf{A}_T^{-1}\right)^\top\cdot[v_i]_{i\in\{1,\dots,T\}}\,.\end{aligned}$$

Therefore, the result follows with

$$v_i = \int_{\mathbb{R}^T} f_\varepsilon'(\mathbf{e}_i^\top\cdot\mathbf{A}_T^{-1}\cdot(\mathbf{x}-\widetilde{\mathbf{x}}))\ \cdot\ \prod_{j\neq i} f_\varepsilon(\mathbf{e}_j^\top\mathbf{A}_T^{-1}(\mathbf{x}-\widetilde{\mathbf{x}}))\,\mathrm{d}\mathbf{P}_{[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell}(\widetilde{\mathbf{x}})\,.$$

Proof of (iii): Since $\det\mathbf{A}_T^{-1}=1/(\det\mathbf{A}_T)=1$, it follows from (ii) that

$$\begin{aligned}&\int_{\mathbb{R}^T}\left|\mathbf{z}^\top\cdot\nabla f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x})\right|\mathrm{d}\mathbf{x} = \int_{\mathbb{R}^T}\left|\mathbf{z}^\top\cdot\nabla f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{A}_T\cdot\mathbf{x})\right|\mathrm{d}\mathbf{x}\\ &= \int_{\mathbb{R}^T}\mathrm{d}\mathbf{x}\\ &\left|\left(\mathbf{A}_T^{-1}\mathbf{z}\right)^\top\left[\mathbf{E}\left[f_\varepsilon'(x_{i-1}-\mathbf{e}_i^\top\cdot\mathbf{A}_T^{-1}\cdot[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell)\prod_{j\neq i}f_\varepsilon(x_{j-1}-\mathbf{e}_j^\top\cdot\mathbf{A}_T^{-1}\cdot[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell)\right]\right]_{i=1,\dots,T}\right|\\ &\lesssim \sum_{i=1}^{T}\int_{\mathbb{R}^T}\mathbf{E}\left[\left|f_\varepsilon'(x_{i-1}-\mathbf{e}_i^\top\cdot\mathbf{A}_T^{-1}\cdot[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell)\right|\prod_{j\neq i}f_\varepsilon(x_{j-1}-\mathbf{e}_j^\top\cdot\mathbf{A}_T^{-1}\cdot[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell)\right]\mathrm{d}\mathbf{x}\,,\end{aligned}$$

where the multiplicative constant in $\lesssim$ depends only on $\mathbf{z}$ and the coefficients (a_i) . Integrating out the density terms, together with Assumption 1 and Lemma A.2(i) gives, for $\gamma\in(0,\alpha-1)$, the upper bound

$$\begin{aligned}&\sum_{i=1}^{T}\int_{\mathbb{R}}\mathbf{E}\left[\left|f_\varepsilon'(x_{i-1}-\mathbf{e}_i^\top\cdot\mathbf{A}_T^{-1}\cdot[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell)\right|\right]\mathrm{d}x_{i-1}\\ &\le \sum_{i=1}^{T}\int_{\mathbb{R}}\mathbf{E}[g_\gamma(x_{i-1}-\mathbf{e}_i^\top\cdot\mathbf{A}_T^{-1}\cdot[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell)]\,\mathrm{d}x_{i-1}\\ &\le \sum_{i=1}^{T}\int_{\mathbb{R}}\mathbf{E}\left[\left(1\vee\left|\mathbf{e}_i^\top\cdot\mathbf{A}_T^{-1}\cdot[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell\right|\right)^{1+\gamma}\right]g_\gamma(x_{i-1})\,\mathrm{d}x_{i-1}\\ &< \infty\,,\end{aligned}$$

since $X_{\ell,k+\ell}-X_{\ell,\ell}\in L^{1+\gamma}(\mathbf{P})$ for all $\ell\in\{0,\dots,T-1\}$. Therefore $\mathbf{z}^\top\cdot\nabla f_{[X_{\ell,k+\ell}]_\ell}\in L^1(\mathbb{R}^T)$. Continuity follows immediately from $f_\varepsilon', f_\varepsilon\in C^0(\mathbb{R})\cap L^\infty(\mathbb{R})$ and dominated convergence.

Proof of (iv): Fix $i\in\{1,\dots,T\}$. Since $\det\mathbf{A}_T^{-1}=1/(\det\mathbf{A}_T)=1$,

$$\begin{aligned}&\int_{\mathbb{R}^T}\left|x_{i-1}\frac{\partial}{\partial x_{i-1}}f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x})\right|\mathrm{d}\mathbf{x} = \int_{\mathbb{R}^T}\left|\left(\mathbf{e}_i^\top\cdot\mathbf{A}_T\cdot\mathbf{x}\right)\cdot\mathbf{e}_i^\top\nabla f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{A}_T\cdot\mathbf{x})\right|\mathrm{d}\mathbf{x}\\ &= \int_{\mathbb{R}^T}\left|\left(\sum_{l=1}^{i}a_{i-l}x_{l-1}\right)\mathbf{e}_i^\top\nabla f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{A}_T\cdot\mathbf{x})\right|\mathrm{d}\mathbf{x} \lesssim \sum_{l=1}^{i}\int_{\mathbb{R}^T}\left|x_{l-1}\frac{\partial}{\partial x_{i-1}}f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{A}_T\cdot\mathbf{x})\right|\mathrm{d}\mathbf{x}\,,\end{aligned}\tag{17}$$

where the multiplicative constant in $\lesssim$ depends only on the coefficients (a_i). If $l \in \{1,\dots,i-1\}$, it follows from (ii) that

$$\begin{aligned}
&\int_{\mathbb{R}^T} \left| x_{l-1} \frac{\partial}{\partial x_{i-1}} f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{A}_T \cdot \mathbf{x}) \right| \mathrm{d}\mathbf{x} \\
&= \int_{\mathbb{R}^T} \mathrm{d}\mathbf{x} \\
&\quad \left| x_{l-1} \mathbf{E}\left[f_\varepsilon'(x_{i-1} - \mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot [X_{\ell,k+\ell} - X_{\ell,\ell}]_\ell) \prod_{j \neq i} f_\varepsilon(x_{j-1} - \mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot [X_{\ell,k+\ell} - X_{\ell,\ell}]_\ell) \right] \right| \\
&\leq \int_{\mathbb{R}} \mathbf{E}\left[\left| x_{l-1} \cdot f_\varepsilon(x_{l-1} - \mathbf{e}_l^\top \cdot \mathbf{A}_T^{-1} \cdot [X_{\ell,k+\ell} - X_{\ell,\ell}]_\ell) \right| \right] \mathrm{d}x_{l-1} \int_{\mathbb{R}} |f_\varepsilon'(x)| \, \mathrm{d}x \\
&\leq \left(\mathbf{E}[|\varepsilon|] + \mathbf{E}\left[|\mathbf{e}_l^\top \cdot \mathbf{A}_T^{-1} \cdot [X_{\ell,k+\ell} - X_{\ell,\ell}]_\ell| \right] \right) \int_{\mathbb{R}} |f_\varepsilon'(x)| \, \mathrm{d}x \,.
\end{aligned}$$

If $l = i$, then

$$\begin{aligned}
&\int_{\mathbb{R}^T} \left| x_{i-1} \frac{\partial}{\partial x_{i-1}} f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{A}_T \cdot \mathbf{x}) \right| \mathrm{d}\mathbf{x} \\
&= \int_{\mathbb{R}} \mathbf{E}\left[\left| x_{i-1} \cdot f_\varepsilon'(x_{i-1} - \mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot [X_{\ell,k+\ell} - X_{\ell,\ell}]_\ell) \right| \right] \mathrm{d}x_{i-1} \\
&\leq \int_{\mathbb{R}} |x_{i-1}||f_\varepsilon'(x_{i-1})| \, \mathrm{d}x_{i-1} + \mathbf{E}\left[|\mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot [X_{\ell,k+\ell} - X_{\ell,\ell}]_\ell| \right] \int_{\mathbb{R}} |f_\varepsilon'(x_{i-1})| \, \mathrm{d}x_{i-1} \,.
\end{aligned}$$

By Assumption 5,

$$\int_{\mathbb{R}} |x||f_\varepsilon'(x)| \, \mathrm{d}x \;<\; \infty \,,$$

by Assumption 1,

$$\int_{\mathbb{R}} |f_\varepsilon'(x)| \, \mathrm{d}x \;\lesssim\; \|g_{\alpha-1}(x)\|_{L^1(\mathbb{R})} \;<\; \infty \,,$$

and since $\varepsilon \in L^1(\mathbf{P})$ and $X_{\ell,k+\ell} - X_{\ell,\ell} \in L^1(\mathbf{P})$ for all $\ell \in \{0,\dots,T-1\}$, we obtain

$$\mathbf{E}[|\varepsilon|] < \infty \quad \text{and} \quad \mathbf{E}\left[|\mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot [X_{\ell,k+\ell} - X_{\ell,\ell}]_\ell| \right] \;\lesssim\; \mathbf{E}[\|[X_{\ell,k+\ell} - X_{\ell,\ell}]_\ell\|_1] \;<\; \infty \,.$$

Therefore,

$$\int_{\mathbb{R}^T} \left| x_{l-1} \frac{\partial}{\partial x_{i-1}} f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{A}_T \cdot \mathbf{x}) \right| \mathrm{d}\mathbf{x} \;<\; \infty \qquad \text{for all } l \in \{1,\dots,i\},$$

so that the statement follows from (17). □

B.2. Gradient of $G_{k,n}$

We first require an auxiliary result about L^1-continuity with respect to scaling and translation.

Lemma B.2 (L^1-continuity of scaling and translation)**.** *Let* $\mathfrak{G} \in L^1(\mathbb{R}^T)$, $\mathbf{y} \in \mathbb{R}^T$, *and* $s \in \mathbb{R} \setminus \{0\}$. *Suppose that* $(\mathbf{y}_n) \subset \mathbb{R}^T$, *and* $(s_n) \subset \mathbb{R} \setminus \{0\}$ *satisfy* $\mathbf{y}_n \to \mathbf{y}$ *and* $s_n \to s$ *as* $n \to \infty$. *Then*

$$\|\mathfrak{G}(s_n(\cdot) - \mathbf{y}_n) \;-\; \mathfrak{G}(s(\cdot) - \mathbf{y})\|_{L^1(\mathbb{R}^T)} \;\to\; 0 \qquad \textit{as } n \to \infty \,.$$

Proof. By Lemma 3.26(2) in Alt (2016), there exists a sequence of continuous functions $(\mathfrak{G}_m)$ with compact support such that

$$\|\mathfrak{G} - \mathfrak{G}_m\|_{L^1(\mathbb{R}^T)} \;\to\; 0 \qquad \text{as } m \to \infty\,.$$

We make the decomposition

$$\begin{aligned}
&\|\mathfrak{G}(s_n(\cdot) - \mathbf{y}_n) \;-\; \mathfrak{G}(s(\cdot) - \mathbf{y})\|_{L^1(\mathbb{R}^T)} \\
&\leq\; \|\mathfrak{G}(s_n(\cdot) - \mathbf{y}_n) \;-\; \mathfrak{G}(s_n(\cdot) - \mathbf{y})\|_{L^1(\mathbb{R}^T)} \\
&\quad +\; \|\mathfrak{G}(s_n(\cdot) - \mathbf{y}) \;-\; \mathfrak{G}_m(s_n(\cdot) - \mathbf{y})\|_{L^1(\mathbb{R}^T)} \;+\; \|\mathfrak{G}_m(s_n(\cdot) - \mathbf{y}) \;-\; \mathfrak{G}_m(s(\cdot) - \mathbf{y})\|_{L^1(\mathbb{R}^T)} \\
&\quad +\; \|\mathfrak{G}_m(s(\cdot) - \mathbf{y}) \;-\; \mathfrak{G}(s(\cdot) - \mathbf{y})\|_{L^1(\mathbb{R}^T)}\,.
\end{aligned}$$

By change of variables,

$$\begin{aligned}
\|\mathfrak{G}(s_n(\cdot) - \mathbf{y}_n) \;-\; \mathfrak{G}(s_n(\cdot) - \mathbf{y})\|_{L^1(\mathbb{R}^T)} &= |s_n|^{-T}\, \|\mathfrak{G}((\cdot) + (\mathbf{y} - \mathbf{y}_n)) \;-\; \mathfrak{G}\|_{L^1(\mathbb{R}^T)}\,, \\
\|\mathfrak{G}(s_n(\cdot) - \mathbf{y}) \;-\; \mathfrak{G}_m(s_n(\cdot) - \mathbf{y})\|_{L^1(\mathbb{R}^T)} &= |s_n|^{-T}\, \|\mathfrak{G} \;-\; \mathfrak{G}_m\|_{L^1(\mathbb{R}^T)}\,, \\
\|\mathfrak{G}_m(s(\cdot) - \mathbf{y}) \;-\; \mathfrak{G}(s(\cdot) - \mathbf{y})\|_{L^1(\mathbb{R}^T)} &= |s|^{-T}\, \|\mathfrak{G}_m \;-\; \mathfrak{G}\|_{L^1(\mathbb{R}^T)}\,.
\end{aligned}$$

By the continuity and compact support of $\mathfrak{G}_m$, the dominated convergence theorem provides for any fixed m

$$\|\mathfrak{G}_m(s_n(\cdot) - \mathbf{y}) \;-\; \mathfrak{G}_m(s(\cdot) - \mathbf{y})\|_{L^1(\mathbb{R}^T)} \;\to\; 0 \qquad \text{as } n \to \infty\,,$$

and by Lemma 4.15(1) in Alt (2016)

$$\|\mathfrak{G}((\cdot) + (\mathbf{y} - \mathbf{y}_n)) \;-\; \mathfrak{G}\|_{L^1(\mathbb{R}^T)} \;\to\; 0 \qquad \text{as } n \to \infty\,.$$

It follows that

$$\limsup_{n\to\infty} \|\mathfrak{G}(s_n(\cdot) - \mathbf{y}_n) \;-\; \mathfrak{G}(s(\cdot) - \mathbf{y})\|_{L^1(\mathbb{R}^T)} \;\leq\; 2|s|^{-T}\, \|\mathfrak{G} - \mathfrak{G}_m\|_{L^1(\mathbb{R}^T)} \;\to\; 0 \qquad \text{as } m \to \infty\,.$$

This proves the claimed convergence. □

Lemma B.3 (Gradient of $G_{k,n}$). *Let Assumptions 1 and 2 hold. Then, for all $k \in \mathbb{N}_0 \cup \{\infty\}$, all sets $A \subset \mathbb{R}^T$, the following hold:*

(i) For all $s_1, s_2 \in \mathbb{R} \cup \{\infty\}$ satisfying $s_1 < s_2$,

$$\mathbf{y} \;\mapsto\; G_{k,n}(\mathbf{y}, s_1, s_2, A) \;:=\; \mathbf{E}[G_n([X_{\ell,k+\ell}]_\ell + \mathbf{y}, s_1, s_2, A)] \;=\; \int_{\mathbb{R}^T} G_n(\mathbf{x}, s_1, s_2, A) f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x} - \mathbf{y})\, \mathrm{d}\mathbf{x}$$

is continuously differentiable, with gradient

$$\mathbf{y} \;\mapsto\; \nabla_{\mathbf{y}} G_{k,n}(\mathbf{y}, s_1, s_2, A) \;=\; -\int_{\mathbb{R}^T} G_n(\mathbf{x}, s_1, s_2, A) \nabla f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x} - \mathbf{y})\, \mathrm{d}\mathbf{x}\,.$$

(ii) If Assumption 5 holds, then

$$\mathbb{R}^T \times (0, \infty) \;\to\; \mathbb{R}\,, \qquad (\mathbf{y}, s) \;\mapsto\; G_{k,n}(\mathbf{y}, s, \infty, A)$$

is continuously differentiable, with partial derivatives with respect to $\mathbf{y}$ given in (i), and with partial derivative with respect to s given as

$$\begin{aligned}
&\frac{\partial}{\partial s} G_{k,n}(\mathbf{y}, s, \infty, A) \\
&= \frac{1}{s}\left(T\, \mathbf{P}[[X_{\ell,k+\ell}]_\ell + \mathbf{y} \in u_n \cdot s \cdot A] \;+\; \int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n \cdot s \cdot A\} \cdot \mathbf{x}^\top \cdot \nabla f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x} - \mathbf{y})\, \mathrm{d}\mathbf{x} \right).
\end{aligned}$$

Proof. We prove the two statements separately.

Proof of (i): We calculate each partial derivative $\partial/\partial y_{i-1}$ for $i \in \{1, \ldots, T\}$. To streamline the analysis, we write throughout $G_n(\mathbf{x})$ for $G_n(\mathbf{x}, s_1, s_2, A)$. For all $i \in \{1, \ldots, T\}$ and $h \neq 0$,

$$\begin{aligned}&\frac{\mathbf{E}[G_n([X_{\ell,k+\ell}]_\ell + \mathbf{y} + h\mathbf{e}_i)] - \mathbf{E}[G_n([X_{\ell,k+\ell}]_\ell + \mathbf{y})]}{h}\\ &= \int_{\mathbb{R}^T} G_n(\mathbf{x}) \frac{f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x} - \mathbf{y} - h\mathbf{e}_i) - f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x} - \mathbf{y})}{h} \, \mathrm{d}\mathbf{x},\end{aligned}$$

so that, if we may interchange the limit $h \to 0$ with the integral, we find

$$\begin{aligned}&\frac{\partial}{\partial y_{i-1}} \mathbf{E}[G_n([X_{\ell,k+\ell}]_\ell + \mathbf{y})]\\ &= -\int_{\mathbb{R}^T} G_n(\mathbf{x}) \lim_{h \to 0} \frac{f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x} - \mathbf{y} - h\mathbf{e}_i) - f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x} - \mathbf{y})}{-h} \, \mathrm{d}\mathbf{x}\\ &= -\int_{\mathbb{R}^T} G_n(\mathbf{x}) \frac{\partial}{\partial z_{i-1}} f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{z})\Big|_{\mathbf{z}=\mathbf{x}-\mathbf{y}} \, \mathrm{d}\mathbf{x},\end{aligned}$$

as required. We now show that we may interchange limit and integral. Since

$$\frac{f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x} - \mathbf{y}) - f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x} - \mathbf{y} - h\mathbf{e}_i)}{h} = \int_0^1 \frac{\partial}{\partial z_{i-1}} f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{z})\Big|_{\mathbf{z}=\mathbf{x}-\mathbf{y}-\xi h\mathbf{e}_i} \, \mathrm{d}\xi,$$

it holds, by Tonelli's theorem,

$$\begin{aligned}&\left| \frac{\mathbf{E}[G_n([X_{\ell,k+\ell}]_\ell + \mathbf{y} + h\mathbf{e}_i)] - \mathbf{E}[G_n([X_{\ell,k+\ell}]_\ell + \mathbf{y})]}{h} + \int_{\mathbb{R}^T} G_n(\mathbf{x}) \frac{\partial}{\partial z_{i-1}} f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{z})\Big|_{\mathbf{z}=\mathbf{x}-\mathbf{y}} \mathrm{d}\mathbf{x} \right|\\ &\leq \int_{\mathbb{R}^T} G_n(\mathbf{x}) \left(\int_0^1 \left| \frac{\partial}{\partial z_{i-1}} f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{z})\Big|_{\mathbf{z}=\mathbf{x}-\mathbf{y}} - \frac{\partial}{\partial z_{i-1}} f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{z})\Big|_{\mathbf{z}=\mathbf{x}-\mathbf{y}-\xi h\mathbf{e}_i} \right| \mathrm{d}\xi \right) \mathrm{d}\mathbf{x}\\ &\leq \int_0^1 \left\| \frac{\partial}{\partial z_{i-1}} f_{[X_{\ell,k+\ell}]_\ell}\Big|_{\mathbf{z}=(\cdot)} - \frac{\partial}{\partial z_{i-1}} f_{[X_{\ell,k+\ell}]_\ell}\Big|_{\mathbf{z}=(\cdot)-\xi h\mathbf{e}_i} \right\|_{L^1(\mathbb{R}^T)} \mathrm{d}\xi.\end{aligned}$$

Since $\partial f_{[X_{\ell,k+\ell}]_\ell}/\partial z_{i-1} \in L^1(\mathbb{R}^T)$ by Lemma B.1(iii), it follows from Lemma B.2 that

$$\int_0^1 \left\| \frac{\partial}{\partial z_{i-1}} f_{[X_{\ell,k+\ell}]_\ell}\Big|_{\mathbf{z}=(\cdot)} - \frac{\partial}{\partial z_{i-1}} f_{[X_{\ell,k+\ell}]_\ell}\Big|_{\mathbf{z}=(\cdot)-\xi h\mathbf{e}_i} \right\|_{L^1(\mathbb{R}^T)} \mathrm{d}\xi \to 0$$

as $h \to 0$.

Proof of continuity in part (i): Let $\mathbf{h} \in \mathbb{R}^T$. Since $\partial f_{[X_{\ell,k+\ell}]_\ell}/\partial z_{i-1} \in L^1(\mathbb{R}^T)$ by Lemma B.1(iii), it follows from Lemma B.2 that

$$\begin{aligned}&\left\| \frac{\partial}{\partial z_{i-1}} G_{k,n}\Big|_{\mathbf{z}=(\cdot)} - \frac{\partial}{\partial z_{i-1}} G_{k,n}\Big|_{\mathbf{z}=(\cdot)+\mathbf{h}} \right\|_{L^\infty(\mathbb{R}^T)}\\ &\leq \left\| \frac{\partial}{\partial z_{i-1}} f_{[X_{\ell,k+\ell}]_\ell}\Big|_{\mathbf{z}=(\cdot)} - \frac{\partial}{\partial z_{i-1}} f_{[X_{\ell,k+\ell}]_\ell}\Big|_{\mathbf{z}=(\cdot)-\mathbf{h}} \right\|_{L^1(\mathbb{R}^T)} \to 0,\end{aligned}$$

as $\mathbf{h} \to 0$. This proves (uniform) continuity.

Proof of (ii): To streamline the proof, we write $G_{k,n}(\mathbf{y}, s)$ for $G_{k,n}(\mathbf{y}, s, \infty, A)$. By the change of variables $\mathbf{x} \mapsto s\mathbf{x}$,

$$G_{k,n}(\mathbf{y}, s) = s^T \int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n A\} f_{[X_{\ell,k+\ell}]_\ell}(s\mathbf{x} - \mathbf{y}) \, \mathrm{d}\mathbf{x}.$$

Existence of $\frac{\partial}{\partial s} G_{k,n}(\mathbf{y}, s)$ in part (ii): The partial derivative with respect to $s > 0$, if it exists, is

$$\begin{aligned}
&\frac{\partial}{\partial s} G_{k,n}(\mathbf{y}, s) \\
&= T s^{T-1} \int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n A\} f_{[X_{\ell,k+\ell}]_\ell}(s\mathbf{x} - \mathbf{y}) \,\mathrm{d}\mathbf{x} \;+\; s^T \frac{\partial}{\partial s} \left(\int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n A\} f_{[X_{\ell,k+\ell}]_\ell}(s\mathbf{x} - \mathbf{y}) \,\mathrm{d}\mathbf{x} \right).
\end{aligned}$$

It therefore remains to show, for the second term, that we can swap differentiation and integration. Since

$$\frac{\partial}{\partial s} \left(f_{[X_{\ell,k+\ell}]_\ell}(s\mathbf{x} - \mathbf{y}) \right) \;=\; \mathbf{x}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(s\mathbf{x} - \mathbf{y})$$

if we are allowed to swap differentiation and integration, we will then get, as required,

$$\begin{aligned}
&\frac{\partial}{\partial s} G_{k,n}(\mathbf{y}, s) \\
&= T s^{T-1} \int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n A\} f_{[X_{\ell,k+\ell}]_\ell}(s\mathbf{x} - \mathbf{y}) \,\mathrm{d}\mathbf{x} \;+\; s^T \int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n A\} \mathbf{x}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(s\mathbf{x} - \mathbf{y}) \,\mathrm{d}\mathbf{x} \\
&= \frac{1}{s} \left(T\, \mathbf{P}[[X_{\ell,k+\ell}]_\ell + \mathbf{y} \in u_n s A] \;+\; \int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n s A\} \mathbf{x}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x} - \mathbf{y}) \,\mathrm{d}\mathbf{x} \right).
\end{aligned}$$

Let us focus on the swap. Let $h \in \mathbb{R}$ satisfy $|h| < s/2$. Then

$$\begin{aligned}
&\frac{1}{h} \left(\int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n A\} f_{[X_{\ell,k+\ell}]_\ell}((s+h)\mathbf{x} - \mathbf{y}) \,\mathrm{d}\mathbf{x} \;-\; \int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n A\} f_{[X_{\ell,k+\ell}]_\ell}(s\mathbf{x} - \mathbf{y}) \,\mathrm{d}\mathbf{x} \right) \\
&= \int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n A\} \left(\int_0^1 \mathbf{x}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}((s+h\xi)\mathbf{x} - \mathbf{y}) \,\mathrm{d}\xi \right) \mathrm{d}\mathbf{x} \\
&= \int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n A\} \left(\int_0^1 \frac{((s+h\xi)\mathbf{x})^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}((s+h\xi)\mathbf{x} - \mathbf{y})}{s+h\xi} \,\mathrm{d}\xi \right) \mathrm{d}\mathbf{x} \\
&= \int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n A\} \left(\int_0^1 \frac{\mathfrak{H}((s+h\xi)\mathbf{x} - \mathbf{y})}{s+h\xi} \,\mathrm{d}\xi \right) \mathrm{d}\mathbf{x} \\
&\quad + \int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n A\} \left(\int_0^1 \frac{\mathbf{y}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}((s+h\xi)\mathbf{x} - \mathbf{y})}{s+h\xi} \,\mathrm{d}\xi \right) \mathrm{d}\mathbf{x}, \qquad (18)
\end{aligned}$$

where

$$\mathfrak{H}(\mathbf{x}) \;:=\; \mathbf{x}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x}), \qquad \mathbf{x} \in \mathbb{R}^T.$$

The function $\mathfrak{H}$ is integrable and continuous, by Lemma B.1(iv). We expand

$$\begin{aligned}
&\int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n A\} \mathbf{x}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(s\mathbf{x} - \mathbf{y}) \,\mathrm{d}\mathbf{x} \\
&= \int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n A\} \left(\int_0^1 \frac{\mathfrak{H}(s\mathbf{x} - \mathbf{y})}{s} \,\mathrm{d}\xi \right) \mathrm{d}\mathbf{x} \;+\; \int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n A\} \int_0^1 \left(\frac{\mathbf{y}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(s\mathbf{x} - \mathbf{y})}{s} \,\mathrm{d}\xi \right) \mathrm{d}\mathbf{x}. \qquad (19)
\end{aligned}$$

It then holds, by (18) and (19),

$$\begin{aligned}
&\left| \frac{1}{h}\left(\int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x}\in u_nA\} f_{[X_{\ell,k+\ell}]_\ell}((s+h)\mathbf{x}-\mathbf{y})\,\mathrm{d}\mathbf{x} \;-\; \int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x}\in u_nA\} f_{[X_{\ell,k+\ell}]_\ell}(s\mathbf{x}-\mathbf{y})\,\mathrm{d}\mathbf{x}\right)\right.\\
&\qquad \left. -\int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x}\in u_nA\}\mathbf{x}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(s\mathbf{x}-\mathbf{y})\,\mathrm{d}\mathbf{x}\right|\\
&\le \int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x}\in u_nA\}\Bigg(\int_0^1 \left|\frac{\mathfrak{H}(s\mathbf{x}-\mathbf{y})}{s} - \frac{\mathfrak{H}((s+h\xi)\mathbf{x}-\mathbf{y})}{s+h\xi}\right|\\
&\qquad + \left|\frac{\mathbf{y}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(s\mathbf{x}-\mathbf{y})}{s} - \frac{\mathbf{y}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}((s+h\xi)\mathbf{x}-\mathbf{y})}{s+h\xi}\right|\,\mathrm{d}\xi\Bigg)\,\mathrm{d}\mathbf{x}\,.
\end{aligned}$$

Note that

$$\begin{aligned}
&\left|\frac{\mathfrak{H}(s\mathbf{x}-\mathbf{y})}{s} - \frac{\mathfrak{H}((s+h\xi)\mathbf{x}-\mathbf{y})}{s+h\xi}\right| + \left|\frac{\mathbf{y}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(s\mathbf{x}-\mathbf{y})}{s} - \frac{\mathbf{y}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}((s+h\xi)\mathbf{x}-\mathbf{y})}{s+h\xi}\right|\\
&\le \frac{1}{s}\Big(|\mathfrak{H}(s\mathbf{x}-\mathbf{y}) - \mathfrak{H}((s+h\xi)\mathbf{x}-\mathbf{y})| + \left|\mathbf{y}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(s\mathbf{x}-\mathbf{y}) - \mathbf{y}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}((s+h\xi)\mathbf{x}-\mathbf{y})\right|\Big)\\
&\quad + \left|\frac{1}{s} - \frac{1}{s+h\xi}\right|\Big(|\mathfrak{H}((s+h\xi)\mathbf{x}-\mathbf{y})| + \left|\mathbf{y}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}((s+h\xi)\mathbf{x}-\mathbf{y})\right|\Big)\,.
\end{aligned}$$

Therefore, by Tonelli's theorem,

$$\begin{aligned}
&\int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x}\in u_nA\}\Bigg(\int_0^1 \left|\frac{\mathfrak{H}(s\mathbf{x}-\mathbf{y})}{s} - \frac{\mathfrak{H}((s+h\xi)\mathbf{x}-\mathbf{y})}{s+h\xi}\right|\\
&\qquad + \left|\frac{\mathbf{y}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(s\mathbf{x}-\mathbf{y})}{s} - \frac{\mathbf{y}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}((s+h\xi)\mathbf{x}-\mathbf{y})}{s+h\xi}\right|\,\mathrm{d}\xi\Bigg)\,\mathrm{d}\mathbf{x}\\
&\le \int_0^1 \Bigg(\frac{1}{s}\Big(\|\mathfrak{H}(s(\cdot)-\mathbf{y}) - \mathfrak{H}((s+h\xi)(\cdot)-\mathbf{y})\|_{L^1(\mathbb{R}^T)}\\
&\qquad + \left\|\mathbf{y}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(s(\cdot)-\mathbf{y}) - \mathbf{y}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}((s+h\xi)(\cdot)-\mathbf{y})\right\|_{L^1(\mathbb{R}^T)}\Big)\\
&\qquad + \left|\frac{1}{s} - \frac{1}{s+h\xi}\right|\Big(\|\mathfrak{H}((s+h\xi)(\cdot)-\mathbf{y})\|_{L^1(\mathbb{R}^T)}\\
&\qquad\quad + \left\|\mathbf{y}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}((s+h\xi)(\cdot)-\mathbf{y})\right\|_{L^1(\mathbb{R}^T)}\Big)\Bigg)\,\mathrm{d}\xi.
\end{aligned} \tag{20}$$

Note that

$$\begin{aligned}
&\|\mathfrak{H}((s+h\xi)(\cdot)-\mathbf{y})\|_{L^1(\mathbb{R}^T)} + \left\|\mathbf{y}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}((s+h\xi)(\cdot)-\mathbf{y})\right\|_{L^1(\mathbb{R}^T)}\\
&= |s+h\xi|^{-T}\left(\|\mathfrak{H}\|_{L^1(\mathbb{R}^T)} + \left\|\mathbf{y}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}\right\|_{L^1(\mathbb{R}^T)}\right),
\end{aligned}$$

so that

$$\begin{aligned}
&\int_0^1 \left|\frac{1}{s} - \frac{1}{s+h\xi}\right|\left(\|\mathfrak{H}((s+h\xi)(\cdot)-\mathbf{y})\|_{L^1(\mathbb{R}^T)} + \left\|\mathbf{y}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}((s+h\xi)(\cdot)-\mathbf{y})\right\|_{L^1(\mathbb{R}^T)}\right)\mathrm{d}\xi\\
&\le \left(\|\mathfrak{H}\|_{L^1(\mathbb{R}^T)} + \left\|\mathbf{y}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}\right\|_{L^1(\mathbb{R}^T)}\right)\sup_{\xi\in(0,1)}\left(\left|\frac{1}{s} - \frac{1}{s+h\xi}\right| |s+h\xi|^{-T}\right) \to 0
\end{aligned}$$

as $h \to 0$. Let us deal with the first pair of integrals in the right-hand side of (20). Since $\mathfrak{H}$ and $\mathbf{y}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell} \in L^1(\mathbb{R}^T) \cap C^0(\mathbb{R}^T)$, it follows from Lemma B.2 that

$$\begin{aligned} &\int_0^1 \frac{1}{s}\bigg(\|\mathfrak{H}(s(\cdot)-\mathbf{y}) \;-\; \mathfrak{H}((s+h\xi)(\cdot)-\mathbf{y})\|_{L^1(\mathbb{R}^T)} \\ &\qquad\qquad + \left\|\mathbf{y}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(s(\cdot)-\mathbf{y}) \;-\; \mathbf{y}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}((s+h\xi)(\cdot)-\mathbf{y})\right\|_{L^1(\mathbb{R}^T)}\bigg)\,\mathrm{d}\xi \\ &\to\; 0 \end{aligned}$$

as $h \to 0$, which proves the claim.

Proof of continuity of $\nabla_{\mathbf{y}} G_{k,n}$ in part (ii): Let $(\mathbf{y}, s) \in \mathbb{R}^T \times (0,\infty)$ and $(\mathbf{y}_m, s_m)_{m\in\mathbb{N}} \subset \mathbb{R}^T \times (s/2, 3s/2)$ with $(\mathbf{y}_m, s_m) \to (\mathbf{y}, s)$ as $m \to \infty$. By part (i) and the change of variables $\mathbf{x} \mapsto s\mathbf{x}$,

$$\nabla_{\mathbf{y}} G_{k,n}(\mathbf{y}, s) \;=\; -s^T \int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n A\} \nabla f_{[X_{\ell,k+\ell}]_\ell}(s\mathbf{x}-\mathbf{y})\,\mathrm{d}\mathbf{x}\,.$$

Then, for all $\mathbf{z} \in \mathbb{R}^T$,

$$\begin{aligned} &\left|\mathbf{z}^\top \nabla_{\mathbf{y}} G_{k,n}(\mathbf{y}, s) \;-\; \mathbf{z}^\top \nabla_{\mathbf{y}} G_{k,n}(\mathbf{y}_m, s_m)\right| \\ &\le\; |s^T - s_m^T|\, \left\|\mathbf{z}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(s(\cdot)-\mathbf{y})\right\|_{L^1(\mathbb{R}^T)} \\ &\quad +\; |s_m|^T \left\|\mathbf{z}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(s_m(\cdot)-\mathbf{y}_m) \;-\; \mathbf{z}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(s(\cdot)-\mathbf{y})\right\|_{L^1(\mathbb{R}^T)} \\ &\to\; 0\,, \end{aligned}$$

as $m \to \infty$, with similar arguments as before.

Proof of continuity of $\frac{\partial}{\partial s} G_{k,n}$ in part (ii): To prove continuity of the partial derivative in s, note that

$$\begin{aligned} &\frac{\partial}{\partial s} G_{k,n}(\mathbf{y}_m, s_m) \\ &=\; \frac{1}{s_m}\left(T\,\mathbf{P}[[X_{\ell,k+\ell}]_\ell + \mathbf{y}_m \in u_n s_m A] \;+\; \int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n s_m A\}\mathbf{x}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x}-\mathbf{y}_m)\,\mathrm{d}\mathbf{x}\right). \end{aligned}$$

For the first term, we have

$$\mathbf{P}[[X_{\ell,k+\ell}]_\ell + \mathbf{y}_m \in u_n s_m A] \;=\; s_m^T \int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n A\} f_{[X_{\ell,k+\ell}]_\ell}(s_m\mathbf{x}-\mathbf{y}_m)\,\mathrm{d}\mathbf{x}$$

and, by Lemma B.2,

$$\begin{aligned} &\left|\int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n A\} f_{[X_{\ell,k+\ell}]_\ell}(s_m\mathbf{x}-\mathbf{y}_m)\,\mathrm{d}\mathbf{x} \;-\; \int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n A\} f_{[X_{\ell,k+\ell}]_\ell}(s\mathbf{x}-\mathbf{y})\,\mathrm{d}\mathbf{x}\right| \\ &\le\; \left\|f_{[X_{\ell,k+\ell}]_\ell}(s_m(\cdot)-\mathbf{y}_m) \;-\; f_{[X_{\ell,k+\ell}]_\ell}(s(\cdot)-\mathbf{y})\right\|_{L^1(\mathbb{R}^T)} \;\to\; 0\,, \end{aligned}$$

so that

$$\frac{1}{s_m} T\,\mathbf{P}[[X_{\ell,k+\ell}]_\ell + \mathbf{y}_m \in u_n s_m A] \;\to\; \frac{1}{s} T\,\mathbf{P}[[X_{\ell,k+\ell}]_\ell + \mathbf{y} \in u_n s A]$$

as $m \to \infty$. For the second term, note that

$$\int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n s_m A\}\mathbf{x}^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x}-\mathbf{y}_m)\,\mathrm{d}\mathbf{x} \;=\; s_m^T \int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n A\}(s_m\mathbf{x})^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(s_m\mathbf{x}-\mathbf{y}_m)\,\mathrm{d}\mathbf{x}\,,$$

so that it remains to prove that

$$\left| \int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n A\} (s_m \mathbf{x})^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(s_m \mathbf{x} - \mathbf{y}_m)\, \mathrm{d}\mathbf{x} \;-\; \int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n A\} (s \mathbf{x})^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(s \mathbf{x} - \mathbf{y})\, \mathrm{d}\mathbf{x} \right| \;\to\; 0$$

as $m \to \infty$. Note that for m sufficiently large $\|\mathbf{y}_m - \mathbf{y}\|_\infty \le 1$, so that

$$\begin{aligned}
&\left| \int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n A\} (s_m \mathbf{x})^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(s_m \mathbf{x} - \mathbf{y}_m)\, \mathrm{d}\mathbf{x} \;-\; \int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x} \in u_n A\} (s \mathbf{x})^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(s \mathbf{x} - \mathbf{y})\, \mathrm{d}\mathbf{x} \right| \\
&\le \left\| s_m(\cdot)^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(s_m(\cdot) - \mathbf{y}_m) \;-\; s(\cdot)^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(s(\cdot) - \mathbf{y}) \right\|_{L^1(\mathbb{R}^T)} \\
&\le \left\| \mathfrak{H}(s_m(\cdot) - \mathbf{y}_m) \;-\; \mathfrak{H}(s(\cdot) - \mathbf{y}) \right\|_{L^1(\mathbb{R}^T)} \\
&\quad + \left\| \mathbf{y}^\top \left(\nabla f_{[X_{\ell,k+\ell}]_\ell}(s_m(\cdot) - \mathbf{y}_m) \;-\; \nabla f_{[X_{\ell,k+\ell}]_\ell}(s(\cdot) - \mathbf{y}) \right) \right\|_{L^1(\mathbb{R}^T)} \\
&\quad + \left\| (\mathbf{y}_m - \mathbf{y})^\top \left(\nabla f_{[X_{\ell,k+\ell}]_\ell}(s_m(\cdot) - \mathbf{y}_m) \;-\; \nabla f_{[X_{\ell,k+\ell}]_\ell}(s(\cdot) - \mathbf{y}) \right) \right\|_{L^1(\mathbb{R}^T)} \\
&\quad + \left\| (\mathbf{y}_m - \mathbf{y})^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(s(\cdot) - \mathbf{y}) \right\|_{L^1(\mathbb{R}^T)} \\
&\le \left\| \mathfrak{H}(s_m(\cdot) - \mathbf{y}_m) \;-\; \mathfrak{H}(s(\cdot) - \mathbf{y}) \right\|_{L^1(\mathbb{R}^T)} \\
&\quad + \left\| \mathbf{y}^\top \left(\nabla f_{[X_{\ell,k+\ell}]_\ell}(s_m(\cdot) - \mathbf{y}_m) \;-\; \nabla f_{[X_{\ell,k+\ell}]_\ell}(s(\cdot) - \mathbf{y}) \right) \right\|_{L^1(\mathbb{R}^T)} \\
&\quad + \sum_{j=1}^{T} \left\| \mathbf{e}_j^\top \left(\nabla f_{[X_{\ell,k+\ell}]_\ell}(s_m(\cdot) - \mathbf{y}_m) \;-\; \nabla f_{[X_{\ell,k+\ell}]_\ell}(s(\cdot) - \mathbf{y}) \right) \right\|_{L^1(\mathbb{R}^T)} \\
&\quad + \sum_{j=1}^{T} \left\| \mathbf{e}_j^\top \nabla f_{[X_{\ell,k+\ell}]_\ell}(s(\cdot) - \mathbf{y}) \right\|_{L^1(\mathbb{R}^T)} \|\mathbf{y}_m - \mathbf{y}\|_\infty \,.
\end{aligned}$$

By Lemma B.2, this upper bound tends to 0 as $m \to \infty$. □

B.3. Central limit theorem for multivariate partial sums

Next we provide a proof of Theorem 3.1, which extends a well known central limit theorem for univariate partial sums of long memory linear time series to the multivariate setting relevant in this work.

Proof of Theorem 3.1: Since

$$\mathbf{E}\left[\left\| n^{-d-1/\alpha} \left(\sum_{t=1}^{n} [X_{t+\ell}]_\ell \;-\; \sum_{t=1}^{n-T+1} [X_{t+\ell}]_\ell \right) \right\|_1 \right] \;\le\; n^{-d-1/\alpha} \cdot T^2 \cdot \mathbf{E}[|X_0|] \;\to\; 0 \qquad \text{as } n \to \infty\,,$$

we show the result without loss of generality for

$$\sum_{t=1}^{n} [X_{t+\ell}]_\ell \qquad \text{instead of} \qquad \sum_{t=1}^{n-T+1} [X_{t+\ell}]_\ell \,.$$

Define

$$[\Delta_{t,t+\ell}]_\ell \;:=\; [X_{t+\ell} - X_t]_\ell \;=\; [X_{t+\ell}]_\ell \;-\; X_t \cdot \mathbf{1}_T \qquad \text{for } t \ge 1\,.$$

Then

$$\sum_{t=1}^{n} [X_{t+\ell}]_\ell \;=\; \left(\sum_{t=1}^{n} X_t \right) \cdot \mathbf{1}_T \;+\; \sum_{t=1}^{n} [\Delta_{t,t+\ell}]_\ell \,.$$

We show that, as $n \to \infty$,

$$n^{-d-1/\alpha} \sum_{t=1}^{n} [\Delta_{t,t+\ell}]_\ell \ \to_{\mathbf{P}} \ \mathbf{0}_T \,. \tag{21}$$

Note that by a telescoping sum argument

$$\sum_{t=1}^{n} [\Delta_{t,t+\ell}]_\ell \ = \ \left[\sum_{t=1}^{n} (X_{t+\ell} \ - \ X_t) \right]_\ell \ = \ \left[\sum_{t=1}^{\ell} (X_{n+t} \ - \ X_t) \right]_\ell \,.$$

Since for all $\ell \in \{0, \ldots, T-1\}$

$$\mathbf{E}\left[\left| n^{-d-1/\alpha} \sum_{t=1}^{\ell} (X_{n+t} \ - \ X_t) \right| \right] \ \leq \ n^{-d-1/\alpha} \cdot (2 \cdot T \cdot \mathbf{E}[|X_0|]) \ \to \ 0 \qquad \text{as } n \to \infty \,,$$

it readily follows (21). The result hence follows from Slutsky's theorem combined with the standard asymptotic theory for $n^{-d-1/\alpha} \sum_{t=1}^{n} X_t$, see, for example, Theorem 3.1 in Scheffel et al. (2026). $\square$

C. Separable product bounds featuring u_n and (s_1, s_2)

By (10), there exist $\{\mathfrak{a}_j\} \subset (0, \infty)$ such that

$$\begin{aligned} G_n(\mathbf{A}_T \cdot \mathbf{z}, s_1, s_2, A) \ &= \ \mathbb{1}\{\mathbf{A}_T \cdot \mathbf{z} \in u_n \cdot ((s_1 \cdot A) \setminus (s_2 \cdot A))\} \\ &\leq \ \sum_{j=1}^{T} \mathbb{1}\left\{ \left| \sum_{m=1}^{j} a_{j-m} z_{m-1} \right| \in u_n \cdot \mathfrak{a}_j \cdot [s_1, s_2] \right\} . \end{aligned}$$

The main purpose of this section is to separate the threshold u_n from the uniformity interval $[s_1, s_2]$ in a way that preserves the structure required for the later estimates. We formulate the argument for general index sets $I \subset \{1, \ldots, T\}$ and coefficients $(c_m)_{m \in I}$, where in the later sections we mostly have $I = \{1, \ldots, T\}$ and $c_m = a_{j-m}$. The first lemma contains the basic separation argument. In the rest of the section we then build the complexity needed for the subsequent analysis around this.

Lemma C.1 (Separable product bound I). *Let $\emptyset \neq I \subset \{1, \ldots, T\}$, $(c_m)_{m \in I} \subset \mathbb{R}$, $\mathfrak{a} \in (0, \infty)$, $\gamma \in (0, \infty)$. Then*

$$\int_{\mathbb{R}^{|I|}} \mathbb{1}\left\{ \left| \sum_{m \in I} c_m z_{m-1} \right| \in u_n \cdot \mathfrak{a} \cdot [s_1, s_2] \right\} \cdot \prod_{m \in I} g_\gamma(z_{m-1}) \, \mathrm{d}\mathbf{z} \ \lesssim \ u_n^{-\gamma} \cdot ((s_2 - s_1) \wedge 1) \,,$$

where $\lesssim$ means inequality up to a constant that only depends on T, (c_m), $\mathfrak{a}$, γ, and $\underline{s}$, and is independent of n and s_1, s_2.

Proof. Since

$$\begin{aligned} &\int_{\mathbb{R}^{|I|}} \mathbb{1}\left\{ \left| \sum_{m \in I} c_m z_{m-1} \right| \in u_n \cdot \mathfrak{a} \cdot [s_1, s_2] \right\} \cdot \prod_{m \in I} g_\gamma(z_{m-1}) \, \mathrm{d}\mathbf{z} \\ &\leq \int_{\mathbb{R}^{|I|}} \left(\mathbb{1}\left\{ \sum_{m \in I} c_m z_{m-1} \in u_n \cdot \mathfrak{a} \cdot [s_1, s_2] \right\} \ + \ \mathbb{1}\left\{ \sum_{m \in I} (-c_m) z_{m-1} \in u_n \cdot \mathfrak{a} \cdot [s_1, s_2] \right\} \right) \cdot \prod_{m \in I} g_\gamma(z_{m-1}) \, \mathrm{d}\mathbf{z} \,, \end{aligned}$$

it is sufficient to prove

$$\int_{\mathbb{R}^{|I|}} \mathbb{1}\left\{\sum_{m\in I} c_m z_{m-1} \in u_n \cdot \mathfrak{a} \cdot [s_1, s_2]\right\} \cdot \prod_{m\in I} g_\gamma\left(z_{m-1}\right) \,\mathrm{d}\mathbf{z} \;\lesssim\; u_n^{-\gamma} \cdot ((s_2 - s_1) \wedge 1)\;,$$

for $(c_m)_{m\in I} \subset \mathbb{R}$, since this automatically covers $(-c_m)_{m\in I} \subset \mathbb{R}$.

First consider $I \subset \{1, \ldots, T\}$ with $|I| = 1$. Without loss of generality, we assume $1 \in I$, $c_1 \neq 0$. Making the change of variables

$$z_0 \;\mapsto\; \frac{u_n \cdot \mathfrak{a}}{c_1} z_0$$

and using Lemma A.2(iii), we get

$$\begin{aligned}
&\int_{\mathbb{R}} \mathbb{1}\{c_1 z_0 \in u_n \cdot \mathfrak{a} \cdot [s_1, s_2]\} g_\gamma(z_0) \,\mathrm{d}z_0 \;=\; u_n \frac{\mathfrak{a}}{|c_1|} \int_{\mathbb{R}} \mathbb{1}\{z_0 \in [s_1, s_2]\} g_\gamma\left(u_n \cdot z_0 \frac{\mathfrak{a}}{c_1}\right) \mathrm{d}z_0 \\
&\lesssim\; u_n \int_{s_1}^{s_2} \mathbb{1}\{z_0 \in [s_1, s_2]\} g_\gamma(u_n \cdot z_0) \,\mathrm{d}z_0\,,
\end{aligned}$$

by the symmetry of g_γ. The multiplicative constant in $\lesssim$ depends only on γ, $\mathfrak{a}$, and the coefficient c_1. We bound this in two different ways. First,

$$\begin{aligned}
&u_n \int_{s_1}^{s_2} g_\gamma(u_n \cdot z_0) \,\mathrm{d}z_0 \;\leq\; u_n \cdot g_\gamma(u_n \cdot s_1) \cdot (s_2 - s_1) \;=\; \frac{u_n}{(1 + s_1 \cdot u_n)^{1+\gamma}} (s_2 - s_1) \\
&\lesssim\; u_n^{-\gamma}(s_2 - s_1)\,,
\end{aligned}$$

since $s_1 > \underline{s} > 0$. Here the constant in $\lesssim$ depends only on c_1, $\mathfrak{a}$, and $\underline{s}$. The second bound is

$$u_n \int_{s_1}^{s_2} g_\gamma(u_n \cdot z_0) \,\mathrm{d}z_0 \;\lesssim\; u_n \int_{\underline{s}}^{\infty} g_\gamma(u_n \cdot z_0) \,\mathrm{d}z_0 \;\lesssim\; (1 + u_n \cdot \underline{s})^{-\gamma} \;\lesssim\; u_n^{-\gamma}\,,$$

where the multiplicative constant in $\lesssim$ depends only on $\underline{s}$. This proves the case with $|I| = 1$. We proceed by mathematical induction: suppose that there exists $j \in \{1, \ldots, T\}$ such that the statement holds for all $I \subset \{1, \ldots, T\}$ with $|I| = j$, that is,

$$\int_{\mathbb{R}^{|I|}} \mathbb{1}\left\{\sum_{m\in I} c_m z_{m-1} \in u_n \cdot \mathfrak{a} \cdot [s_1, s_2]\right\} \cdot \prod_{m\in I} g_\gamma\left(z_{m-1}\right) \,\mathrm{d}\mathbf{z} \;\lesssim\; u_n^{-\gamma} \cdot ((s_2 - s_1) \wedge 1) \tag{22}$$

for $I \subset \{1, \ldots, T\}, |I| = j$, where the multiplicative constant in $\lesssim$ depends only on T, γ, $\mathfrak{a}$, $\underline{s}$, and the coefficients (c_m). Let now $I \subset \{1, \ldots, T\}$ with $|I| = j + 1$. If $c_m = 0$ for some $m \in I$, then this coordinate is not active in the indicator, so that flushing that inactive variable out of the integral using $\|g_\gamma\|_{L^1(\mathbb{R})} < \infty$, we have only j active variables left, so (22) applies. Therefore, we assume that $c_m \neq 0$ for all $m \in I$. Write

$$\begin{aligned}
&\left\{\sum_{m\in I} c_m z_{m-1} \in u_n \cdot \mathfrak{a} \cdot [s_1, s_2]\right\} \\
&\subset \bigcup_{\mathfrak{m}\in I} \left\{c_{\mathfrak{m}} z_{\mathfrak{m}-1} \;\geq\; \frac{u_n \cdot \mathfrak{a} \cdot s_1}{j+1}\right\} \cap \bigcup_{\mathfrak{n}\in I} \left\{c_{\mathfrak{n}} z_{\mathfrak{n}-1} \;\leq\; \frac{u_n \cdot \mathfrak{a} \cdot s_2}{j+1}\right\} \\
&= \bigcup_{\substack{\mathfrak{m},\mathfrak{n}\in I \\ \mathfrak{m}\neq\mathfrak{n}}} \left\{c_{\mathfrak{m}} z_{\mathfrak{m}-1} \;\geq\; \frac{u_n \cdot \mathfrak{a} \cdot s_1}{j+1} \quad \text{and} \quad c_{\mathfrak{n}} z_{\mathfrak{n}-1} \;\leq\; \frac{u_n \cdot \mathfrak{a} \cdot s_2}{j+1}\right\} \setminus \left(\bigcup_{\mathfrak{m}\in I} \left\{c_{\mathfrak{m}} z_{\mathfrak{m}-1} \in \frac{u_n \cdot \mathfrak{a}}{j+1} \cdot [s_1, s_2]\right\}\right) \\
&\quad \cup \bigcup_{\mathfrak{m}\in I} \left\{c_{\mathfrak{m}} z_{\mathfrak{m}-1} \in \frac{u_n \cdot \mathfrak{a}}{j+1} \cdot [s_1, s_2]\right\}.
\end{aligned}$$

Since

$$\left(\bigcup_{\mathfrak{m}\in I}\left\{c_{\mathfrak{m}}z_{\mathfrak{m}-1}\in\frac{u_n\cdot\mathfrak{a}}{j+1}\cdot[s_1,s_2]\right\}\right)^c = \bigcap_{\mathfrak{m}\in I}\left\{c_{\mathfrak{m}}z_{\mathfrak{m}-1} > \frac{u_n\cdot\mathfrak{a}}{j+1}s_2 \quad \text{or} \quad c_{\mathfrak{m}}z_{\mathfrak{m}-1} < \frac{u_n\cdot\mathfrak{a}}{j+1}s_1\right\},$$

it holds on the intersection with this event that

$$\begin{aligned}&\bigcup_{\substack{\mathfrak{m},\mathfrak{n}\in I\\ \mathfrak{m}\neq\mathfrak{n}}}\left\{c_{\mathfrak{m}}z_{\mathfrak{m}-1} \geq \frac{u_n\cdot\mathfrak{a}\cdot s_1}{j+1} \quad \text{and} \quad c_{\mathfrak{n}}z_{\mathfrak{n}-1} \leq \frac{u_n\cdot\mathfrak{a}\cdot s_2}{j+1}\right\}\\ &= \bigcup_{\substack{\mathfrak{m},\mathfrak{n}\in I\\ \mathfrak{m}\neq\mathfrak{n}}}\left\{c_{\mathfrak{m}}z_{\mathfrak{m}-1} > \frac{u_n\cdot\mathfrak{a}\cdot s_2}{j+1} \quad \text{and} \quad c_{\mathfrak{n}}z_{\mathfrak{n}-1} < \frac{u_n\cdot\mathfrak{a}\cdot s_1}{j+1}\right\}.\end{aligned}$$

It follows that

$$\begin{aligned}&\left\{\sum_{m\in I}c_m z_{m-1}\in u_n\cdot\mathfrak{a}\cdot[s_1,s_2]\right\}\\ &\subset \bigcup_{\substack{\mathfrak{m},\mathfrak{n}\in I\\ \mathfrak{m}\neq\mathfrak{n}}}\left\{c_{\mathfrak{m}}z_{\mathfrak{m}-1} > \frac{u_n\cdot\mathfrak{a}\cdot s_2}{j+1} \quad \text{and} \quad c_{\mathfrak{n}}z_{\mathfrak{n}-1} < \frac{u_n\cdot\mathfrak{a}\cdot s_1}{j+1}\right\}\\ &\quad \cup \bigcup_{\mathfrak{m}\in I}\left\{c_{\mathfrak{m}}z_{\mathfrak{m}-1}\in\frac{u_n\cdot\mathfrak{a}}{j+1}\cdot[s_1,s_2]\right\}.\end{aligned}$$

Then

$$\begin{aligned}&\int_{\mathbb{R}^{|I|}}\mathbb{1}\left\{\sum_{m\in I}c_m z_{m-1}\in u_n\cdot\mathfrak{a}\cdot[s_1,s_2]\right\}\cdot\prod_{m\in I}g_\gamma\left(z_{m-1}\right)\,\mathrm{d}\mathbf{z}\\ &\leq \sum_{\mathfrak{m}\in I}\int_{\mathbb{R}^{|I|}}\mathbb{1}\left\{c_{\mathfrak{m}}z_{\mathfrak{m}-1}\in\frac{u_n\cdot\mathfrak{a}}{j+1}\cdot[s_1,s_2]\right\}\cdot\prod_{m\in I}g_\gamma\left(z_{m-1}\right)\,\mathrm{d}\mathbf{z}\\ &+\sum_{\substack{\mathfrak{m},\mathfrak{n}\in I\\ \mathfrak{m}\neq\mathfrak{n}}}\int_{\mathbb{R}^{|I|}}\mathbb{1}\left\{c_{\mathfrak{m}}z_{\mathfrak{m}-1} > \frac{u_n\cdot\mathfrak{a}\cdot s_2}{j+1} \quad \text{and} \quad c_{\mathfrak{n}}z_{\mathfrak{n}-1} < \frac{u_n\cdot\mathfrak{a}\cdot s_1}{j+1}\right\}\\ &\qquad\times\mathbb{1}\left\{\sum_{m\in I}c_m z_{m-1}\in u_n\cdot\mathfrak{a}\cdot[s_1,s_2]\right\}\cdot\prod_{m\in I}g_\gamma\left(z_{m-1}\right)\,\mathrm{d}\mathbf{z}\end{aligned}$$

and we control each term. On the diagonal $\mathfrak{m}=\mathfrak{n}$, the bound follows immediately by (22) for $j=1$ by setting

$$\widetilde{s}_1 = \frac{s_1}{j+1} \quad \text{and} \quad \widetilde{s}_2 = \frac{s_2}{j+1}, \qquad \text{so that} \qquad \left(\left(\widetilde{s}_2-\widetilde{s}_1\right)\wedge 1\right)\leq\left(\left(s_2-s_1\right)\wedge 1\right),$$

and integrating out all other variables. Next we lay out how we tackle the off-diagonal part. If $s_2=\infty$, there is nothing to do. Therefore, assume that $s_2<\infty$. For the part off the diagonal we draw the smaller variable into the g_γ of the larger variable. Then we use monotonicity on the half-line of g_γ to fix the smaller variable to a constant. We then transform back and distribute the constant to the interval $[s_1,s_2]$. Finally, we use the assumption of the mathematical induction. Let us execute this. Let $\mathfrak{m},\mathfrak{n}\in I$ satisfying $\mathfrak{m}\neq\mathfrak{n}$ such that

$$c_{\mathfrak{m}}z_{\mathfrak{m}-1} > \frac{u_n\cdot\mathfrak{a}\cdot s_2}{j+1} \qquad \text{and} \qquad c_{\mathfrak{n}}z_{\mathfrak{n}-1} < \frac{u_n\cdot\mathfrak{a}\cdot s_1}{j+1}.$$

Then

$$c_{\mathfrak{m}} z_{\mathfrak{m}-1} \; - \; c_{\mathfrak{n}} z_{\mathfrak{n}-1} \; > \; c_{\mathfrak{m}} z_{\mathfrak{m}-1} \; - \; \frac{u_n \cdot \mathfrak{a} \cdot s_1}{j+1} \; > \; \frac{u_n \cdot \mathfrak{a}}{j+1} (s_2 - s_1) \; \geq \; 0 \, ,$$

so that by the monotonicity of g_γ on $[0, \infty)$,

$$g_\gamma \left(c_{\mathfrak{m}} z_{\mathfrak{m}-1} \; - \; c_{\mathfrak{n}} z_{\mathfrak{n}-1} \right) \; \leq \; g_\gamma \left(c_{\mathfrak{m}} z_{\mathfrak{m}-1} \; - \; \frac{u_n \cdot \mathfrak{a} \cdot s_1}{j+1} \right) .$$

By

$$g_\gamma(z_{\mathfrak{m}-1}) \; \lesssim \; g_\gamma(c_{\mathfrak{m}} z_{\mathfrak{m}-1}) \; \lesssim \; g_\gamma(z_{\mathfrak{m}-1}) \, , \tag{23}$$

where the multiplicative constant in $\lesssim$ depends only on $c_{\mathfrak{m}}$ and γ. With the change of variables

$$z_{\mathfrak{m}-1} \; \mapsto \; z_{\mathfrak{m}-1} \; - \; \frac{c_{\mathfrak{n}}}{c_{\mathfrak{m}}} z_{\mathfrak{n}-1} \, ,$$

we get

$$\begin{aligned}
&\int_{\mathbb{R}^{|I|}} \mathbb{1} \left\{ c_{\mathfrak{m}} z_{\mathfrak{m}-1} \; > \; \frac{u_n \cdot \mathfrak{a} \cdot s_2}{j+1} \qquad \text{and} \qquad c_{\mathfrak{n}} z_{\mathfrak{n}-1} < \frac{u_n \cdot \mathfrak{a} \cdot s_1}{j+1} \right\} \\
&\qquad \times \; \mathbb{1} \left\{ \sum_{m \in I} c_m z_{m-1} \in u_n \cdot \mathfrak{a} \cdot [s_1, s_2] \right\} \cdot \prod_{m \in I} g_\gamma \left(z_{m-1} \right) \, \mathrm{d}\mathbf{z} \\
&\lesssim \; \int_{\mathbb{R}^{|I|}} \mathbb{1} \left\{ c_{\mathfrak{m}} z_{\mathfrak{m}-1} \; > \; \frac{u_n \cdot \mathfrak{a} \cdot s_2}{j+1} \qquad \text{and} \qquad c_{\mathfrak{n}} z_{\mathfrak{n}-1} < \frac{u_n \cdot \mathfrak{a} \cdot s_1}{j+1} \right\} \\
&\qquad \times \; \mathbb{1} \left\{ \sum_{m \in I} c_m z_{m-1} \in u_n \cdot \mathfrak{a} \cdot [s_1, s_2] \right\} g_\gamma(c_{\mathfrak{m}} z_{\mathfrak{m}-1}) \cdot \prod_{\substack{m \in I \\ m \neq \mathfrak{m}}} g_\gamma \left(z_{m-1} \right) \, \mathrm{d}\mathbf{z} \\
&= \; \int_{\mathbb{R}^{|I|}} \mathbb{1} \left\{ c_{\mathfrak{m}} z_{\mathfrak{m}-1} \; > \; \frac{u_n \cdot \mathfrak{a} \cdot s_2}{j+1} \qquad \text{and} \qquad c_{\mathfrak{n}} z_{\mathfrak{n}-1} < \frac{u_n \cdot \mathfrak{a} \cdot s_1}{j+1} \right\} \\
&\qquad \times \; \mathbb{1} \left\{ \sum_{\substack{m \in I \\ m \neq \mathfrak{n}}} c_m z_{m-1} \in u_n \cdot \mathfrak{a} \cdot [s_1, s_2] \right\} g_\gamma \left(c_{\mathfrak{m}} z_{\mathfrak{m}-1} \; - \; c_{\mathfrak{n}} z_{\mathfrak{n}-1} \right) \cdot \prod_{\substack{m \in I \\ m \neq \mathfrak{m}}} g_\gamma \left(z_{m-1} \right) \, \mathrm{d}\mathbf{z} \\
&\leq \; \int_{\mathbb{R}^{|I|}} \mathbb{1} \left\{ \sum_{\substack{m \in I \\ m \neq \mathfrak{n}}} c_m z_{m-1} \in u_n \cdot \mathfrak{a} \cdot [s_1, s_2] \right\} g_\gamma \left(c_{\mathfrak{m}} z_{\mathfrak{m}-1} \; - \; \frac{u_n \cdot \mathfrak{a}}{j+1} s_1 \right) \cdot \prod_{\substack{m \in I \\ m \neq \mathfrak{m}}} g_\gamma \left(z_{m-1} \right) \, \mathrm{d}\mathbf{z} \, ,
\end{aligned}$$

where the multiplicative constant in $\lesssim$ depends only on $c_{\mathfrak{m}}$ and γ. By the change of variables

$$z_{\mathfrak{m}-1} \; \mapsto \; z_{\mathfrak{m}-1} \; + \; \frac{1}{c_{\mathfrak{m}}} \frac{u_n \cdot \mathfrak{a}}{j+1} \cdot s_1$$

and Lemma A.2(iii) again, this equals

$$\begin{aligned}
&\int_{\mathbb{R}^{|I|}} \mathbb{1}\left\{\sum_{\substack{m\in I\\ m\neq \mathfrak{n}}} c_m z_{m-1} \in u_n\cdot\mathfrak{a}\cdot\left[s_1-\frac{s_1}{j+1}\,,s_2-\frac{s_1}{j+1}\right]\right\} g_\gamma(c_{\mathfrak{m}} z_{\mathfrak{m}-1}) \prod_{\substack{m\in I\\ m\neq \mathfrak{m}}} g_\gamma\left(z_{m-1}\right)\,\mathrm{d}\mathbf{z}\\
\lesssim\ &\int_{\mathbb{R}^{|I|}} \mathbb{1}\left\{\sum_{\substack{m\in I\\ m\neq \mathfrak{n}}} c_m z_{m-1} \in u_n\cdot\mathfrak{a}\cdot\left[s_1-\frac{s_1}{j+1}\,,s_2-\frac{s_1}{j+1}\right]\right\} \prod_{m\in I} g_\gamma\left(z_{m-1}\right)\,\mathrm{d}\mathbf{z}\\
\lesssim\ &\|g_\gamma\|_{L^1(\mathbb{R})}\int_{\mathbb{R}^{|I|-1}} \mathbb{1}\left\{\sum_{\substack{m\in I\\ m\neq \mathfrak{n}}} c_m z_{m-1} \in u_n\cdot\mathfrak{a}\cdot\left[s_1-\frac{s_1}{j+1}\,,s_2-\frac{s_1}{j+1}\right]\right\} \prod_{\substack{m\in I\\ m\neq \mathfrak{n}}} g_\gamma\left(z_{m-1}\right)\,\mathrm{d}\mathbf{z}\\
\lesssim\ &u_n^{-\gamma}\left(\left(s_2-\frac{s_1}{j+1}\ -\ \left(s_1-\frac{s_1}{j+1}\right)\right)\wedge 1\right)\\
=\ &u_n^{-\gamma}\cdot\left((s_2-s_1)\wedge 1\right),
\end{aligned}$$

where the second to last inequality follows from (23), and the last inequality follows from $|I\setminus\{\mathfrak{n}\}| = j$ and (22) with

$$\widetilde{s}_1 \ =\ s_1\ -\ \frac{s_1}{j+1} \qquad\text{and}\qquad \widetilde{s}_2\ =\ s_2\ -\ \frac{s_1}{j+1}\,.$$

The multiplicative constant in $\lesssim$ depends only on T, (c_m), $\mathfrak{a}$, γ, and $\underline{s}$. Applying mathematical induction finishes the proof. $\square$

Next we provide the variants of the bounds needed for the subsequent analysis.

Lemma C.2 (Separable product bound II). *Let $\emptyset \neq I\subset\{1,\ldots,T\}$, $(c_m)_{m\in I}\subset\mathbb{R}$, $\mathfrak{a}\in(0,\infty)$, $\gamma\in(0,\alpha-1)$, and let $\varphi,\psi:\{1,\ldots,T\}\to\mathbb{R}$ and $\mathfrak{x}\in\mathbb{R}$. Then*

(i)

$$\begin{aligned}
&(1\wedge|\mathfrak{x}|)\int_{\mathbb{R}^{|I|}}\mathbb{1}\left\{\left|\sum_{m\in I}c_m z_{m-1}\right|\in u_n\cdot\mathfrak{a}\cdot[s_1,s_2]\right\}\cdot\prod_{m\in I} g_\gamma\left(z_{m-1}+\varphi(m)+\psi(m)\right)\,\mathrm{d}\mathbf{z}\\
&\lesssim\ u_n^{-\gamma}\cdot\left((s_2-s_1)\wedge 1\right)\cdot\left(1+\left|\sum_{m\in I}c_m\varphi(m)\right|^{1+\gamma}\right)\cdot\left(|\mathfrak{x}|+\left|\sum_{m\in I}c_m\psi(m)\right|^{1+\gamma}\right),
\end{aligned}$$

(ii)

$$\begin{aligned}
&\int_{\mathfrak{b}\wedge\mathfrak{c}}^{\mathfrak{b}\vee\mathfrak{c}}(1\wedge|s\cdot\mathfrak{x}|)\left(\int_{\mathbb{R}^{|I|}}\mathbb{1}\left\{\left|\sum_{m\in I}c_m z_{m-1}\right|\in u_n\cdot\mathfrak{a}\cdot[s_1,s_2]\right\}\cdot\prod_{m\in I} g_\gamma\left(z_{m-1}+\varphi(m)+s\cdot\psi(m)\right)\,\mathrm{d}\mathbf{z}\right)\mathrm{d}s\\
&\lesssim\ u_n^{-\gamma}\cdot\left((s_2-s_1)\wedge 1\right)\\
&\quad\times\ \left(\left|\sum_{m\in I}c_m\psi(m)\right|^{\gamma}\ +\ |\mathfrak{x}|^\gamma\right)\ \cdot\ \left(1+\left|\sum_{m\in I}c_m\varphi(m)\right|^{1+\gamma}\right)\cdot\left(|\mathfrak{b}|^{1+\gamma}+|\mathfrak{c}|^{1+\gamma}\right)
\end{aligned}$$

where $\lesssim$ means inequality up to a multiplicative constant that depends only on T, (c_m), $\mathfrak{a}$, γ, and $\underline{s}$, and is independent of φ, ψ, $\mathfrak{x}$, n, s_1 and s_2.

Proof. If $c_m = 0$ for all $m \in I$,

$$\mathbb{1}\left\{\left|\sum_{m\in I} c_m z_{m-1}\right| \in u_n \cdot \mathfrak{a} \cdot [s_1, s_2]\right\} \;=\; 0\,,$$

so there is nothing to show. Therefore we assume $|\{m \in I \mid c_m \neq 0\}| > 0$. Since

$$\sum_{m\in I} c_m z_{m-1} \;=\; \sum_{\substack{m\in I\\ c_m\neq 0}} c_m z_{m-1}\,,$$

and

$$\int_{\mathbb{R}^{|\{m\in I|c_m=0\}|}} \prod_{\substack{m\in I\\ c_m=0}} g_\gamma(z_{m-1} + \varphi(m))\,\mathrm{d}\mathbf{z} \;=\; \|g_\gamma\|_{L^1(\mathbb{R})}^{|\{m\in I|c_m=0\}|} \;<\infty\,,$$

we may assume without loss of generality that $c_m \neq 0$ for all $m \in I$.

Proof of (i): For all $m \in I$ we make the change of variables

$$z_{m-1} \;\mapsto\; z_{m-1} - (\varphi(m) + \psi(m))$$

to get

$$\begin{aligned}
&\int_{\mathbb{R}^{|I|}} \mathbb{1}\left\{\left|\sum_{m\in I} c_m z_{m-1}\right| \in u_n \cdot \mathfrak{a} \cdot [s_1, s_2]\right\} \cdot \prod_{m\in I} g_\gamma\left(z_{m-1} + \varphi(m) + \psi(m)\right)\,\mathrm{d}\mathbf{z}\\
&= \int_{\mathbb{R}^{|I|}} \mathbb{1}\left\{\left|\sum_{m\in I} c_m\left(z_{m-1} - \varphi(m) - \psi(m)\right)\right| \in u_n \cdot \mathfrak{a} \cdot [s_1, s_2]\right\} \cdot \prod_{m\in I} g_\gamma\left(z_{m-1}\right)\,\mathrm{d}\mathbf{z}\,.
\end{aligned}$$

Next we invoke Lemma A.2(i) to separate the $\mathbf{z}$ variable from φ and ψ in g_γ. We cannot do this separately for each $m \in I$, because this would yield a product that is unsuitable for the subsequent analysis. We therefore fix $\mathfrak{m} \in I$, shift all the relevant terms to the corresponding g_γ, and invoke Lemma A.2(i) once. Making the change of variables

$$z_{\mathfrak{m}-1} \;\mapsto\; z_{\mathfrak{m}-1} \;+\; \frac{1}{c_{\mathfrak{m}}}\sum_{m\in I} c_m\left(\varphi(m) + \psi(m)\right)$$

this equals

$$\int_{\mathbb{R}^{|I|}} \mathbb{1}\left\{\left|\sum_{m\in I} c_m z_{m-1}\right| \in u_n \cdot \mathfrak{a} \cdot [s_1, s_2]\right\} \cdot \prod_{\substack{m\in I\\ m\neq\mathfrak{m}}} g_\gamma\left(z_{m-1}\right) \cdot g_\gamma\left(z_{\mathfrak{m}-1} \;+\; \frac{1}{c_{\mathfrak{m}}}\sum_{m\in I} c_m\left(\varphi(m) \;+\; \psi(m)\right)\right)\,\mathrm{d}\mathbf{z}\,.$$

Applying Lemma A.2(i) twice,

$$\begin{aligned}
&g_\gamma\left(z_{\mathfrak{m}-1} \;+\; \frac{1}{c_{\mathfrak{m}}}\sum_{m\in I} c_m\varphi(m) \;+\; \frac{1}{c_{\mathfrak{m}}}\sum_{m\in I} c_m\psi(m)\right)\\
&\lesssim\; g_\gamma(z_{\mathfrak{m}-1})\left(1 \;+\; \left|\frac{1}{c_{\mathfrak{m}}}\sum_{m\in I} c_m\varphi(m)\right|^{1+\gamma}\right)\left(1 \;+\; \left|\frac{1}{c_{\mathfrak{m}}}\sum_{m\in I} c_m\psi(m)\right|^{1+\gamma}\right)\\
&\lesssim\; g_\gamma(z_{\mathfrak{m}-1})\left(1 \;+\; \left|\sum_{m\in I} c_m\varphi(m)\right|^{1+\gamma}\right)\left(1 \;+\; \left|\sum_{m\in I} c_m\psi(m)\right|^{1+\gamma}\right)\,,
\end{aligned}$$

where the multiplicative constant in $\lesssim$ depends only on γ. The inequality

$$(1 \wedge |\mathfrak{x}|) \cdot \left(1 \;+\; \left|\sum_{m\in I} c_m \psi(m)\right|^{1+\gamma}\right) \;\leq\; \left(|\mathfrak{x}| \;+\; \left|\sum_{m\in I} c_m \psi(m)\right|^{1+\gamma}\right),$$

together with Lemma C.1 yields the claimed inequality.

Proof of (ii): For all $m \in I$ we make the change of variables

$$z_{m-1} \;\mapsto\; z_{m-1} - (\varphi(m) + s \cdot \psi(m))$$

to get

$$\begin{aligned}&\int_{\mathbb{R}^{|I|}} \mathbb{1}\left\{\left|\sum_{m\in I} c_m z_{m-1}\right| \in u_n \cdot \mathfrak{a} \cdot [s_1, s_2]\right\} \cdot \prod_{m\in I} g_\gamma\left(z_{m-1} + \varphi(m) + s \cdot \psi(m)\right) \mathrm{d}\mathbf{z}\\ &= \int_{\mathbb{R}^{|I|}} \mathbb{1}\left\{\left|\sum_{m\in I} c_m \left(z_{m-1} - \varphi(m) - s \cdot \psi(m)\right)\right| \in u_n \cdot \mathfrak{a} \cdot [s_1, s_2]\right\} \cdot \prod_{m\in I} g_\gamma\left(z_{m-1}\right) \mathrm{d}\mathbf{z}\,.\end{aligned}$$

Fix $\mathfrak{m} \in I$. Making the change of variables

$$z_{\mathfrak{m}-1} \;\mapsto\; z_{\mathfrak{m}-1} \;+\; \frac{1}{c_\mathfrak{m}} \sum_{m\in I} c_m \left(\varphi(m) + s \cdot \psi(m)\right)$$

this equals

$$\int_{\mathbb{R}^{|I|}} \mathbb{1}\left\{\left|\sum_{m\in I} c_m z_{m-1}\right| \in u_n \cdot \mathfrak{a} \cdot [s_1, s_2]\right\} \cdot \prod_{\substack{m\in I\\ m\neq\mathfrak{m}}} g_\gamma\left(z_{m-1}\right) \cdot g_\gamma\left(z_{\mathfrak{m}-1} \;+\; \frac{1}{c_\mathfrak{m}} \sum_{m\in I} c_m \left(\varphi(m) \;+\; s\psi(m)\right)\right) \mathrm{d}\mathbf{z}\,.$$

By Lemma A.2(iv), with

$$\mathfrak{x} = \mathfrak{x}\,, \qquad \mathfrak{y} \;=\; \frac{1}{c_\mathfrak{m}} \sum_{m\in I} c_m \psi(m)\,, \qquad \mathfrak{z} \;=\; z_{\mathfrak{m}-1} \;+\; \frac{1}{c_\mathfrak{m}} \sum_{m\in I} c_m \varphi(m)\,,$$

and then Lemma A.2(i),

$$\begin{aligned}&\int_{\mathfrak{b}\wedge\mathfrak{c}}^{\mathfrak{b}\vee\mathfrak{c}} (1 \wedge |s \cdot \mathfrak{x}|) \cdot g_\gamma\left(z_{\mathfrak{m}-1} \;+\; \frac{1}{c_\mathfrak{m}} \sum_{m\in I} c_m \left(\varphi(m) + s \cdot \psi(m)\right)\right) \mathrm{d}s\\ &\lesssim \left(\left|\sum_{m\in I} c_m \psi(m)\right|^{\gamma} \;+\; |\mathfrak{x}|^\gamma\right) \cdot \left(1 + \left|\sum_{m\in I} c_m \varphi(m)\right|^{1+\gamma}\right) \cdot \left(|\mathfrak{b}|^{1+\gamma} + |\mathfrak{c}|^{1+\gamma}\right) \cdot g_\gamma(z_{\mathfrak{m}-1})\,,\end{aligned}$$

where the multiplicative constant in $\lesssim$ depends only on γ. Applying Lemma C.1 yields the result. □

The next lemma flushes the inactive variables out of the analysis, while keeping enough degrees of freedom for the subsequent analysis.

Lemma C.3 (Flushing inactive variables)**.** *Let $\gamma \in (0, \infty)$ and $\mathfrak{y}_1, \ldots, \mathfrak{y}_T \in \mathbb{R}$. Then*

$$\begin{aligned}&\int_{\mathbb{R}^T} G_n(\mathbf{A}_T \cdot \mathbf{z}, s_1, s_2, A) \prod_{m=1}^{T} g_\gamma(z_{m-1} + \mathfrak{y}_m)\, \mathrm{d}\mathbf{z}\\ &\lesssim \sum_{j=1}^{T} \int_{\mathbb{R}^j} \mathbb{1}\left\{\left|\sum_{m=1}^{j} a_{j-m} z_{m-1}\right| \in u_n \cdot \mathfrak{a}_j \cdot [s_1, s_2]\right\} \prod_{m=1}^{j} g_\gamma(z_{m-1} + \mathfrak{y}_m)\, \mathrm{d}\mathbf{z}\,,\end{aligned}$$

where $\lesssim$ means inequality up to a multiplicative constant that depends only on T and γ.

Proof. By (10),

$$G_n(\mathbf{A}_T \cdot \mathbf{z}, s_1, s_2, A) \;\le\; \sum_{j=1}^{T} \mathbb{1}\left\{ \left| \sum_{m=1}^{j} a_{j-m} z_{m-1} \right| \in u_n \cdot \mathfrak{a}_j \cdot [s_1, s_2] \right\},$$

so that

$$\begin{aligned}
&\int_{\mathbb{R}^T} G_n(\mathbf{A}_T \cdot \mathbf{z}, s_1, s_2, A) \prod_{m=1}^{T} g_\gamma(z_{m-1} + \mathfrak{y}_m)\, \mathrm{d}\mathbf{z} \\
&\le \sum_{j=1}^{T} \left(\int_{\mathbb{R}^j} \mathbb{1}\left\{ \left| \sum_{m=1}^{j} a_{j-m} z_{m-1} \right| \in u_n \cdot \mathfrak{a}_j \cdot [s_1, s_2] \right\} \prod_{m=1}^{j} g_\gamma(z_{m-1} + \mathfrak{y}_m)\, \mathrm{d}\mathbf{z} \right. \\
&\quad \left. \times \int_{\mathbb{R}^{T-j}} \prod_{m=j+1}^{T} g_\gamma(z_{m-1} + \mathfrak{y}_m)\, \mathrm{d}\mathbf{z} \right).
\end{aligned}$$

For $m > j$, we make the changes of variables

$$z_{m-1} \;\mapsto\; z_{m-1} - \mathfrak{y}_m,$$

so that

$$\int_{\mathbb{R}^{T-j}} \prod_{m=j+1}^{T} g_\gamma(z_{m-1} + \mathfrak{y}_m)\, \mathrm{d}\mathbf{z} \;=\; \left(\|g_\gamma\|_{L^1(\mathbb{R})} \right)^{T-j} \;<\; \infty.$$

This gives the claimed upper bound. □

We finish the section with a lemma that bounds two single probabilistic terms which will appear in the subsequent analysis. Since the bounds developed above are analytic in nature, the lemma also illustrates how they apply to the probabilistic problem at hand.

Lemma C.4 (Separable product bound III). *Let Assumptions 1 and 2 hold, let $T \ge 1$, let $\gamma \in (0, \alpha - 1)$. Then, for $n \in \mathbb{N}$, we have the following upper bounds:*

(i)

$$\|\nabla G_{\infty,n}(\mathbf{0}_T, s_1, s_2, A)\|_\infty \;\lesssim\; u_n^{-\gamma} \cdot ((s_2 - s_1) \wedge 1),$$

(ii)

$$\mathbf{P}[[X_\ell]_\ell \in u_n \cdot ((s_1 \cdot A) \setminus (s_2 \cdot A))] \;\lesssim\; u_n^{-\gamma} \cdot ((s_2 - s_1) \wedge 1).$$

In both (i) and (ii) $\lesssim$ means inequality up to a multiplicative constant that only depends on A, T, $\underline{s}$, γ, the coefficients (a_i), and the distribution of ε, and is independent of s_1, s_2, and n.

Proof of (i): By Lemma B.3,

$$\begin{aligned}
\|\nabla G_{\infty,n}(\mathbf{0}_T, s_1, s_2, A)\|_\infty \;&\le\; \int_{\mathbb{R}^T} G_n(\mathbf{z}, s_1, s_2, A) \left\| \nabla f_{[X_\ell]_\ell}(\mathbf{z}) \right\|_\infty \mathrm{d}\mathbf{z} \\
&=\; \int_{\mathbb{R}^T} G_n(\mathbf{A}_T \cdot \mathbf{z}, s_1, s_2, A) \left\| \nabla f_{[X_\ell]_\ell}(\mathbf{A}_T \cdot \mathbf{z}) \right\|_\infty \mathrm{d}\mathbf{z},
\end{aligned}$$

where the last equality follows from the change of variables $\mathbf{z} \mapsto \mathbf{A}_T \cdot \mathbf{z}$, with $\det \mathbf{A}_T = 1$. By Lemma B.1 and Assumption 1,

$$\begin{aligned}\left\|\nabla f_{[X_\ell]_\ell}(\mathbf{A}_T \cdot \mathbf{z})\right\|_\infty &\lesssim \sum_{m=1}^{T} \mathbf{E}\left[\left|f_\varepsilon'(z_{m-1} - \mathbf{e}_m^\top \cdot \mathbf{A}_T^{-1} \cdot [X_\ell - X_{\ell,\ell}]_\ell)\right| \prod_{h\neq m} f_\varepsilon(z_{h-1} - \mathbf{e}_h^\top \cdot \mathbf{A}_T^{-1} \cdot [X_\ell - X_{\ell,\ell}]_\ell)\right] \\ &\lesssim \mathbf{E}\left[\prod_{m=1}^{T} g_\gamma(z_{m-1} - \mathbf{e}_m^\top \cdot \mathbf{A}_T^{-1} \cdot [X_\ell - X_{\ell,\ell}]_\ell)\right],\end{aligned}$$

where the multiplicative constant in $\lesssim$ depends only on T, the coefficients $(\mathfrak{a}_i)$, and the distribution of ε. By Lemma C.3,

$$\begin{aligned}&\int_{\mathbb{R}^T} G_n(\mathbf{A}_T \cdot \mathbf{z}, s_1, s_2, A) \prod_{m=1}^{T} g_\gamma(z_{m-1} - \mathbf{e}_m^\top \cdot \mathbf{A}_T^{-1} \cdot [X_\ell - X_{\ell,\ell}]_\ell)\,\mathrm{d}\mathbf{z} \\ &\lesssim \sum_{j=1}^{T} \int_{\mathbb{R}^j} \mathbb{1}\left\{\left|\sum_{m=1}^{j} a_{j-m} z_{m-1}\right| \in u_n \cdot \mathfrak{a}_j \cdot [s_1, s_2]\right\} \prod_{m=1}^{j} g_\gamma(z_{m-1} - \mathbf{e}_m^\top \cdot \mathbf{A}_T^{-1} \cdot [X_\ell - X_{\ell,\ell}]_\ell)\,\mathrm{d}\mathbf{z},\end{aligned}$$

Since

$$\sum_{m=1}^{j} a_{j-m} \mathbf{e}_m^\top \;=\; \mathbf{e}_j^\top \mathbf{A}_T$$

it holds

$$\sum_{m=1}^{j} a_{j-m} \cdot \mathbf{e}_m^\top \cdot \mathbf{A}_T^{-1} \cdot [X_\ell - X_{\ell,\ell}]_\ell \;=\; \mathbf{e}_j^\top \cdot [X_\ell - X_{\ell,\ell}]_\ell \;=\; X_{j-1} - X_{j-1,j-1}.$$

Applying Lemma C.2(i) with

$$\varphi(m) \;=\; -\,\mathbf{e}_m^\top \cdot \mathbf{A}_T^{-1} \cdot [X_\ell - X_{\ell,\ell}]_\ell, \qquad \psi \;\equiv\; 0, \qquad m \in \{1, \ldots, j\},$$

we get

$$\begin{aligned}&\sum_{j=1}^{T} \int_{\mathbb{R}^j} \mathbb{1}\left\{\left|\sum_{m=1}^{j} a_{j-m} z_{m-1}\right| \in u_n \cdot \mathfrak{a}_j \cdot [s_1, s_2]\right\} \prod_{m=1}^{j} g_\gamma(z_{m-1} - \mathbf{e}_m^\top \cdot \mathbf{A}_T^{-1} \cdot [X_\ell - X_{\ell,\ell}]_\ell)\,\mathrm{d}\mathbf{z} \\ &\lesssim u_n^{-\gamma} \cdot ((s_2 - s_1) \wedge 1) \cdot \sum_{j=1}^{T} \left(1 + |X_{j-1} - X_{j-1,j-1}|^{1+\gamma}\right).\end{aligned}$$

To obtain the claimed upper bound $u_n^{-\gamma}((s_2 - s_1) \wedge 1)$, note that $1 + \gamma < \alpha$, so that

$$\mathbf{E}\left[\sum_{j=1}^{T} \left(1 + |X_{j-1} - X_{j-1,j-1}|^{1+\gamma}\right)\right] \;=\; T\left(1 \;+\; \mathbf{E}[|X_0 - X_{0,0}|^{1+\gamma}]\right) \;<\; \infty.$$

Proof of (ii): It holds $u_n \cdot ((s_1 \cdot A) \setminus (s_2 \cdot A)) \subset \bigcup_{j=1}^{T} \{|x_{j-1}| \in \mathfrak{a}_j \cdot u_n \cdot [s_1, s_2]\}$, so that, by stationarity of (X_t),

$$\mathbf{P}[[X_\ell]_\ell \in u_n \cdot (s_1 \cdot A \setminus s_2 \cdot A)] \;\leq\; \sum_{j=1}^{T} \mathbf{P}[|X_0| \in \mathfrak{a}_j \cdot u_n \cdot [s_1, s_2]].$$

Then, for $j \in \{1, \ldots, T\}$,

$$\begin{aligned}
\mathbf{P}[|X_0| \in \mathfrak{a}_j \cdot u_n \cdot [s_1, s_2]] &= \int_{\mathbb{R}} \mathbb{1}\{|z_0| \in \mathfrak{a}_j \cdot u_n \cdot [s_1, s_2]\} \cdot f_{X_0}(z_0) \, \mathrm{d}z_0 \\
&= \int_{\mathbb{R}} \mathbb{1}\{|z_0| \in \mathfrak{a}_j \cdot u_n \cdot [s_1, s_2]\} \cdot \mathbf{E}[f_\varepsilon(z_0 - (X_0 - X_{0,0}))] \, \mathrm{d}z_0 \\
&\lesssim \mathbf{E}[1 \vee |X_0 - X_{0,0}|^{1+\gamma}] \int_{\mathbb{R}} \mathbb{1}\{|z_0| \in \mathfrak{a}_j \cdot u_n \cdot [s_1, s_2]\} \cdot g_\gamma(z_0) \, \mathrm{d}z_0 \\
&\lesssim u_n^{-\gamma} \cdot ((s_2 - s_1) \wedge 1) \,,
\end{aligned}$$

since $1 + \gamma < \alpha$, by Lemma A.2(i), and

$$\int_{\mathbb{R}} \mathbb{1}\{|z_0| \in \mathfrak{a}_j \cdot u_n \cdot [s_1, s_2]\} \cdot g_\gamma(z_0) \, \mathrm{d}z_0 \;\lesssim\; u_n^{-\gamma} \cdot ((s_2 - s_1) \wedge 1) \,,$$

by Lemma C.1, where the constant in $\lesssim$ depends only on A, $\underline{s}$, γ, the coefficients $(\mathfrak{a}_i)$, and the distribution of ε, and is independent of s_1, s_2, and n. $\square$

D. Analysis of gradient differences

In the subsequent analysis we have to control differences of gradients of the form

$$\nabla_{\mathbf{z}} G_{k,n}(\mathbf{z}, s_1, s_2, A)\Big|_{\mathbf{z}=\mathbf{x}+\mathbf{y}} - \nabla_{\mathbf{z}} G_{k,n}(\mathbf{z}, s_1, s_2, A)\Big|_{\mathbf{z}=\mathbf{x}} \quad \text{or} \quad \nabla_{\mathbf{x}} G_{\infty,n}(\mathbf{x}, s_1, s_2, A) - \nabla_{\mathbf{x}} G_{k,n}(\mathbf{x}, s_1, s_2, A) \,,$$

where $\nabla G_{k,n}$ is defined in Lemma B.3 for $k \in \mathbb{N} \cup \{\infty\}$. These contain differences of products that create algebraic complexity. To organize this, we introduce the $\widetilde{R}$-notation. The following lemma then explains how this notation will be used. Part (i) rewrites the relevant differences of products in terms of $\widetilde{R}$-terms. Part (ii) gives a bound for these terms that is useful in the broader analysis. The final two lemmas of this section express the gradient differences in $\widetilde{R}$-notation, and apply the bounds of the last section to yield useful bounds for the final analysis.

Definition D.1 ($\widetilde{R}$-notation)**.** We define, for functions $\varphi, \psi : \{1, \ldots, T\} \to \mathbb{R}$ and $i, l \in \{1, \ldots, T\}$,

$$\widetilde{R}_{i,l}(\varphi, \psi)$$
$$:= \begin{cases}
\left(\prod_{j<l} f_\varepsilon(\varphi(j))\right) \cdot (f_\varepsilon(\varphi(l)) - f_\varepsilon(\varphi(l) + \psi(l))) \cdot f'_\varepsilon(\varphi(i) + \psi(i)) \cdot \left(\prod_{\substack{j>l \\ j \neq i}} f_\varepsilon(\varphi(j) + \psi(j))\right) \,, & l < i \,, \\
\left(\prod_{j<i} f_\varepsilon(\varphi(j))\right) \cdot (f'_\varepsilon(\varphi(i)) - f'_\varepsilon(\varphi(i) + \psi(i))) \cdot \left(\prod_{j>i} f_\varepsilon(\varphi(j) + \psi(j))\right) \,, & l = i \,, \\
f'_\varepsilon(\varphi(i)) \cdot \left(\prod_{\substack{j<l \\ j \neq i}} f_\varepsilon(\varphi(j))\right) \cdot (f_\varepsilon(\varphi(l)) - f_\varepsilon(\varphi(l) + \psi(l))) \cdot \left(\prod_{j>l} f_\varepsilon(\varphi(j) + \psi(j))\right) \,, & l > i \,.
\end{cases}$$

Lemma D.2 (Properties of $\widetilde{R}_{i,l}$)**.** *Let Assumption 1 hold. Then, for all functions $\varphi, \psi : \{1, \ldots, T\} \to \mathbb{R}$, and all $i, l \in \{1, \ldots, T\}$, the following statements about $\widetilde{R}_{i,l}(\varphi, \psi)$ are true:*

(i)

$$\sum_{l=1}^{T} \widetilde{R}_{i,l}(\varphi, \psi) \;=\; f'_\varepsilon(\varphi(i)) \cdot \prod_{\substack{j=1 \\ j \neq i}}^{T} f_\varepsilon(\varphi(j)) \;-\; f'_\varepsilon(\varphi(i) + \psi(i)) \cdot \prod_{\substack{j=1 \\ j \neq i}}^{T} f_\varepsilon(\varphi(j) + \psi(j))$$

(ii) For all $\gamma \in (0, \alpha - 1)$,

$$\left|\widetilde{R}_{i,l}(\varphi,\psi)\right| \lesssim (1 \wedge |\psi(l)|) \cdot \prod_{m=1}^{T} \left(g_\gamma(\varphi(m)) + g_\gamma(\varphi(m)+\psi(m))\right)$$
$$= (1 \wedge |\psi(l)|) \sum_{I \subset \{1,\ldots,T\}} \left(\prod_{m \notin I} g_\gamma(\varphi(m))\right) \left(\prod_{m \in I} g_\gamma(\varphi(m)+\psi(m))\right) .$$

In (ii) $\lesssim$ *means inequality up to a multiplicative constant that only depends on the distribution of* ε *, and is independent of* T *,* γ *,* φ *,* ψ *,* i *and* ℓ *.*

Proof. Fix $i, l \in \{1, \ldots, T\}$.
Proof of (i): The statement follows from the definition of $\widetilde{R}_{i,l}$ and

$$c_i \cdot \prod_{\substack{j=1 \\ j \neq i}}^{T} a_j - d_i \cdot \prod_{\substack{j=1 \\ j \neq i}}^{T} b_j$$
$$= \sum_{l=1}^{i-1} \left(\prod_{j<l} a_j\right) (a_l - b_l) \cdot d_i \cdot \left(\prod_{\substack{j>l \\ j \neq i}} b_j\right)$$
$$+ \left(\prod_{j<i} a_j\right) (c_i - d_i) \left(\prod_{j>i} b_j\right)$$
$$+ \sum_{l=i+1}^{T} c_i \cdot \left(\prod_{\substack{j<l \\ j \neq i}} a_j\right) (a_l - b_l) \left(\prod_{\substack{j>l \\ j \neq i}} b_j\right) ,$$

with

$$a_j = f_\varepsilon(\varphi(j)) , \qquad b_j = f_\varepsilon(\varphi(j)+\psi(j)) , \qquad c_i = f'_\varepsilon(\varphi(i)) , \qquad d_i = f'_\varepsilon(\varphi(i)+\psi(i)) .$$

Proof of (ii): Expanding the product

$$\prod_{m=1}^{T} (a_m + b_m) = \sum_{I \subset \{1,\ldots,T\}} \prod_{m \notin I} a_m \prod_{m \in I} b_m , \tag{24}$$

with

$$a_m = g_\gamma(\varphi(m)) , \qquad b_m = g_\gamma(\varphi(m)+\psi(m)) , \qquad m \in \{1, \ldots, T\} ,$$

it remains to prove that the left-hand side of (24) is a valid upper bound for $|\widetilde{R}_{i,l}(\varphi,\psi)|$.
Case $|\psi(l)| < 1$ *:* In the product structure of $\widetilde{R}_{i,l}$ there appears exactly one factor that is a difference term, namely the l-th term. We begin by bounding the absolute value of this difference term. If this features f'_ε , we use Assumption 1, and if this features f_ε , we use Lemma A.1. The difference term is then bounded above by

$$|\psi(l)| \cdot g_\gamma(\varphi(l)) \lesssim (1 \wedge |\psi(l)|) \cdot \left(g_\gamma(\varphi(l)) + g_\gamma(\varphi(l)+\psi(l))\right) ,$$

where the multiplicative constant in $\lesssim$ depends only on the distribution of ε. By Assumption 1, any remaining factor $j \in \{1, \ldots, T\} \setminus \{l\}$ is bounded above by

$$g_\gamma(\varphi(j)) \; + \; g_\gamma(\varphi(j) + \psi(j)) \, .$$

up to a multiplicative constant that depends only on the distribution of ε.
Case $|\psi(l)| \geq 1$*:* We use the triangle inequality and Assumption 1 to bound the absolute value of the difference term above by

$$g_\gamma(\varphi(l)) \; + \; g_\gamma(\varphi(l) + \psi(l)) \, ,$$

up to a multiplicative constant that depends only on the distribution of ε. All remaining terms are bounded as in the previous case. Since $|\psi(l)| \geq 1$, it holds $(1 \wedge |\psi(l)|) = 1$, so we multiply this factor to the bound for free. □

The next lemma expresses gradient differences in terms of $\widetilde{R}$-notation.

Lemma D.3 (Gradient differences and $\widetilde{R}$-notation)**.** *Let Assumptions 1 and 2 hold. Then, for all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^T$ *and all* $k \in \mathbb{N}_0 \cup \{\infty\}$ *we have the following identities:*

(i)

$$\begin{aligned}
&\nabla_{\mathbf{z}} G_{k,n}(\mathbf{z}, s_1, s_2, A)\Big|_{\mathbf{z}=\mathbf{x}+\mathbf{y}} \; - \; \nabla_{\mathbf{z}} G_{k,n}(\mathbf{z}, s_1, s_2, A)\Big|_{\mathbf{z}=\mathbf{x}} \\
&= \left(\mathbf{A}_T^{-1}\right)^\top \\
&\quad \times \Bigg[\int_{\mathbb{R}^T} G_n(\mathbf{A}_T \cdot \mathbf{z}, s_1, s_2, A) \, \mathrm{d}\mathbf{z} \int_{\mathbb{R}^T} \mathrm{d}\mathbf{P}_{[X_{\ell,k+\ell} - X_{\ell,\ell}]_\ell}(\widetilde{\mathbf{x}}) \\
&\qquad \widetilde{R}_{i,l} \left(\left(j \;\mapsto\; z_{j-1} - \mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot (\mathbf{x} + \widetilde{\mathbf{x}}) \right), \left(j \;\mapsto\; -\mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot \mathbf{y} \right) \right) \Bigg]_{i,l \in \{1,\ldots,T\}} \cdot \mathbf{1}_T \, .
\end{aligned}$$

(ii)

$$\begin{aligned}
&\nabla_{\mathbf{x}} G_{\infty,n}(\mathbf{x}, s_1, s_2, A) \; - \; \nabla_{\mathbf{x}} G_{k,n}(\mathbf{x}, s_1, s_2, A) \\
&= \left(\mathbf{A}_T^{-1}\right)^\top \\
&\quad \times \Bigg[\int_{\mathbb{R}^T} G_n(\mathbf{A}_T \cdot \mathbf{z}, s_1, s_2, A) \, \mathrm{d}\mathbf{z} \int_{\mathbb{R}^T} \mathrm{d}\mathbf{P}_{[X_\ell - X_{\ell,k+\ell}]_\ell}(\widetilde{\mathbf{x}}) \int_{\mathbb{R}^T} \mathrm{d}\mathbf{P}_{[X_{\ell,k+\ell} - X_{\ell,\ell}]_\ell}(\widetilde{\mathbf{y}}) \\
&\qquad \widetilde{R}_{i,l} \left(\left(j \;\mapsto\; z_{j-1} \;-\; \mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot (\mathbf{x} \;+\; \widetilde{\mathbf{y}}) \right), \left(j \;\mapsto\; -\mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot \widetilde{\mathbf{x}} \right) \right) \Bigg]_{i,l \in \{1,\ldots,T\}} \cdot \mathbf{1}_T
\end{aligned}$$

Proof of (i): Note that for arbitrary $\mathbf{x} \in \mathbb{R}^T$ it holds that

$$\begin{aligned}
&\nabla_{\mathbf{x}} G_{k,n}(\mathbf{x}, s_1, s_2, A) \\
&= -\int_{\mathbb{R}^T} G_n(\mathbf{z}, s_1, s_2, A) \nabla f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{z} - \mathbf{x}) \, \mathrm{d}\mathbf{z} \\
&= -\sum_{i=1}^T \left(\mathbf{A}_T^{-1}\right)^\top \mathbf{e}_i \\
&\qquad \int_{\mathbb{R}^T} G_n(\mathbf{z}, s_1, s_2, A) \, \mathrm{d}\mathbf{z} \int_{\mathbb{R}^T} \mathrm{d}\mathbf{P}_{[X_{\ell,k+\ell} - X_{\ell,\ell}]_\ell}(\widetilde{\mathbf{x}}) \\
&\qquad f_\varepsilon' \left(\mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot (\mathbf{z} - \mathbf{x} - \widetilde{\mathbf{x}}) \right) \prod_{j \neq i} f_\varepsilon \left(\mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot (\mathbf{z} - \mathbf{x} - \widetilde{\mathbf{x}}) \right)
\end{aligned}$$

$$
\begin{aligned}
&= -\sum_{i=1}^{T} \left(\mathbf{A}_T^{-1}\right)^\top \mathbf{e}_i \\
&\qquad \int_{\mathbb{R}^T} G_n(\mathbf{A}\cdot\mathbf{z}, s_1, s_2, A)\,\mathrm{d}\mathbf{z} \int_{\mathbb{R}^T} \mathrm{d}\mathbf{P}_{[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell}(\widetilde{\mathbf{x}}) \\
&\qquad f_\varepsilon'\left(z_{i-1} - \mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1}\cdot(\mathbf{x}+\widetilde{\mathbf{x}})\right) \prod_{j\neq i} f_\varepsilon\left(z_{j-1} - \mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1}\cdot(\mathbf{x}+\widetilde{\mathbf{x}})\right)\,,
\end{aligned}
$$

where the first equality follows from Lemma B.3, the second equality follows from Lemma B.1(ii) and the third equality follows from the change of variables $\mathbf{A}_T^{-1}\mathbf{z}\mapsto\mathbf{z}$, where $\det\mathbf{A}_T^{-1} = (\det\mathbf{A}_T)^{-1} = 1$. Therefore, for arbitrary $\mathbf{y}\in\mathbb{R}^T$,

$$
\begin{aligned}
&\nabla_{\mathbf{z}} G_{k,n}(\mathbf{z}, s_1, s_2, A)\Big|_{\mathbf{z}=\mathbf{x}+\mathbf{y}} - \nabla_{\mathbf{z}} G_{k,n}(\mathbf{z}, s_1, s_2, A)\Big|_{\mathbf{z}=\mathbf{x}} \\
&= \sum_{i=1}^{T} \left(\mathbf{A}_T^{-1}\right)^\top \mathbf{e}_i \\
&\qquad \int_{\mathbb{R}^T} G_n(\mathbf{A}_T\cdot\mathbf{z}, s_1, s_2, A)\,\mathrm{d}\mathbf{z} \int_{\mathbb{R}^T} \mathrm{d}\mathbf{P}_{[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell}(\widetilde{\mathbf{x}}) \\
&\qquad \Bigg[f_\varepsilon'\left(z_{i-1} - \mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1}\cdot(\mathbf{x}+\widetilde{\mathbf{x}})\right) \prod_{j\neq i} f_\varepsilon\left(z_{j-1} - \mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1}\cdot(\mathbf{x}+\widetilde{\mathbf{x}})\right) \\
&\qquad - f_\varepsilon'\left(z_{i-1} - \mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1}\cdot(\mathbf{x}+\mathbf{y}+\widetilde{\mathbf{x}})\right) \prod_{j\neq i} f_\varepsilon\left(z_{j-1} - \mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1}\cdot(\mathbf{x}+\mathbf{y}+\widetilde{\mathbf{x}})\right) \Bigg]\,,
\end{aligned}
$$

which, by Lemma D.2(i), is equal to

$$
\begin{aligned}
&\sum_{i=1}^{T} \left(\mathbf{A}_T^{-1}\right)^\top \mathbf{e}_i \\
&\qquad \int_{\mathbb{R}^T} G_n(\mathbf{A}_T\cdot\mathbf{z}, s_1, s_2, A)\,\mathrm{d}\mathbf{z} \int_{\mathbb{R}^T} \mathrm{d}\mathbf{P}_{[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell}(\widetilde{\mathbf{x}}) \\
&\qquad \sum_{l=1}^{T} \widetilde{R}_{i,l}\left(\left(j \;\mapsto\; z_{j-1} - \mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1}\cdot(\mathbf{x}+\widetilde{\mathbf{x}})\right), \left(j \;\mapsto\; -\mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1}\cdot\mathbf{y}\right)\right) \\
&= \left(\mathbf{A}_T^{-1}\right)^\top \\
&\quad \times \Bigg[\int_{\mathbb{R}^T} G_n(\mathbf{A}_T\cdot\mathbf{z}, s_1, s_2, A)\,\mathrm{d}\mathbf{z} \int_{\mathbb{R}^T} \mathrm{d}\mathbf{P}_{[X_{\ell,k+\ell}-X_{\ell,\ell}]_\ell}(\widetilde{\mathbf{x}}) \\
&\qquad \widetilde{R}_{i,l}\left(\left(j \;\mapsto\; z_{j-1} - \mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1}\cdot(\mathbf{x}+\widetilde{\mathbf{x}})\right), \left(j \;\mapsto\; -\mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1}\cdot\mathbf{y}\right)\right)\Bigg]_{i,l\in\{1,\ldots,T\}} \cdot \mathbf{1}_T\,.
\end{aligned}
$$

Proof of (ii): From Lemma B.3 it follows for all $\mathbf{x}\in\mathbb{R}^T$ that

$$
\begin{aligned}
&\nabla_{\mathbf{x}} G_{\infty,n}(\mathbf{x}, s_1, s_2, A) - \nabla_{\mathbf{x}} G_{k,n}(\mathbf{x}, s_1, s_2, A) \\
&= \int_{\mathbb{R}^T} G_n(\mathbf{z}, s_1, s_2, A)\left(\nabla f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{z}-\mathbf{x}) - \nabla f_{[X_\ell]_\ell}(\mathbf{z}-\mathbf{x})\right)\mathrm{d}\mathbf{z} \\
&= \int_{\mathbb{R}^T} G_n(\mathbf{A}_T\cdot\mathbf{z}, s_1, s_2, A)\left(\nabla f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{A}_T\cdot\mathbf{z}-\mathbf{x}) - \nabla f_{[X_\ell]_\ell}(\mathbf{A}_T\cdot\mathbf{z}-\mathbf{x})\right)\mathrm{d}\mathbf{z}\,,
\end{aligned}
\tag{25}
$$

where for the change of variables in the last step we used again $\det\mathbf{A}_T = 1$. Note that by Lemma B.1(ii) again,

$$\begin{aligned}
&\nabla f_{[X_{\ell,k+\ell}]_\ell}(\mathbf{A}_T \cdot \mathbf{z} - \mathbf{x}) \;-\; \nabla f_{[X_\ell]_\ell}(\mathbf{A}_T \cdot \mathbf{z} - \mathbf{x}) \\
&= \left(\mathbf{A}_T^{-1}\right)^\top \\
&\times \Bigg[\mathbf{E}\Bigg[f_\varepsilon'(z_{i-1} \;-\; \mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot (\mathbf{x} + [X_{\ell,k+\ell} - X_{\ell,\ell}]_\ell)) \prod_{j\neq i} f_\varepsilon(z_{j-1} \;-\; \mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot (\mathbf{x} + [X_{\ell,k+\ell} - X_{\ell,\ell}]_\ell)) \\
&\qquad - f_\varepsilon'(z_{i-1} \;-\; \mathbf{e}_i^\top \cdot \mathbf{A}_T^{-1} \cdot (\mathbf{x} + [X_\ell - X_{\ell,\ell}]_\ell)) \prod_{j\neq i} f_\varepsilon(z_{j-1} \;-\; \mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot (\mathbf{x} + [X_\ell - X_{\ell,\ell}]_\ell))\Bigg]\Bigg]_{i\in\{1,\dots,T\}} .
\end{aligned}$$

By Lemma D.2(i) and independence of the random vectors $[X_{\ell,k+\ell} - X_{\ell,\ell}]_\ell$ and $[X_\ell - X_{\ell,k+\ell}]_\ell$, this equals

$$\begin{aligned}
&\left(\mathbf{A}_T^{-1}\right)^\top \\
&\times \Bigg[\mathbf{E}\Bigg[\sum_{l=1}^T \widetilde{R}_{i,l}\Big(\Big(j \;\mapsto\; z_{j-1} \;-\; \mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot (\mathbf{x} \;+\; [X_{\ell,k+\ell} - X_{\ell,\ell}]_\ell)\Big)\;, \\
&\qquad\qquad \Big(j \;\mapsto\; -\mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot [X_\ell - X_{\ell,k+\ell}]_\ell\Big)\Big)\Bigg]\Bigg]_{i\in\{1,\dots,T\}} \\
&= \left(\mathbf{A}_T^{-1}\right)^\top \\
&\quad \times \Bigg[\int_{\mathbb{R}^T} \mathrm{d}\mathbf{P}_{[X_\ell - X_{\ell,k+\ell}]_\ell}(\widetilde{\mathbf{x}}) \int_{\mathbb{R}^T} \mathrm{d}\mathbf{P}_{[X_{\ell,k+\ell} - X_{\ell,\ell}]_\ell}(\widetilde{\mathbf{y}}) \\
&\qquad \widetilde{R}_{i,l}\Big(\Big(j \;\mapsto\; z_{j-1} \;-\; \mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot (\mathbf{x} \;+\; \widetilde{\mathbf{y}})\Big)\;, \Big(j \;\mapsto\; -\mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot \widetilde{\mathbf{x}}\Big)\Big)\Bigg]_{i\in\{1,\dots,T\}} \cdot \mathbf{1}_T
\end{aligned}$$

The proof is complete. □

The final result of this section bounds the right-hand sides of the integrands in the gradient difference identities of Lemma D.3, and their variants needed later. It thereby shows how the analytic estimates developed so far are brought into the probabilistic setting considered in the next section.

Lemma D.4 (Separable product bounds IV)**.** *Let Assumptions 1 and 2 hold, and let* $\mathfrak{b}, \mathfrak{c} \in \mathbb{R}$ *,* $\varphi, \psi : \{1, \dots, T\} \to \mathbb{R}$ *,* $\gamma \in (0, \alpha - 1)$ *, and* $i, l \in \{1, \dots, T\}$ *. Then:*

(i)

$$\begin{aligned}
&\int_{\mathbb{R}^T} G_n(\mathbf{A}_T \cdot \mathbf{z}, s_1, s_2, A) \left|\widetilde{R}_{i,l}\left(j \mapsto z_{j-1} + \varphi(j), \psi\right)\right| \mathrm{d}\mathbf{z} \\
&\lesssim u_n^{-\gamma} \cdot ((s_2 - s_1) \wedge 1) \sum_{j=1}^T \sum_{I\subset\{1,\dots,j\}} \left(1 + \left|\sum_{m=1}^j a_{j-m}\varphi(m)\right|^{1+\gamma}\right) \left(|\psi(l)| + \left|\sum_{m\in I} a_{j-m}\psi(m)\right|^{1+\gamma}\right)
\end{aligned}$$

(ii)

$$\begin{aligned}
&\int_{\mathfrak{b}\wedge\mathfrak{c}}^{\mathfrak{b}\vee\mathfrak{c}} \left(\int_{\mathbb{R}^T} G_n(\mathbf{A}_T \cdot \mathbf{z}, s_1, s_2, A) \left|\widetilde{R}_{i,l}(j \mapsto z_{j-1} + \varphi(j), s \cdot \psi)\right| \mathrm{d}\mathbf{z}\right) \mathrm{d}s \\
&\lesssim u_n^{-\gamma} \cdot ((s_2 - s_1) \wedge 1) \\
&\times \sum_{j=1}^T \sum_{I\subset\{1,\dots,j\}} \left(\left|\sum_{m\in I} a_{j-m}\psi(m)\right|^\gamma \;+\; |\psi(l)|^\gamma\right) \;\times\; \left(1 \;+\; \left|\sum_{m=1}^j a_{j-m}\varphi(m)\right|^{1+\gamma}\right) \left(|\mathfrak{b}|^{1+\gamma} + |\mathfrak{c}|^{1+\gamma}\right) .
\end{aligned}$$

In both (i) and (ii), $\lesssim$ *means inequality up to a multiplicative constant that depends only on* T *,* (a_m) *,* $(\mathfrak{a}_j)$ *,* γ *,* $\underline{s}$ *and the distribution of* ε *.*

Proof of (i): By Lemma D.2(ii),

$$\begin{aligned}&\left|\widetilde{R}_{i,l}\left(j \mapsto z_{j-1}+\varphi(j), \psi\right)\right| \\ &\lesssim (1 \wedge |\psi(l)|) \sum_{I \subset\{1, \ldots, T\}}\left(\prod_{m \notin I} g_\gamma(z_{m-1}+\varphi(m))\right)\left(\prod_{m \in I} g_\gamma(z_{m-1}+\varphi(m)+\psi(m))\right),\end{aligned}$$

where the multiplicative constant in $\lesssim$ depends only on the distribution of ε. We invoke this bound inside the integral to get

$$\begin{aligned}&\int_{\mathbb{R}^T} G_n(\mathbf{A}_T \cdot \mathbf{z}, s_1, s_2, A)\left|\widetilde{R}_{i,l}\left(j \mapsto z_{j-1}+\varphi(j), \psi\right)\right| \mathrm{d}\mathbf{z} \\ &\lesssim (1 \wedge |\psi(l)|) \\ &\quad \times \sum_{I \subset\{1, \ldots, T\}} \int_{\mathbb{R}^T} G_n(\mathbf{A}_T \cdot \mathbf{z}, s_1, s_2, A)\left(\prod_{m \notin I} g_\gamma(z_{m-1}+\varphi(m))\right)\left(\prod_{m \in I} g_\gamma(z_{m-1}+\varphi(m)+\psi(m))\right) \mathrm{d}\mathbf{z},\end{aligned}$$

where the multiplicative constant in $\lesssim$ depends only on the distribution of ε. Applying Lemma C.3, for all $I \subset \{1, \ldots, T\}$, with

$$\mathfrak{y}_m = \varphi(m) + \mathbb{1}\{m \in I\} \cdot \psi(m), \qquad m \in \{1, \ldots, T\},$$

this is bounded above by

$$\begin{aligned}&(1 \wedge |\psi(l)|) \sum_{I \subset\{1, \ldots, T\}} \sum_{j=1}^{T} \\ &\int_{\mathbb{R}^j} \mathbb{1}\left\{\left|\sum_{m=1}^{j} a_{j-m} z_{m-1}\right| \cdot u_n \mathfrak{a}_j \cdot [s_1, s_2]\right\}\left(\prod_{\substack{m \notin I \\ m \leq j}} g_\gamma(z_{m-1}+\varphi(m))\right)\left(\prod_{\substack{m \in I \\ m \leq j}} g_\gamma(z_{m-1}+\varphi(m)+\psi(m))\right) \mathrm{d}\mathbf{z},\end{aligned}$$

up to a multiplicative constant that depends only on T, γ, and the distribution of ε. Since sets $I, I' \subset \{1, \ldots, T\}$ with $I \cap \{1, \ldots, j\} = I' \cap \{1, \ldots, j\}$ yield the same summand, this is bounded above by a multiple of

$$\begin{aligned}&(1 \wedge |\psi(l)|) \sum_{j=1}^{T} \sum_{I \subset\{1, \ldots, j\}} \\ &\int_{\mathbb{R}^j} \mathbb{1}\left\{\left|\sum_{m=1}^{j} a_{j-m} z_{m-1}\right| \cdot u_n \cdot \mathfrak{a}_j \cdot [s_1, s_2]\right\}\left(\prod_{m \notin I} g_\gamma(z_{m-1}+\varphi(m))\right)\left(\prod_{m \in I} g_\gamma(z_{m-1}+\varphi(m)+\psi(m))\right) \mathrm{d}\mathbf{z}.\end{aligned}$$

By Lemma C.2(i), this is then bounded above by

$$u_n^{-\gamma} \cdot ((s_2 - s_1) \wedge 1) \sum_{j=1}^{T} \sum_{I \subset\{1, \ldots, j\}}\left(1+\left|\sum_{m=1}^{j} a_{j-m} \varphi(m)\right|^{1+\gamma}\right)\left(|\psi(l)|+\left|\sum_{m \in I} a_{j-m} \psi(m)\right|^{1+\gamma}\right),$$

up to a multiplicative constant that depends only on T, (a_m), $(\mathfrak{a}_j)$, γ, $\underline{s}$ and the distribution of ε.

Proof of (ii): By Lemma D.2(ii),

$$\begin{aligned}&\left|\widetilde{R}_{i,l}(j \mapsto z_{j-1}+\varphi(j), s \cdot \psi)\right| \\ &\lesssim (1 \wedge |s \cdot \psi(l)|) \sum_{I \subset\{1, \ldots, T\}}\left(\prod_{m \notin I} g_\gamma(z_{m-1}+\varphi(m))\right)\left(\prod_{m \in I} g_\gamma(z_{m-1}+\varphi(m)+s \cdot \psi(m))\right),\end{aligned}$$

where the multiplicative constant in $\lesssim$ depends only on the distribution of ε. Applying Lemma C.3, for all $I \subset \{1, \dots, T\}$, with

$$\mathfrak{y}_m \;=\; \varphi(m) \;+\; \mathbb{1}\{m \in I\} \cdot s \cdot \psi(m)\,, \qquad m \in \{1, \dots, T\}\,,$$

and proceeding as in the proof of (i), we get

$$\begin{aligned}
&\int_{\mathfrak{b}\wedge\mathfrak{c}}^{\mathfrak{b}\vee\mathfrak{c}} \left(\int_{\mathbb{R}^T} G_n(\mathbf{A}_T \cdot \mathbf{z}, s_1, s_2, A) \left| \widetilde{R}_{i,l}(j \mapsto z_{j-1} + \varphi(j), s \cdot \psi) \right| \mathrm{d}\mathbf{z} \right) \mathrm{d}s \\
&\lesssim \sum_{j=1}^{T} \sum_{I \subset \{1,\dots,j\}} \int_{\mathfrak{b}\wedge\mathfrak{c}}^{\mathfrak{b}\vee\mathfrak{c}} (1 \wedge |s \cdot \psi(l)|)\, \mathrm{d}s \\
&\qquad \times \left(\int_{\mathbb{R}^j} \mathbb{1}\left\{ \left| \sum_{m=1}^{j} a_{j-m} z_{m-1} \right| \in u_n \cdot \mathfrak{a}_j \cdot [s_1, s_2] \right\} \right. \\
&\qquad\qquad \left. \times \prod_{m \notin I} g_\gamma(z_{m-1} + \varphi(m)) \cdot \prod_{m \in I} g_\gamma(z_{m-1} + \varphi(m) + s \cdot \psi(m))\, \mathrm{d}\mathbf{z} \right) \\
&\lesssim u_n^{-\gamma} \cdot ((s_2 - s_1) \wedge 1) \\
&\qquad \times \sum_{j=1}^{T} \sum_{I \subset \{1,\dots,j\}} \left(\left| \sum_{m \in I} a_{j-m} \psi(m) \right|^{\gamma} \;+\; |\psi(l)|^{\gamma} \right) \;\times\; \left(1 \;+\; \left| \sum_{m=1}^{j} a_{j-m} \varphi(m) \right|^{1+\gamma} \right) \left(|\mathfrak{b}|^{1+\gamma} + |\mathfrak{c}|^{1+\gamma} \right)\,,
\end{aligned}$$

where the last step follows from Lemma C.2(ii). The multiplicative constant in $\lesssim$ depends only on T, (a_m), $(\mathfrak{a}_j)$, γ, $\underline{s}$ and the distribution of ε. □

E. Reduction Principle

We are ready to analyze the probabilistic part of the problem. We first formulate the main result of the section, which is Lemma E.1. We next provide a short proof of Theorem 3.2, assuming that Lemma E.1 holds, and we then proceed to prove Lemma E.1 in several steps.

Lemma E.1 (L^r-reduction principle)**.** *Let Assumptions 1 and 2 hold. For all $\gamma \in (0,1)$ and $r \in [1,2]$ satisfying*

$$\gamma \;<\; \min\left\{ \frac{d}{1-d}\,, 1 - \frac{1}{r(1-d)}\,, \frac{\alpha}{r} - 1 \right\} \qquad \textit{and} \qquad \frac{1}{1-d} \;<\; r \;<\; \alpha\,, \tag{26}$$

it holds for $n \in \mathbb{N}$,

$$\begin{aligned}
&\mathbf{E}\left[\left| \sum_{t=1}^{n-T+1} G_n([X_{t+\ell}]_\ell, s_1, s_2, A) \;-\; \mathbf{E}[G_n([X_{t+\ell}]_\ell, s_1, s_2, A)] \;-\; \left(\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s_1, s_2, A) \Big|_{\mathbf{y}=\mathbf{0}_T} \right)^{\top} \cdot [X_{t+\ell}]_\ell \right|^{r} \right] \\
&\lesssim\; ((s_2 - s_1) \wedge 1) \cdot n^{r+1-(1-d)(1+\gamma)r}\,,
\end{aligned}$$

where $\lesssim$ means inequality up to a multiplicative constant that depends only on A, T, $\underline{s}$, γ, r, the coefficients (a_i), and the distribution of ε, and is independent of s_1, s_2, and n.

Proof of Theorem 3.2: Following Remark 3.4 in Scheffel et al. (2026), a version of the upper bound in Lemma E.1 with optimal parameter values (γ_0, r_0) is given in Theorem 3.2, where

$$(\gamma_0, r_0) := \begin{cases} \left(\dfrac{d}{1-d},\; \alpha(1-d) \right) & \text{if } \dfrac{1}{(1-d)(1-2d)} < \alpha\,, \\[2ex] \left(\dfrac{\alpha(1-d) - 1}{\alpha(1-d) + 1},\; \dfrac{1}{2}\left(\alpha + \dfrac{1}{1-d} \right) \right) & \text{otherwise,} \end{cases} \tag{27}$$

yielding the optimal value $\kappa_0 := \kappa(\gamma_0, r_0) = 1 + 1/r_0 - (1-d)(1+\gamma_0)$ as

$$\kappa_0 = \begin{cases} \dfrac{1}{\alpha(1-d)} \quad \text{if} \quad \dfrac{1}{(1-d)(1-2d)} < \alpha\,, \\ \dfrac{2(1-d) + (1-\alpha(1-d)(1-2d))}{\alpha(1-d)+1} = d + (1-d)\dfrac{3-\alpha(1-d)}{\alpha(1-d)+1} \quad \text{otherwise.} \end{cases} \tag{28}$$

We remark that, while γ_0 lies on the boundary of the admissible parameter range (26), the value $\gamma_0 - \delta$ lies inside this range for each $\delta > 0$ sufficiently small. We may therefore apply Lemma E.1 with $(\gamma, r) = (\gamma_0 - \delta, r_0)$. The multiplicative constant in $\lesssim$ then also depends on δ. □

E.1. Martingale decomposition

The derivation of the martingale decomposition follows the argument that leads to Lemma B.3 in Scheffel et al. (2026), where its conceptual motivation is explained in detail. Define, for $n \in \mathbb{N}$, $t \in \mathbb{Z}$,

$$\begin{aligned} &\mathcal{U}_n([X_{t+\ell}]_\ell, s_1, s_2, A) \\ &:= \; G_n([X_{t+\ell}]_\ell, s_1, s_2, A) \;-\; \mathbf{E}[G_n([X_{t+\ell}]_\ell, s_1, s_2, A)] \;-\; \left(\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=\mathbf{0}_T}\right)^{\top} \cdot [X_{t+\ell}]_\ell \end{aligned}$$

where the gradient $\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=\mathbf{0}_T}$ is well-defined under the assumptions of Lemma B.3. Let $\mathcal{F}_k = \sigma(\varepsilon_k, \varepsilon_{k-1}, \ldots)$ be the past σ-algebra generated by the sequence (ε_t) up to index k. For $Y \in L^1(\mathbf{P})$, we define the projection

$$P_k Y \;:=\; \mathbf{E}[Y \mid \mathcal{F}_k] \;-\; \mathbf{E}[Y \mid \mathcal{F}_{k-1}] \qquad \text{for all } k \in \mathbb{Z}\,.$$

Lemma E.2 (Moment bounds for the centered process $\mathcal{U}_n$ and its projection)**.** *Let Assumptions 1 and 2 hold. Then, the following holds for all $r \in [1, \alpha)$:*

(i) $\|\mathcal{U}_n([X_\ell]_\ell, s_1, s_2, A)\|_{L^r(\mathbf{P})} \lesssim ((s_2 - s_1) \wedge 1)^{1/r}$*, and in particular* $\mathcal{U}_n([X_\ell]_\ell, s_1, s_2, A) \in L^r(\mathbf{P})$.

(ii) $\|P_k \mathcal{U}_n([X_\ell]_\ell, s_1, s_2, A)\|_{L^r(\mathbf{P})} \lesssim \|\mathcal{U}_n([X_\ell]_\ell, s_1, s_2, A)\|_{L^r(\mathbf{P})}$*, and* $P_k \mathcal{U}_n([X_\ell]_\ell, s_1, s_2, A) \in L^r(\mathbf{P})$ *for all* $k \in \mathbb{Z}$.

In both (i) and (ii) $\lesssim$ means inequality up to a multiplicative constant that depends only on A, T, $\underline{s}$, γ, r, the coefficients (a_i), and the distribution of ε, and is independent of s_1, s_2, n, and k.

Remark E.3 (Dropping $u_n^{-\gamma/r}$ rate)**.** A sharper bound with a factor $u_n^{-\gamma/r} \in (0,1)$ is possible. Since the final analysis does not crucially depend on this factor and the gains of keeping it are small, we drop it.

Proof of (i): Since, for all $\mathbf{x} \in \mathbb{R}^T$,

$$\left\|\mathbf{x}^\top \cdot [X_\ell]_\ell\right\|_{L^r(\mathbf{P})} \;\le\; \sum_{\ell=0}^{T-1} |x_\ell|\, \|X_\ell\|_{L^r(\mathbf{P})} \;\le\; T\, \|\mathbf{x}\|_\infty\, \|X_0\|_{L^r(\mathbf{P})}\,,$$

by stationarity of (X_t), it follows from the definition of $\mathcal{U}_n$ that

$$\begin{aligned}
&\|\mathcal{U}_n([X_\ell]_\ell, s_1, s_2, A)\|_{L^r(\mathbf{P})} \\
&\le \left(\mathbf{P}[[X_\ell]_\ell \in u_n \cdot ((s_1 \cdot A) \setminus (s_2 \cdot A))]\right)^{1/r} + \mathbf{P}[[X_\ell]_\ell \in u_n \cdot ((s_1 \cdot A) \setminus (s_2 \cdot A))] \\
&\quad + \left\| \left(\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=\mathbf{0}_T} \right)^\top \cdot [X_\ell]_\ell \right\|_{L^r(\mathbf{P})} \\
&\le 2\left(\mathbf{P}[[X_\ell]_\ell \in u_n \cdot ((s_1 \cdot A) \setminus (s_2 \cdot A))]\right)^{1/r} + T \left\| \nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=\mathbf{0}_T} \right\|_\infty \|X_0\|_{L^r(\mathbf{P})} \\
&\lesssim \left(\mathbf{P}[[X_\ell]_\ell \in u_n \cdot ((s_1 \cdot A) \setminus (s_2 \cdot A))]\right)^{1/r} + \left\| \nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=\mathbf{0}_T} \right\|_\infty
\end{aligned}$$

since $X_0 \in L^r(\mathbf{P})$ by Lemma A.4. The constant in $\lesssim$ depends only on T, r, the coefficients (a_i), and the distribution of ε. By Lemma C.4, dropping the $u_n^{-\gamma}$ rate (see Remark E.3), this is bounded above by

$$((s_2 - s_1) \wedge 1)^{1/r} + ((s_2 - s_1) \wedge 1) \lesssim ((s_2 - s_1) \wedge 1)^{1/r},$$

up to a multiplicative constant that depends only on A, T, $\underline{s}$, r, γ, the coefficients (a_i), and the distribution of ε.

Proof of (ii): By Jensen's inequality and the properties of conditional expectation,

$$\|P_k Y\|_{L^r(\mathbf{P})} \le \|\mathbf{E}[Y \mid \mathcal{F}_k]\|_{L^r(\mathbf{P})} + \|\mathbf{E}[Y \mid \mathcal{F}_{k-1}]\|_{L^r(\mathbf{P})} \le 2\|Y\|_{L^r(\mathbf{P})} < \infty$$

for $Y \in L^r(\mathbf{P})$. The statement then follows from part (i). □

We define

$$\begin{aligned}
\mathcal{S}_n(s_1, s_2, A) &:= \sum_{t=1}^{n-T+1} \mathcal{U}_n([X_{t+\ell}]_\ell, s_1, s_2, A) \\
&= \sum_{t=1}^{n-T+1} \left(G_n([X_{t+\ell}]_\ell, s_1, s_2, A) - \mathbf{E}[G_n([X_{t+\ell}]_\ell, s_1, s_2, A)] - \left(\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=\mathbf{0}_T} \right)^\top [X_{t+\ell}]_\ell \right).
\end{aligned}$$

Lemma E.4 (Moment bounds after martingale decomposition)**.** *Let Assumptions 1 and 2 hold. Then, for all $r \in [1, \alpha)$,*

$$\mathbf{E}\,|\mathcal{S}_n(s_1, s_2, A)|^r \lesssim \sum_{k \le n} \left(\sum_{t=(k-T+1)\vee 1}^{n-T+1} \|P_{k-t}\,\mathcal{U}_n([X_\ell]_\ell, s_1, s_2, A)\|_{L^r(\mathbf{P})} \right)^r,$$

where $\lesssim$ means inequality up to a multiplicative constant that is independent of all other quantities in this work, in particular, independent of r and n.

Proof. For a proof with $T = 1$ see Lemma B.2 in Scheffel et al. (2026). The proof extends readily to the case $T > 1$ by considering $[X_0, \ldots, X_{T-1}]$ instead of X_0 and noting that $\mathcal{S}_n(s_1, s_2, A)$ is $\mathcal{F}_n$-measurable, and $P_k\,\mathcal{U}_n([X_{t+\ell}]_\ell, s_1, s_2, A) = 0$ for $k - T + 1 > t$. We omit the details. □

Lemma E.5 (Error decomposition)**.** *Let Assumptions 1 and 2 hold. Then, for all* $k \in \mathbb{N}$ *, it holds* $P_{-k}\,\mathcal{U}_n([X_\ell]_\ell, s_1, s_2, A) = R_1 + R_2 + R_3$ *, with*

$$\begin{aligned}
R_1 \;&:=\; G_{k-1,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=[X_\ell - X_{\ell,k-1+\ell}]_\ell} \;-\; \int_{\mathbb{R}} G_{k-1,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=[X_\ell - X_{\ell,k+\ell} + z\cdot a_{k+\ell}]_\ell} \,\mathrm{d}\mathbf{P}_\varepsilon(z) \\
&\quad -\; \varepsilon_{-k}\cdot [a_{k+\ell}]_\ell^\top \cdot \nabla_{\mathbf{y}} G_{k-1,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=[X_\ell - X_{\ell,k+\ell}]_\ell}\,, \\
R_2 \;&:=\; \varepsilon_{-k}\cdot [a_{k+\ell}]_\ell^\top \cdot \left(\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=[X_\ell - X_{\ell,k+\ell}]_\ell} \;-\; \nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=\mathbf{0}_T}\right)\,, \\
R_3 \;&:=\; \varepsilon_{-k}\cdot [a_{k+\ell}]_\ell^\top \cdot \left(\nabla_{\mathbf{y}} G_{k-1,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=[X_\ell - X_{\ell,k+\ell}]_\ell} \;-\; \nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=[X_\ell - X_{\ell,k+\ell}]_\ell}\right)\,.
\end{aligned}$$

Proof. By definition,

$$\begin{aligned}
&R_1 \;+\; R_2 \;+\; R_3 \\
&\;=\; G_{k-1,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=[X_\ell - X_{\ell,k-1+\ell}]_\ell} \;-\; \int_{\mathbb{R}} G_{k-1,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=[X_\ell - X_{\ell,k+\ell} + z\cdot a_{k+\ell}]_\ell} \,\mathrm{d}\mathbf{P}_\varepsilon(z) \\
&\quad -\; \varepsilon_{-k}\cdot [a_{k+\ell}]_\ell^\top \cdot \nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=\mathbf{0}_T}\,.
\end{aligned}$$

For fixed $j, k \in \mathbb{Z}$, observe that

$$P_{-k}\varepsilon_{-j} \;=\; \begin{cases} 0 & \text{if } j \neq k\,, \\ \varepsilon_{-k} & \text{if } j = k\,, \end{cases}$$

so, by dominated convergence, it follows that

$$P_k\, X_\ell \;=\; \sum_{i=0}^{\infty} a_i P_k\, \varepsilon_{\ell-i} \;=\; \begin{cases} 0 & \text{if } k > \ell\,, \\ a_{\ell-k}\varepsilon_k & \text{if } k \leq \ell\,. \end{cases}$$

Thus,

$$\begin{aligned}
&P_k\,\mathcal{U}_n([X_\ell]_\ell, s_1, s_2, A) \\
&\;=\; P_k\left(G_n([X_\ell]_\ell, s_1, s_2, A) \;-\; \mathbf{E}[G_n([X_\ell]_\ell, s_1, s_2, A)] \;-\; \nabla_{\mathbf{y}}\left(G_{\infty,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=\mathbf{0}_T}\right)^\top \cdot [X_\ell]_\ell\right) \\
&\;=\; \mathbf{E}[G_n([X_\ell]_\ell, s_1, s_2, A)\,|\,\mathcal{F}_k] \;-\; \mathbf{E}[G_n([X_\ell]_\ell, s_1, s_2, A)\,|\,\mathcal{F}_{k-1}] \\
&\qquad -\; \varepsilon_k\cdot\left(\left(\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=\mathbf{0}_T}\right)^\top \cdot [a_{\ell-k}\cdot \mathbb{1}\{k \leq \ell\}]_\ell\right)\,.
\end{aligned} \tag{29}$$

Since we assume $k \geq 1$ and $\ell \in \{0, \ldots, T-1\}$, it holds $\mathbb{1}\{-k \leq \ell\} = 1$, so that by (29),

$$\begin{aligned}
&P_{-k}\,\mathcal{U}_n([X_\ell]_\ell, s_1, s_2, A) \\
&\;=\; \mathbf{E}[G_n([X_\ell]_\ell, s_1, s_2, A)\,|\,\mathcal{F}_{-k}] \;-\; \mathbf{E}[G_n([X_\ell]_\ell, s_1, s_2, A)\,|\,\mathcal{F}_{-(k+1)}] \\
&\qquad -\; \varepsilon_{-k}\cdot [a_{k+\ell}]_\ell^\top \cdot \nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=\mathbf{0}_T}\,.
\end{aligned}$$

Note that $X_{t,t+k-1} = \sum_{j=0}^{t+k-1} a_j\varepsilon_{t-j}$ is independent of $\mathcal{F}_{-k}$, and $X_t - X_{t,t+k-1} = \sum_{j=t+k}^{\infty} a_j\varepsilon_{t-j}$ is $\mathcal{F}_{-k}$-measurable. Therefore,

$$\begin{aligned}
\mathbf{E}[G_n([X_\ell]_\ell, s_1, s_2, A)\,|\,\mathcal{F}_{-k}] \;&=\; \mathbf{E}[G_n([X_{\ell,k-1+\ell} + (X_\ell - X_{\ell,k-1+\ell})]_\ell, s_1, s_2, A \mid \mathcal{F}_{-k}] \\
&=\; \int_{\mathbb{R}^T} G_n(\mathbf{x} + [X_\ell - X_{\ell,k-1+\ell}]_\ell, s_1, s_2, A)\,\mathrm{d}\mathbf{P}_{[X_{\ell,k-1+\ell}]_\ell}(\mathbf{x}) \\
&=\; G_{k-1,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=[X_\ell - X_{\ell,k-1+\ell}]_\ell}\,.
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\mathbf{E}[G_n([X_\ell]_\ell, s_1, s_2, A) \mid \mathcal{F}_{-(k+1)}] \\
&= \int_{\mathbb{R}^T} G_n(\mathbf{x} + [X_\ell - X_{\ell,k+\ell}]_\ell, s_1, s_2, A)\, \mathrm{d}\mathbf{P}_{[X_{\ell,k+\ell}]_\ell}(\mathbf{x}) \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}^T} G_n(\mathbf{x} + [X_\ell - X_{\ell,k+\ell} + z \cdot a_{k+\ell}]_\ell, s_1, s_2, A)\, \mathrm{d}\mathbf{P}_{[X_{\ell,k-1+\ell}]_\ell}(\mathbf{x}) \right) \mathrm{d}\mathbf{P}_\varepsilon(z) \\
&= \int_{\mathbb{R}} G_{k-1,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=[X_\ell - X_{\ell,k+\ell} + z\cdot a_{k+\ell}]_\ell} \mathrm{d}\mathbf{P}_\varepsilon(z)\,.
\end{aligned}$$

Consequently, combining the expressions gives

$$\begin{aligned}
&P_{-k}\,\mathcal{U}_n([X_\ell]_\ell, s_1, s_2, A) \\
&= G_{k-1,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=[X_\ell - X_{\ell,k-1+\ell}]_\ell} - \int_{\mathbb{R}} G_{k-1,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=[X_\ell - X_{\ell,k+\ell}+z\cdot a_{k+\ell}]_\ell} \mathrm{d}\mathbf{P}_\varepsilon(z) \\
&\quad - \varepsilon_{-k} \cdot [a_{k+\ell}]_\ell^\top \cdot \nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=\mathbf{0}_T} \\
&= R_1 + R_2 + R_3\,,
\end{aligned}$$

as claimed. □

E.2. Analysis of the remainder terms R_j

In this subsection we use the tools developed in the previous sections. In the next lemma we provide intermediate bounds that are closely linked to the bounds of gradient differences provided in the last section.

Lemma E.6 (Intermediate bounds of the remainder terms)**.** *Let Assumptions 1 and 2 hold, and let $k \in \mathbb{N}$. Then the following statements are true:*

(i)

$$\begin{aligned}
&|R_1| \cdot k^{1-d} \\
&\lesssim \max_{i,l\in\{1,\dots,T\}} \Bigg[\int_{\mathbb{R}} \mathrm{d}\mathbf{P}_\varepsilon(z) \int_{\mathbb{R}^T} \mathrm{d}\mathbf{P}_{[X_{\ell,k-1+\ell}-X_{\ell,\ell}]}(\widetilde{\mathbf{x}}) \int_{\varepsilon_{-k}\wedge z}^{\varepsilon_{-k}\vee z} \mathrm{d}s \int_{\mathbb{R}^T} G_n(\mathbf{A}_T \cdot \mathbf{z}, s_1, s_2, A)\, \mathrm{d}\mathbf{z} \\
&\qquad \left| \widetilde{R}_{i,l} \left(j \mapsto z_{j-1} - \mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot ([X_\ell - X_{\ell,k+\ell}]_\ell + \widetilde{\mathbf{x}})\ , j \mapsto -s \cdot \mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot [a_{k+\ell}]_\ell \right) \right| \Bigg]\,.
\end{aligned}$$

(ii)

$$\begin{aligned}
&|R_2| \cdot k^{1-d} \\
&\lesssim |\varepsilon_{-k}| \\
&\quad \times \max_{i,l\in\{1,\dots,T\}} \Bigg[\int_{\mathbb{R}^T} \mathrm{d}\mathbf{P}_{[X_\ell - X_{\ell,\ell}]_\ell}(\widetilde{\mathbf{x}}) \int_{\mathbb{R}^T} G_n(\mathbf{A}_T \cdot \mathbf{z}, s_1, s_2, A)\, \mathrm{d}\mathbf{z} \\
&\qquad \left| \widetilde{R}_{i,l} \left(\left(j \mapsto z_{j-1} - \mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot \widetilde{\mathbf{x}} \right), \left(j \mapsto -\mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot [X_\ell - X_{\ell,k+\ell}]_\ell \right) \right) \right| \Bigg]\,.
\end{aligned}$$

(iii)

$$\begin{aligned}
&|R_3| \cdot k^{1-d} \\
&\lesssim |\varepsilon_{-k}| \\
&\times \max_{i,l\in\{1,\dots T\}} \Bigg[\int_{\mathbb{R}^T} \mathrm{d}\mathbf{P}_{[X_\ell - X_{\ell,k-1+\ell}]_\ell}(\widetilde{\mathbf{x}}) \int_{\mathbb{R}^T} \mathrm{d}\mathbf{P}_{[X_{\ell,k-1+\ell}-X_{\ell,\ell}]_\ell}(\widetilde{\mathbf{y}}) \int_{\mathbb{R}^T} G_n(\mathbf{A}_T \cdot \mathbf{z})\, \mathrm{d}\mathbf{z} \\
&\qquad \left| \widetilde{R}_{i,l} \left(\left(j \mapsto z_{j-1} - \mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot ([X_\ell - X_{\ell,k+\ell}]_\ell + \widetilde{\mathbf{y}}) \right), \left(j \mapsto -\mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot \widetilde{\mathbf{x}} \right) \right) \right| \Bigg]
\end{aligned}$$

In particular, in (ii) and (iii), ε_{-k} *is independent of the second factor. The multiplicative constant in* $\lesssim$ *depends throughout (i)-(iii) only on* T *and the coefficients* (a_i) *.*

Proof. Throughout the proof we use that

$$\left\|\mathbf{A}_T^{-1}\cdot[a_{k+\ell}]_\ell\right\|_\infty \;\lesssim\; k^{-(1-d)}\,, \tag{30}$$

where the multiplicative constant in $\lesssim$ depends only on T and the coefficients (a_i) .

Proof of (i): By definition,

$$\begin{aligned}
R_1 \;&=\; G_{k-1,n}(\mathbf{y},s_1,s_2,A)\Big|_{\mathbf{y}=[X_\ell-X_{\ell,k-1+\ell}]_\ell} \;-\; \int_{\mathbb{R}} G_{k-1,n}(\mathbf{y},s_1,s_2,A)\Big|_{\mathbf{y}=[X_\ell-X_{\ell,k+\ell}+z\cdot a_{k+\ell}]_\ell}\,\mathrm{d}\mathbf{P}_\varepsilon(z)\\
&\quad -\; \varepsilon_{-k}\cdot[a_{k+\ell}]_\ell^\top\cdot\nabla_{\mathbf{y}}G_{k-1,n}(\mathbf{y},s_1,s_2,A)\Big|_{\mathbf{y}=[X_\ell-X_{\ell,k+\ell}]_\ell}\\
&=\; \int_{\mathbb{R}}\Bigg[G_{k-1,n}(\mathbf{y},s_1,s_2,A)\Big|_{\mathbf{y}=[X_\ell-X_{\ell,k+\ell}+\varepsilon_{-k}\cdot a_{k+\ell}]_\ell} \;-\; G_{k-1,n}(\mathbf{y},s_1,s_2,A)\Big|_{\mathbf{y}=[X_\ell-X_{\ell,k+\ell}+z\cdot a_{k+\ell}]_\ell}\\
&\quad -\; \varepsilon_{-k}\cdot[a_{k+\ell}]_\ell^\top\cdot\nabla_{\mathbf{y}}G_{k-1,n}(\mathbf{y},s_1,s_2,A)\Big|_{\mathbf{y}=[X_\ell-X_{\ell,k+\ell}]_\ell}\Bigg]\,\mathrm{d}\mathbf{P}_\varepsilon(z)\\
&=\; \int_{\mathbb{R}}\Bigg[\left(\int_z^{\varepsilon_{-k}}[a_{k+\ell}]_\ell^\top\cdot\nabla_{\mathbf{y}}G_{k-1,n}(\mathbf{y},s_1,s_2,A)\Big|_{\mathbf{y}=[X_\ell-X_{\ell,k+\ell}+s\cdot a_{k+\ell}]_\ell}\,\mathrm{d}s\right)\\
&\quad -\; (\varepsilon_{-k}-z)\cdot[a_{k+\ell}]_\ell^\top\cdot\nabla_{\mathbf{y}}G_{k-1,n}(\mathbf{y},s_1,s_2,A)\Big|_{\mathbf{y}=[X_\ell-X_{\ell,k+\ell}]_\ell}\Bigg]\,\mathrm{d}\mathbf{P}_\varepsilon(z)\\
&=\; \int_{\mathbb{R}}\mathrm{d}\mathbf{P}_\varepsilon(z)\int_z^{\varepsilon_{-k}}\mathrm{d}s\\
&\quad [a_{k+\ell}]_\ell^\top\Bigg[\nabla_{\mathbf{y}}G_{k-1,n}(\mathbf{y},s_1,s_2,A)\Big|_{\mathbf{y}=[X_\ell-X_{\ell,k+\ell}+s\cdot a_{k+\ell}]_\ell} \;-\; \nabla_{\mathbf{y}}G_{k-1,n}(\mathbf{y},s_1,s_2,A)\Big|_{\mathbf{y}=[X_\ell-X_{\ell,k+\ell}]_\ell}\Bigg]\,.
\end{aligned} \tag{31}$$

Note that the third equality follows by applying the fundamental theorem of calculus to $G_{k-1,n}$, which is justified by Lemma B.3. Note also that adding the term

$$z\cdot[a_{k+\ell}]_\ell^\top\cdot\nabla_{\mathbf{y}}G_{k-1,n}(\mathbf{y},s_1,s_2,A)\Big|_{\mathbf{y}=[X_\ell-X_{\ell,k+\ell}]_\ell}$$

inside the integral is valid because ε is centered. Then we use Lemma D.3(i) for $k-1$ with

$$\mathbf{x} \;=\; [X_\ell-X_{\ell,k+\ell}]_\ell\,,\qquad \mathbf{y} \;=\; s\cdot[a_{k+\ell}]_\ell\,,$$

to get

$$\begin{aligned}
&\nabla_{\mathbf{y}}G_{k-1,n}(\mathbf{y},s_1,s_2,A)\Big|_{\mathbf{y}=[X_\ell-X_{\ell,k+\ell}+s\cdot a_{k+\ell}]_\ell} \;-\; \nabla_{\mathbf{y}}G_{k-1,n}(\mathbf{y},s_1,s_2,A)\Big|_{\mathbf{y}=[X_\ell-X_{\ell,k+\ell}]_\ell}\\
&\;=\; \left(\mathbf{A}_T^{-1}\right)^\top\\
&\quad\times\;\Bigg[\int_{\mathbb{R}^T}G_n(\mathbf{A}_T\cdot\mathbf{z},s_1,s_2,A)\,\mathrm{d}\mathbf{z}\int_{\mathbb{R}^T}\mathrm{d}\mathbf{P}_{[X_{\ell,k-1+\ell}-X_{\ell,\ell}]}(\widetilde{\mathbf{x}})\\
&\qquad \widetilde{R}_{i,l}\left(j\mapsto z_{j-1}-\mathbf{e}_j^\top\cdot\mathbf{A}_T^{-1}\cdot\left([X_\ell-X_{\ell,k+\ell}]_\ell+\widetilde{\mathbf{x}}\right)\,,j\mapsto -s\cdot\mathbf{e}_j^\top\cdot\mathbf{A}_T^{-1}\cdot[a_{k+\ell}]_\ell\right)\Bigg]_{i,l\in\{1,\dots,T\}}\cdot\mathbf{1}_T\,.
\end{aligned}$$

Therefore, taking absolute values and using (30) and Fubini's theorem,

$$\begin{aligned}
&|R_1|\\
&\lesssim\; k^{-(1-d)}\\
&\quad\times\max_{i,l\in\{1,\dots,T\}}\Bigg[\int_{\mathbb{R}}\mathrm{d}\mathbf{P}_\varepsilon(z)\int_{\mathbb{R}^T}\mathrm{d}\mathbf{P}_{[X_{\ell,k-1+\ell}-X_{\ell,\ell}]}(\widetilde{\mathbf{x}})\int_{\varepsilon_{-k}\wedge z}^{\varepsilon_{-k}\vee z}\mathrm{d}s\int_{\mathbb{R}^T}G_n(\mathbf{A}_T\cdot\mathbf{z},s_1,s_2,A)\,\mathrm{d}\mathbf{z}\\
&\qquad\left|\widetilde{R}_{i,l}\left(j\mapsto z_{j-1}-\mathbf{e}_j^\top\cdot\mathbf{A}_T^{-1}\cdot\left([X_\ell-X_{\ell,k+\ell}]_\ell+\widetilde{\mathbf{x}}\right)\,,j\mapsto -s\cdot\mathbf{e}_j^\top\cdot\mathbf{A}_T^{-1}\cdot[a_{k+\ell}]_\ell\right)\right|\Bigg]\,,
\end{aligned}$$

where the multiplicative constant in $\lesssim$ depends only on T and the coefficients (a_i).
Proof of (ii): By definition

$$R_2 \;=\; \varepsilon_{-k}\cdot[a_{k+\ell}]_\ell^\top\cdot\left(\nabla_{\mathbf{y}}G_{\infty,n}(\mathbf{y},s_1,s_2,A)\Big|_{\mathbf{y}=[X_\ell-X_{\ell,k+\ell}]_\ell} \;-\; \nabla_{\mathbf{y}}G_{\infty,n}(\mathbf{y},s_1,s_2,A)\Big|_{\mathbf{y}=\mathbf{0}_T}\right),$$

and by Lemma D.3(i) for $k=\infty$ with

$$\mathbf{x} \;=\; \mathbf{0}_T \qquad\text{and}\qquad \mathbf{y} \;=\; [X_\ell - X_{\ell,k+\ell}]_\ell\,,$$

one has

$$\begin{aligned}
&\nabla_{\mathbf{y}}G_{\infty,n}(\mathbf{y},s_1,s_2,A)\Big|_{\mathbf{y}=[X_\ell-X_{\ell,k+\ell}]_\ell} \;-\; \nabla_{\mathbf{y}}G_{\infty,n}(\mathbf{y},s_1,s_2,A)\Big|_{\mathbf{y}=\mathbf{0}_T}\\
&\;=\; \int_{\mathbb{R}^T} G_n(\mathbf{A}_T\cdot\mathbf{z},s_1,s_2,A)\,\mathrm{d}\mathbf{z}\int_{\mathbb{R}^T}\mathrm{d}\mathbf{P}_{[X_\ell-X_{\ell,\ell}]_\ell}(\widetilde{\mathbf{x}})\\
&\qquad\left(\mathbf{A}_T^{-1}\right)^\top\left[\widetilde{R}_{i,l}\left(\left(j\;\mapsto\; z_{j-1}-\mathbf{e}_j^\top\cdot\mathbf{A}_T^{-1}\cdot\widetilde{\mathbf{x}}\right),\left(j\;\mapsto\;-\mathbf{e}_j^\top\cdot\mathbf{A}_T^{-1}\cdot[X_\ell-X_{\ell,k+\ell}]_\ell\right)\right)\right]_{i,l\in\{1,\ldots,T\}}\cdot\mathbf{1}_T\,.
\end{aligned}$$

Taking absolute values and using (30) and Fubini's theorem yields the result.
Proof of (iii): By definition

$$R_3 \;:=\; \varepsilon_{-k}\cdot[a_{k+\ell}]_\ell^\top\cdot\left(\nabla_{\mathbf{y}}G_{k-1,n}(\mathbf{y},s_1,s_2,A)\Big|_{\mathbf{y}=[X_\ell-X_{\ell,k+\ell}]_\ell} \;-\; \nabla_{\mathbf{y}}G_{\infty,n}(\mathbf{y},s_1,s_2,A)\Big|_{\mathbf{y}=[X_\ell-X_{\ell,k+\ell}]_\ell}\right),$$

and by Lemma D.3(ii) for $k-1$ with

$$\mathbf{x} \;=\; [X_\ell - X_{\ell,k+\ell}]_\ell\,,$$

one has

$$\begin{aligned}
&\nabla_{\mathbf{y}}G_{\infty,n}(\mathbf{y},s_1,s_2,A)\Big|_{\mathbf{y}=[X_\ell-X_{\ell,k+\ell}]_\ell} \;-\; \nabla_{\mathbf{y}}G_{k-1,n}(\mathbf{y},s_1,s_2,A)\Big|_{\mathbf{y}=[X_\ell-X_{\ell,k+\ell}]_\ell}\\
&\;=\;\left(\mathbf{A}_T^{-1}\right)^\top\\
&\quad\times\left[\int_{\mathbb{R}^T}\mathrm{d}\mathbf{P}_{[X_\ell-X_{\ell,k-1+\ell}]_\ell}(\widetilde{\mathbf{x}})\int_{\mathbb{R}^T}\mathrm{d}\mathbf{P}_{[X_{\ell,k-1+\ell}-X_{\ell,\ell}]_\ell}(\widetilde{\mathbf{y}})\int_{\mathbb{R}^T}G_n(\mathbf{A}_T\cdot\mathbf{z},s_1,s_2,A)\,\mathrm{d}\mathbf{z}\right.\\
&\qquad\left.\widetilde{R}_{i,l}\left(\left(j\;\mapsto\; z_{j-1}\;-\;\mathbf{e}_j^\top\cdot\mathbf{A}_T^{-1}\cdot\left([X_\ell-X_{\ell,k+\ell}]_\ell\;+\;\widetilde{\mathbf{y}}\right)\right),\left(j\;\mapsto\;-\mathbf{e}_j^\top\cdot\mathbf{A}_T^{-1}\cdot\widetilde{\mathbf{x}}\right)\right)\right]_{i,l\in\{1,\ldots,T\}}\cdot\mathbf{1}_T\,.
\end{aligned}$$

Taking absolute values and using (30) and Fubini's theorem yields the result. □

In the next lemma we apply the results of the second to last section to separate the threshold u_n and the uniformity interval $[s_1,s_2]$ and simplify the intermediate bounds.

Lemma E.7 (Separable product bounds V). *Let Assumptions 1 and 2 hold. Then the following bounds hold for $\gamma\in(0,\alpha-1)$:*

(i)

$$\begin{aligned}
&\int_{\mathbb{R}}\mathrm{d}\mathbf{P}_\varepsilon(z)\int_{\mathbb{R}^T}\mathrm{d}\mathbf{P}_{[X_{\ell,k-1+\ell}-X_{\ell,\ell}]_\ell}(\widetilde{\mathbf{x}})\int_{\varepsilon_{-k}\wedge z}^{\varepsilon_{-k}\vee z}\mathrm{d}s\int_{\mathbb{R}^T}G_n(\mathbf{A}_T\cdot\mathbf{z},s_1,s_2,A)\,\mathrm{d}\mathbf{z}\\
&\qquad\qquad\left|\widetilde{R}_{i,l}\left(j\mapsto z_{j-1}-\mathbf{e}_j^\top\cdot\mathbf{A}_T^{-1}\cdot\left([X_\ell-X_{\ell,k+\ell}]_\ell+\widetilde{\mathbf{x}}\right),j\mapsto -s\cdot\mathbf{e}_j^\top\cdot\mathbf{A}_T^{-1}\cdot[a_{k+\ell}]_\ell\right)\right|\\
&\lesssim\; u_n^{-\gamma}\cdot((s_2-s_1)\wedge 1)\cdot k^{-\gamma(1-d)}\cdot\left(1+|\varepsilon_{-k}|^{1+\gamma}\right)\cdot\sum_{j=1}^{T}\left(1\;+\;|X_{j-1}-X_{j-1,k+j-1}|^{1+\gamma}\right).
\end{aligned}$$

(ii)

$$\begin{aligned}
&\int_{\mathbb{R}^T} \mathrm{d}\mathbf{P}_{[X_\ell - X_{\ell,\ell}]_\ell}(\widetilde{\mathbf{x}}) \int_{\mathbb{R}^T} G_n(\mathbf{A}_T \cdot \mathbf{z}, s_1, s_2, A)\,\mathrm{d}\mathbf{z} \\
&\qquad \left| \widetilde{R}_{i,l} \left(\left(j \;\mapsto\; z_{j-1} - \mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot \widetilde{\mathbf{x}} \right) , \left(j \;\mapsto\; -\mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot [X_\ell - X_{\ell,k+\ell}]_\ell \right) \right) \right| \\
&\quad \lesssim\; u_n^{-\gamma} \cdot ((s_2 - s_1) \wedge 1) \cdot \left(\|[X_\ell - X_{\ell,k+\ell}]_\ell\|_1 \;+\; \|[X_\ell - X_{\ell,k+\ell}]_\ell\|_1^{1+\gamma} \right)
\end{aligned}$$

(iii)

$$\begin{aligned}
&\int_{\mathbb{R}^T} \mathrm{d}\mathbf{P}_{[X_\ell - X_{\ell,k-1+\ell}]_\ell}(\widetilde{\mathbf{x}}) \int_{\mathbb{R}^T} \mathrm{d}\mathbf{P}_{[X_{\ell,k-1+\ell} - X_{\ell,\ell}]_\ell}(\widetilde{\mathbf{y}}) \int_{\mathbb{R}^T} G_n(\mathbf{A}_T \cdot \mathbf{z}, s_1, s_2, A)\,\mathrm{d}\mathbf{z} \\
&\qquad \left| \widetilde{R}_{i,l} \left(\left(j \;\mapsto\; z_{j-1} \;-\; \mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot ([X_\ell - X_{\ell,k+\ell}]_\ell \;+\; \widetilde{\mathbf{y}}) \right) , \left(j \;\mapsto\; -\mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot \widetilde{\mathbf{x}} \right) \right) \right| \\
&\quad \lesssim\; u_n^{-\gamma} \cdot ((s_2 - s_1) \wedge 1) \sum_{j=1}^{T} \Bigg[\left(\mathbf{E}[\|[X_\ell - X_{\ell,k-1+\ell}]\|_1] + \mathbf{E}[\|[X_\ell - X_{\ell,k-1+\ell}]\|_1^{1+\gamma}] \right) \\
&\qquad \times \left(1 + |X_{j-1} - X_{j-1,k+j-1}|^{1+\gamma} \right) \Bigg]
\end{aligned}$$

Throughout (i)-(iii), the multiplicative constant in $\lesssim$ depends only on T, (a_i), $(\mathfrak{a}_j)$, γ, $\underline{s}$, and the distribution of ε.

Proof of (i): We want to apply Lemma D.4(ii) with

$$\varphi(j) \;=\; -\mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot ([X_\ell - X_{\ell,k+\ell}]_\ell + \widetilde{\mathbf{x}}) \qquad \text{and} \qquad \psi(j) \;=\; -\mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot [a_{k+\ell}]_\ell\,.$$

Note that

$$\sum_{m=1}^{j} a_{j-m} \mathbf{e}_m^\top \;=\; \mathbf{e}_j^\top \cdot \mathbf{A}_T\,,$$

so that

$$\sum_{m=1}^{j} a_{j-m} \varphi(m) \;=\; -([X_{j-1} - X_{j-1,k+j-1}] + \widetilde{x}_{j-1})\,.$$

Also

$$\left| \mathbf{e}_l^\top \cdot \mathbf{A}_T^{-1} \cdot [a_{k+\ell}]_\ell \right| \;\lesssim\; k^{-(1-d)}\,,$$

and for $I \subset \{1, \ldots, j\}$,

$$\left| \sum_{m \in I} a_{j-m} \psi(m) \right| \;\lesssim\; \max_{m \in I} |\psi(m)| \;\lesssim\; \|[a_{k+\ell}]_\ell\|_\infty \;\lesssim\; k^{-(1-d)},$$

where the multiplicative constant in $\lesssim$ depends only on T, and the coefficients (a_i). Therefore, by Lemma D.4(ii),

$$\begin{aligned}
&\int_{\varepsilon_{-k} \wedge z}^{\varepsilon_{-k} \vee z} \mathrm{d}s \int_{\mathbb{R}^T} G_n(\mathbf{A}_T \cdot \mathbf{z}, s_1, s_2, A)\,\mathrm{d}\mathbf{z} \\
&\left| \widetilde{R}_{i,l} \left(j \mapsto z_{j-1} - \mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot ([X_\ell - X_{\ell,k+\ell}]_\ell + \widetilde{\mathbf{x}}) \,, j \mapsto -s \cdot \mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot [a_{k+\ell}]_\ell \right) \right| \\
&\lesssim\; u_n^{-\gamma} \cdot ((s_2 - s_1) \wedge 1) \\
&\quad \times \sum_{j=1}^{T} \sum_{I \subset \{1,\ldots,j\}} \left(\left| \sum_{m \in I} a_{j-m} \psi(m) \right|^\gamma \;+\; |\mathbf{e}_l^\top \cdot \mathbf{A}_T^{-1} \cdot [a_{k+\ell}]_\ell|^\gamma \right) \\
&\quad \times \left(1 \;+\; |[X_{j-1} - X_{j-1,k+j-1}] + \widetilde{x}_{j-1}|^{1+\gamma} \right) \left(|\varepsilon_{-k}|^{1+\gamma} + |z|^{1+\gamma} \right)\,.
\end{aligned} \tag{32}$$

where the multiplicative constant in $\lesssim$ depends only on T, (a_i), $(\mathfrak{a}_j)$, γ, $\underline{s}$, and the distribution of ε. Note that

$$\left|\sum_{m\in I} a_{j-m}\psi(m)\right|^{\gamma} \ + \ |\mathbf{e}_l^\top \cdot \mathbf{A}_T^{-1} \cdot [a_{k+\ell}]_\ell|^\gamma \ \lesssim \ k^{-\gamma(1-d)},$$

and

$$1 \ + \ |[X_{j-1} - X_{j-1,k+j-1}] + \widetilde{x}_{j-1}|^{1+\gamma} \ \lesssim \ 1 \ + \ |X_{j-1} - X_{j-1,k+j-1}|^{1+\gamma} \ + \ |\widetilde{x}_{j-1}|^{1+\gamma},$$

where the multiplicative constant in $\lesssim$ depends only on T, (a_i), and γ. The bound in (32) expands to

$$u_n^{-\gamma} \cdot ((s_2 - s_1) \wedge 1) \cdot k^{-\gamma(1-d)} \cdot \left(|\varepsilon_{-k}|^{1+\gamma} + |z|^{1+\gamma}\right) \cdot \sum_{j=1}^{T} \left(1 \ + \ |X_{j-1} - X_{j-1,k+j-1}|^{1+\gamma} \ + \ |\widetilde{x}_{j-1}|^{1+\gamma}\right),$$

up to a multiplicative constant that depends only on T, (a_i), γ. Note that, by $1+\gamma < \alpha$,

$$\begin{aligned}
&\sum_{j=1}^{T} \int_{\mathbb{R}} \mathrm{d}\mathbf{P}_\varepsilon(z) \int_{\mathbb{R}^T} \mathrm{d}\mathbf{P}_{[X_{\ell,k-1+\ell} - X_{\ell,\ell}]_\ell}(\widetilde{\mathbf{x}}) \left(1 \ + \ |X_{j-1} - X_{j-1,k+j-1}|^{1+\gamma} \ + \ |\widetilde{x}_{j-1}|^{1+\gamma}\right) \left(|\varepsilon_{-k}|^{1+\gamma} + |z|^{1+\gamma}\right) \\
&\lesssim \ \sum_{j=1}^{T} \left(1 + |X_{j-1} - X_{j-1,k+j-1}|^{1+\gamma}\right) \left(1 + |\varepsilon_{-k}|^{1+\gamma}\right),
\end{aligned}$$

where the multiplicative constant in $\lesssim$ depends only on (a_i), γ, and the distribution of ε. The result follows.

Proof of (ii): We want to apply Lemma D.4(i) with

$$\varphi(m) \ = \ -\mathbf{e}_m^\top \cdot \mathbf{A}_T^{-1} \cdot \widetilde{\mathbf{x}} \qquad \text{and} \qquad \psi(m) \ = \ -\mathbf{e}_m^\top \cdot \mathbf{A}_T^{-1} \cdot [X_\ell - X_{\ell,k+\ell}]_\ell,$$

so that

$$\sum_{m=1}^{j} a_{j-m}\varphi(m) \ = \ -\widetilde{x}_{j-1}.$$

By Lemma D.4(i),

$$\begin{aligned}
&\int_{\mathbb{R}^T} G_n(\mathbf{A}_T \cdot \mathbf{z}, s_1, s_2, A) \left|\widetilde{R}_{i,l}\left(\left(j \ \mapsto \ z_{j-1} - \mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot \widetilde{\mathbf{x}}\,, j \ \mapsto \ -\mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot [X_\ell - X_{\ell,k+\ell}]_\ell\right)\right)\right| \mathrm{d}\mathbf{z} \\
&\lesssim \ u_n^{-\gamma} \cdot ((s_2 - s_1) \wedge 1) \sum_{j=1}^{T} \sum_{I \subset \{1,\dots,j\}} \left(1 + |\widetilde{x}_{j-1}|^{1+\gamma}\right) \left(|\mathbf{e}_l^\top \mathbf{A}_T^{-1} \cdot [X_\ell - X_{\ell,k+\ell}]_\ell| + \left|\sum_{m\in I} a_{j-m}\psi(m)\right|^{1+\gamma}\right),
\end{aligned}$$

where the multiplicative constant in $\lesssim$ depends only on T, (a_i), $(\mathfrak{a}_j)$, γ, $\underline{s}$, and the distribution of ε. Using

$$|\mathbf{e}_l^\top \mathbf{A}_T^{-1} \cdot [X_\ell - X_{\ell,k+\ell}]_\ell| \ \lesssim \ \|[X_\ell - X_{\ell,k+\ell}]_\ell\|_1,$$

one finds, by Lemma A.5,

$$\left|\sum_{m\in I} a_{j-m}\psi(m)\right| \ \lesssim \ \max_{m\in I} |\psi(m)| \ \lesssim \ \|[X_\ell - X_{\ell,k+\ell}]_\ell\|_1 \ \in \ L^{1+\gamma}(\mathbf{P}),$$

and

$$\int_{\mathbb{R}^T} \left(1 + |\widetilde{x}_{j-1}|^{1+\gamma}\right) \mathrm{d}\mathbf{P}_{[X_\ell - X_{\ell,\ell}]_\ell}(\widetilde{\mathbf{x}}) \; = \; 1 + \mathbf{E}[|X_{j-1} - X_{j-1,j-1}|^{1+\gamma}] \; < \; \infty \, .$$

This yields the result.

Proof of (iii): We want to apply Lemma D.4(i) with

$$\varphi(j) \; = \; -\mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot (\widetilde{\mathbf{y}} \; + \; [X_\ell - X_{\ell,k+\ell}]_\ell) \qquad \text{and} \qquad \psi(j) \; = \; -\mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot \widetilde{\mathbf{x}} \, ,$$

so that

$$\sum_{m=1}^{j} a_{j-m} \varphi(m) \; = \; -(\widetilde{y}_{j-1} + (X_{j-1} - X_{j-1,k+j-1})) \, .$$

By Lemma D.4(i),

$$\int_{\mathbb{R}^T} G_n(\mathbf{A}_T \cdot \mathbf{z}, s_1, s_2, A) \left| \widetilde{R}_{i,l} \left(\left(j \; \mapsto \; z_{j-1} - \mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot (\widetilde{\mathbf{y}} + [X_\ell - X_{\ell,k+\ell}]_\ell) \; , j \; \mapsto \; -\mathbf{e}_j^\top \cdot \mathbf{A}_T^{-1} \cdot \widetilde{\mathbf{x}} \right) \right) \right| \mathrm{d}\mathbf{z}$$

$$\lesssim \; u_n^{-\gamma} \cdot ((s_2 - s_1) \wedge 1) \sum_{j=1}^{T} \sum_{I \subset \{1,\ldots,j\}} \left(1 + |\widetilde{y}_{j-1}|^{1+\gamma} + |X_{j-1} - X_{j-1,k+j-1}|^{1+\gamma}\right)$$

$$\times \; \left(|\mathbf{e}_l^\top \cdot \mathbf{A}_T^{-1} \cdot \widetilde{\mathbf{x}}| + \left| \sum_{m \in I} a_{j-m} \psi(m) \right|^{1+\gamma} \right) \, ,$$

where the multiplicative constant in $\lesssim$ depends only on T, (a_i), $(\mathfrak{a}_j)$, γ, $\underline{s}$, and the distribution of ε. Using

$$|\mathbf{e}_l^\top \cdot \mathbf{A}_T^{-1} \cdot \widetilde{\mathbf{x}}| \; \lesssim \; \|\widetilde{\mathbf{x}}\|_1 \, ,$$

$$\left| \sum_{m \in I} a_{j-m} \psi(m) \right| \; \lesssim \; \max_{m \in I} |\psi(m)| \; \lesssim \; \|\widetilde{\mathbf{x}}\|_1$$

and, by Lemma A.5,

$$\int_{\mathbb{R}^T} \left(\|\widetilde{\mathbf{x}}\|_1 + \|\widetilde{\mathbf{x}}\|_1^{1+\gamma} \right) \mathrm{d}\mathbf{P}_{[X_\ell - X_{\ell,k-1+\ell}]}(\widetilde{\mathbf{x}}) \; = \; \mathbf{E}[\|[X_\ell - X_{\ell,k-1+\ell}]\|_1] + \mathbf{E}[\|[X_\ell - X_{\ell,k-1+\ell}]\|_1^{1+\gamma}] \; < \; \infty \, ,$$

$$\int_{\mathbb{R}^T} (1 + |\widetilde{y}_{j-1}|^{1+\gamma}) \, \mathrm{d}\mathbf{P}_{[X_{\ell,k-1+\ell} - X_{\ell,\ell}]}(\widetilde{\mathbf{y}}) \; = \; 1 + \mathbf{E}[|X_{j-1,k-1+j-1} - X_{j-1,j-1}|^{1+\gamma}] \; < \; \infty$$

yields the result. □

This is the main result of the subsection, where we complete the analysis of the remainder terms R_j by taking the $L^r(\mathbf{P})$-norm of the intermediate bounds provided so far.

Lemma E.8 (L^r-bounds of the remainder terms R_j)**.** *Let Assumptions 1 and 2 hold, and let* $\gamma \in (0, \alpha - 1)$ *and* $r \in (1, 2)$ *satisfy*

$$\gamma \; < \; \min\left\{ \frac{d}{1-d} \, , 1 - \frac{1}{r(1-d)} \, , \frac{\alpha}{r} - 1 \right\} \qquad \textit{and} \qquad \frac{1}{1-d} \; < \; r \; < \; \alpha \, .$$

Then, for all $j = 1, 2, 3$,

$$\|R_j\|_{L^r(\mathbf{P})} \; \lesssim \; u_n^{-\gamma} \cdot ((s_2 - s_1) \wedge 1) \cdot k^{-(1+\gamma)(1-d)} \, ,$$

where the multiplicative constant in $\lesssim$ *depends only on* T, (a_i), $(\mathfrak{a}_j)$, r, γ, $\underline{s}$, *and the distribution of* ε.

Proof. We use first Lemmas E.6 and E.7 and treat each R_j separately.
Analysis of R_1: It suffices to bound the $L^r(\mathbf{P})$ norm of

$$\left(1+|\varepsilon_{-k}|^{1+\gamma}\right)\cdot\Big(1 \;+\; |X_{j-1}-X_{j-1,k+j-1}|^{1+\gamma}\Big)$$

for each j. The two factors have a finite r th moment, due to the assumption $(1+\gamma)r<\alpha$ and Lemma A.5, and they are independent, so their product has a finite r th moment as well.
Analysis of R_2: First use the independence property and the fact that ε_{-k} has a finite $L^r(\mathbf{P})$ norm, to find that it is enough to show that

$$k^{-(1-d)}\Big(\|[X_\ell-X_{\ell,k+\ell}]_\ell\|_1 \;+\; \|[X_\ell-X_{\ell,k+\ell}]_\ell\|_1^{1+\gamma}\Big)$$

is bounded in $L^r(\mathbf{P})$ by a multiple of $k^{-(1+\gamma)(1-d)}$. By stationarity of (X_t) and Lemma A.5,

$$\big\|\|[X_\ell-X_{\ell,k+\ell}]_\ell\|_1\big\|_{L^r(\mathbf{P})} \;\lesssim\; \sum_{\ell=0}^{T-1}\|X_\ell-X_{\ell,k+\ell}\|_{L^r(\mathbf{P})} \;\lesssim\; \|X_0-X_{0,k}\|_{L^r(\mathbf{P})} \;\lesssim\; k^{1/r-(1-d)},$$

where the multiplicative constant in $\lesssim$ depends only on T, (a_i), and the distribution of ε. By convexity of $x\mapsto x^{1+\gamma}$,

$$\|[X_\ell-X_{\ell,k+\ell}]_\ell\|_1^{1+\gamma} \;\lesssim\; \sum_{\ell=0}^{T-1}|X_\ell-X_{\ell,k+\ell}|^{1+\gamma},$$

and therefore, by Lemma A.5 again,

$$\Big\|\|[X_\ell-X_{\ell,k+\ell}]_\ell\|_1^{1+\gamma}\Big\|_{L^r(\mathbf{P})} \;\lesssim\; \sum_{\ell=0}^{T-1}\Big\||X_\ell-X_{\ell,k+\ell}|^{1+\gamma}\Big\|_{L^r(\mathbf{P})} \;\lesssim\; \big\||X_0-X_{0,k}|^{1+\gamma}\big\|_{L^r(\mathbf{P})} \;\lesssim\; k^{1/r-(1+\gamma)(1-d)},$$

where we used that by assumption $1/(1-d)<r<(1+\gamma)r<\alpha$. Here the multiplicative constant in $\lesssim$ depends only on T, (a_i), γ, and the distribution of ε. Consequently

$$k^{-(1-d)}\,\big\|\|[X_\ell-X_{\ell,k+\ell}]_\ell\|_1\big\|_{L^r(\mathbf{P})} \;+\; \Big\|\|[X_\ell-X_{\ell,k+\ell}]_\ell\|_1^{1+\gamma}\Big\|_{L^r(\mathbf{P})} \;\lesssim\; k^{1/r-2(1-d)}.$$

Now, since $\gamma<1-1/(r(1-d))$ by assumption,

$$\frac{1}{r} \;-\; 2(1-d) \;=\; -(1-d)\left(1-\frac{1}{r(1-d)}\right) \;-\; (1-d) \;\le\; -(1-d)\gamma \;-\; (1-d) \;=\; -(1-d)(1+\gamma),$$

so that indeed

$$k^{-(1-d)}\,\big\|\|[X_\ell-X_{\ell,k+\ell}]_\ell\|_1\big\|_{L^r(\mathbf{P})} \;+\; \Big\|\|[X_\ell-X_{\ell,k+\ell}]_\ell\|_1^{1+\gamma}\Big\|_{L^r(\mathbf{P})} \;\lesssim\; k^{-(1+\gamma)(1-d)},$$

where the multiplicative constant in $\lesssim$ depends only on T, (a_i), γ, and the distribution of ε.
Analysis of R_3: Again use the independence property and the fact that ε_{-k} has a finite $L^r(\mathbf{P})$ norm, to find that

$$\big\||\varepsilon_{-k}|\cdot\left(1+|X_{j-1}-X_{j-1,k+j-1}|^{1+\gamma}\right)\big\|_{L^r(\mathbf{P})} \;<\; \infty$$

due to $(1+\gamma)r<\alpha$. Then note that by convexity,

$$\begin{aligned}
&\mathbf{E}[\|[X_\ell-X_{\ell,k-1+\ell}]_\ell\|_1] \;+\; \mathbf{E}[\|[X_\ell-X_{\ell,k-1+\ell}]_\ell\|_1^{1+\gamma}]\\
&\le\; \big\|\|[X_\ell-X_{\ell,k-1+\ell}]_\ell\|_1\big\|_{L^r(\mathbf{P})} \;+\; \Big\|\|[X_\ell-X_{\ell,k-1+\ell}]_\ell\|_1^{1+\gamma}\Big\|_{L^r(\mathbf{P})}\\
&\lesssim\; k^{1/r-(1-d)},
\end{aligned}$$

and the proof finishes as in (ii). $\square$

E.3. Proof of Lemma E.1

We are going to apply Lemmas E.4 (moment bounds after martingale decomposition), E.5 (error decomposition), E.8 (L^r -bounds of the remainder terms R_j), and Lemma B.4 in Scheffel et al. (2026), which we are going to recall now.

Lemma E.9 (Rate of divergence of convolution-type sum; see Lemma B.4 in Scheffel et al. (2026))**.** *Let Assumptions 1 and 2 hold, and let $\gamma \in (0, \alpha - 1)$ and $r \in (1, 2)$ satisfy*

$$\gamma \;<\; \min\left\{ \frac{d}{1-d}\,, 1 - \frac{1}{r(1-d)}\,, \frac{\alpha}{r} - 1 \right\} \qquad \textit{and} \qquad \frac{1}{1-d} \;<\; r \;<\; \alpha\,.$$

Then

$$\sum_{k=-\infty}^{n} \left(\sum_{t=k\vee 1}^{n} \left(1 \wedge (t-k)^{-(1-d)(1+\gamma)} \right) \right)^r \;\lesssim\; n^{r+1-(1-d)(1+\gamma)r}\,,$$

where $\lesssim$ denotes inequality up to a multiplicative constant that depends only on γ , α , and d . We adopt the convention $1/0 = \infty$.

Proof of Lemma E.1: By Lemma E.4,

$$\mathbf{E}\, |\mathcal{S}_n(s_1, s_2, A)|^r \;\lesssim\; \sum_{k=-\infty}^{n} \left(\sum_{t=(k-T+1)\vee 1}^{n-T+1} \| P_{k-t}\, \mathcal{U}_n([X_\ell]_\ell, s_1, s_2, A) \|_{L^r(\mathbf{P})} \right)^r ,$$

where the multiplicative constant in $\lesssim$ is independent of all other quantities in this work. We split the inner sum based on whether $t - k \geq 1$ and use the convexity of $x \mapsto x^r$ (due to $r > 1$) to get

$$\begin{aligned} \mathbf{E}\, |\mathcal{S}_n(s_1, s_2, A)|^r \;\lesssim\; &\sum_{k=-\infty}^{n} \left(\sum_{\substack{((k-T+1)\vee 1) \leq t \leq n-T+1 \\ t-k \geq 1}} \| P_{k-t}\, \mathcal{U}_n([X_\ell]_\ell, s_1, s_2, A) \|_{L^r(\mathbf{P})} \right)^r \\ &+ \sum_{k=-\infty}^{n} \left(\sum_{\substack{((k-T+1)\vee 1) \leq t \leq n-T+1 \\ t-k < 1}} \| P_{k-t}\, \mathcal{U}_n([X_\ell]_\ell, s_1, s_2, A) \|_{L^r(\mathbf{P})} \right)^r . \end{aligned}$$

Case $t - k \geq 1$: We apply Lemmas E.5, E.8, and E.9. This yields

$$\begin{aligned} &\sum_{k=-\infty}^{n} \left(\sum_{\substack{((k-T+1)\vee 1) \leq t \leq n-T+1 \\ t-k \geq 1}} \| P_{k-t}\, \mathcal{U}_n([X_\ell]_\ell, s_1, s_2, A) \|_{L^r(\mathbf{P})} \right)^r \\ &\lesssim\; \left(u_n^{-\gamma} \cdot ((s_2 - s_1) \wedge 1) \right)^r \sum_{k=-\infty}^{n} \left(\sum_{t=k\vee 1}^{n} \left(1 \wedge (t-k)^{-(1-d)(1+\gamma)} \right) \right)^r \\ &\lesssim\; ((s_2 - s_1) \wedge 1) \cdot n^{r+1-(1-d)(1+\gamma)r}\,, \end{aligned}$$

since $r > 1$. Here the multiplicative constant in $\lesssim$ depends only on T , (a_i) , $(\mathfrak{a}_j)$, γ , $\underline{s}$, and the distribution of ε .

Case $t-k<1$: Note that the constraints $t<k+1$ and $t\geq((k-T+1)\vee 1)$ ensure that the inner sum is empty when $k\leq 0$. By Lemma E.2,

$$\sum_{k=-\infty}^{n}\left(\sum_{\substack{((k-T+1)\vee 1)\leq t\leq n-T+1\\ t-k<1}}\|P_{k-t}\,\mathcal{U}_n([X_\ell]_\ell,s_1,s_2,A)\|_{L^r(\mathbf{P})}\right)^r$$
$$\lesssim\ \Big(\|\mathcal{U}_n([X_\ell]_\ell,s_1,s_2,A)\|_{L^r(\mathbf{P})}\Big)^r\sum_{k=1}^{n}\left(\#\{t\in\mathbb{N}:\ ((k-T+1)\vee 1)\leq t\leq k+1\}\right)^r$$
$$\lesssim\ ((s_2-s_1)\wedge 1)\cdot n\cdot T^r\ \lesssim\ n\cdot((s_2-s_1)\wedge 1)\,,$$

where the multiplicative constant in $\lesssim$ depends only on T, (a_i), $(\mathfrak{a}_j)$, r, γ, $\underline{s}$, and the distribution of ε. We conclude that

$$\mathbf{E}\,|\mathcal{S}_n(s_1,s_2,A)|^r\ \lesssim\ ((s_2-s_1)\wedge 1)\cdot\Big(n^{r+1-(1-d)(1+\gamma)r}\ +\ n\Big)\,.$$

Observe that $1-(1-d)(1+\gamma)\geq 0$ due to $\gamma\leq d/(1-d)$. Therefore

$$((s_2-s_1)\wedge 1)\cdot\Big(n^{r+1-(1-d)(1+\gamma)r}\ +\ n\Big)\ \leq\ 2\cdot((s_2-s_1)\wedge 1)\cdot n^{r+1-(1-d)(1+\gamma)r}\,.$$

This finishes the proof. □

F. Proof of Theorem 3.4

We adapt the chaining argument in the proof of Theorem 2.1 in Koul and Surgailis (2001) to our setting. Choose $\delta>0$ sufficiently small and set $u_n=o(n^{(d+1/(\nu\wedge 2)-\kappa_0-\delta)/(\nu+1)})$, where κ_0 is defined in (3). We define, for $k\in\mathbb{N}_0$ and $j\in\{0,\dots,2^k\}$ the partition

$$\pi_{j,k}\ =\ \underline{s}\ +\ \frac{2^k-j}{2^k}\,(\overline{s}-\underline{s})\ ,$$

so that

$$\underline{s}\ =\ \pi_{2^k,k}\ <\ \pi_{2^k-1,k}\ <\ \cdots\ <\ \pi_{1,k}\ <\ \pi_{0,k}\ =\ \overline{s}\,.$$

For $k\in\mathbb{N}_0$ and $s\in(\underline{s},\overline{s}]$, we define $j_k^s\in\{0,\dots,2^k-1\}$ to be the unique index such that

$$\pi_{j_k^s+1,k}\ <\ s\ \leq\ \pi_{j_k^s,k}\,.$$

Note that the partitions are nested with increasing k so that one has $s\ \leq\ \pi_{j_{k+1}^s,k+1}\ \leq\ \pi_{j_k^s,k}$. We define a chain linking $\overline{s}$ to each point $s\in(\underline{s},\overline{s}]$ by

$$\pi_{j_K^s+1,K}\ <\ s\ \leq\ \pi_{j_K^s,K}\ \leq\ \cdots\ \leq\ \pi_{j_1^s,1}\ \leq\ \pi_{j_0^s,0}\ =\ \overline{s}\,,$$

where we choose $K=K_n=\lfloor\log_2(n^{2-d-1/\alpha}/f_{X_0}(u_n))\rfloor$, so that $\limsup_{n\to\infty}2^{-K_n}\frac{n^{2-d-1/\alpha}}{f_{X_0}(u_n)}<\infty$. By the assumption on u_n and regular variation of f_{X_0}, we have that $f_{X_0}(u_n)$ decays at most polynomially, so that, by $2-d-1/\alpha\in(1,2)$, $K_n\to\infty$ logarithmically. By construction, each link in the chain comes from a different partition. Since $(\pi_{j,k})\subset(\pi_{j,k+1})$, the next result identifies the possible coordinates of adjacent links of the chain in terms of the larger partition.

Lemma F.1 (Partition coordinates of adjacent chain links)**.** *For all $k \in \mathbb{N}$ and all $s \in (\underline{s}, \overline{s}]$ it holds* $\pi_{j^s_{k+1},k+1}, \pi_{j^s_k,k} \in \{\pi_{2j^s_k,k+1}, \pi_{2j^s_k+1,k+1}\}$.

Proof. By definition,

$$\pi_{j,k} \;=\; \underline{s} \;+\; \frac{2^k - j}{2^k}\,(\overline{s} - \underline{s}) \;=\; \underline{s} \;+\; \frac{2^{k+1} - 2j}{2^{k+1}}\,(\overline{s} - \underline{s}) \;=\; \pi_{2j,k+1}\,,$$

so that

$$\pi_{2(j^s_k+1),k+1} \;=\; \pi_{j^s_k+1,k} < s \;\le\; \pi_{j^s_k,k} \;=\; \pi_{2j^s_k,k+1}\,.$$

Since $\pi_{2(j^s_k+1),k+1} < s \;\le\; \pi_{2j^s_k,k+1}$, it either holds

$$\pi_{2(j^s_k+1),k+1} \;<\; s \;\le\; \pi_{2j^s_k+1,k+1} \;<\; \pi_{2j^s_k,k+1} \qquad \text{or} \qquad \pi_{2(j^s_k+1),k+1} \;<\; \pi_{2j^s_k+1,k+1} \;<\; s \;\le\; \pi_{2j^s_k,k+1}\,,$$

and therefore $j^s_{k+1} \in \{2j^s_k, 2j^s_k + 1\}$. Thus $\pi_{j^s_{k+1},k+1}, \pi_{j^s_k,k} \in \{\pi_{2j^s_k,k+1}, \pi_{2j^s_k+1,k+1}\}$. □

Fix $A \in \mathcal{C}_T$ and recall the notation

$$\begin{aligned}\mathcal{S}_n(s_1, s_2, A) \;=\; \sum_{t=1}^{n-T+1} \Bigg(& G_n([X_{t+\ell}]_\ell, s_1, s_2, A) \;-\; \mathbf{E}[G_n([X_{t+\ell}]_\ell, s_1, s_2, A)] \\ & -\; \bigg(\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=\mathbf{0}_T}\bigg)^{\top} \cdot [X_{t+\ell}]_\ell \Bigg)\end{aligned}$$

from Lemma E.4. To streamline the argument, we will write throughout,

$$\mathcal{S}_n(s_1, s_2) \;:=\; \mathcal{S}_n(s_1, s_2, A)\,.$$

Note that $\mathcal{S}_n(s_1, s_1) = 0$ since $(s_1 \cdot A) \setminus (s_1 \cdot A) = \emptyset$.

Lemma F.2 (Decomposition of $\mathcal{S}_n$)**.** *Let Assumption 1 and 2 hold. Then, for $s_1, s_2, s_3 \in (\underline{s}, \overline{s}] \cup \{\infty\}$ satisfying $s_1 \le s_2 \le s_3$,*

$$\mathcal{S}_n(s_1, s_3) \;=\; \mathcal{S}_n(s_1, s_2) \;+\; \mathcal{S}_n(s_2, s_3)\,.$$

In particular, for all $s \in (\underline{s}, \overline{s}]$,

$$\mathcal{S}_n(s, \overline{s}) \;=\; \sum_{k=1}^{K} \mathcal{S}_n(\pi_{j^s_k,k}, \pi_{j^s_{k-1},k-1}) \;+\; \mathcal{S}_n(s, \pi_{j^s_K,K})\,.$$

Proof. If $s_j = \infty$ for some $j \in \{1, 2, 3\}$, there is nothing to prove. Therefore, we assume that $s_j < \infty$ for all $j \in \{1, 2, 3\}$. Fix $s_1, s_2, s_3 \in (\underline{s}, \infty)$ satisfying $s_1 \le s_2 \le s_3$. Since $A \in \mathcal{C}_T$ it holds $s_3 \cdot A \subset s_2 \cdot A \subset s_1 \cdot A$, so that

$$\begin{aligned}& G_n([X_{t+\ell}]_\ell, s_1, s_3, A) \;=\; \mathbb{1}\{[X_{t+\ell}]_\ell \in u_n((s_1 \cdot A) \setminus (s_3 \cdot A))\} \\ &\;=\; \mathbb{1}\{[X_{t+\ell}]_\ell \in u_n((s_1 \cdot A) \setminus (s_2 \cdot A))\} \;+\; \mathbb{1}\{[X_{t+\ell}]_\ell \in u_n((s_2 \cdot A) \setminus (s_3 \cdot A))\} \\ &\;=\; G_n([X_{t+\ell}]_\ell, s_1, s_2, A) \;+\; G_n([X_{t+\ell}]_\ell, s_2, s_3, A)\,,\end{aligned}$$

and

$$\begin{aligned}
\mathbf{E}[G_n([X_\ell]_\ell, s_1, s_3, A)] &= \mathbf{P}[[X_\ell]_\ell \in u_n((s_1 \cdot A) \setminus (s_3 \cdot A))] \\
&= \mathbf{P}[[X_\ell]_\ell \in u_n((s_1 \cdot A) \setminus (s_2 \cdot A))] + \mathbf{P}[[X_\ell]_\ell \in u_n((s_2 \cdot A) \setminus (s_3 \cdot A))] \\
&= \mathbf{E}[G_n([X_\ell]_\ell, s_1, s_2, A)] + \mathbf{E}[G_n([X_\ell]_\ell, s_2, s_3, A)],
\end{aligned}$$

and, by Lemma B.3,

$$\begin{aligned}
\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s_1, s_3, A)\Big|_{\mathbf{y}=\mathbf{0}_T} &= -\int_{\mathbb{R}^T} G_n(\mathbf{x}, s_1, s_3, A) \nabla f_{[X_\ell]_\ell}(\mathbf{x})\,\mathrm{d}\mathbf{x} \\
&= -\int_{\mathbb{R}^T} \left(G_n(\mathbf{x}, s_1, s_2, A) + G_n(\mathbf{x}, s_2, s_3, A)\right) \nabla f_{[X_\ell]_\ell}(\mathbf{x})\,\mathrm{d}\mathbf{x} \\
&= \nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s_1, s_2, A)\Big|_{\mathbf{y}=\mathbf{0}_T} + \nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s_2, s_3, A)\Big|_{\mathbf{y}=\mathbf{0}_T}.
\end{aligned}$$

The result now follows immediately. □

Now we perform the approximation step outlined in the heuristic argument before Theorem 3.4.

Lemma F.3. *Let the assumptions of Theorem 3.2 hold. Then, for $n \in \mathbb{N}$, the following holds:*

(i) For all $s \in (\underline{s}, \overline{s})$,

$$\left|\mathcal{S}_n(s, \pi_{j_K^s,K})\right| \lesssim \left|\mathcal{S}_n(\pi_{j_K^s+1,K}, \pi_{j_K^s,K})\right| + B_n,$$

where

$$B_n := 2^{-K} \cdot \sum_{t=1}^{n-T+1} (1 + \|[X_{t+\ell}]_\ell\|_1).$$

(ii) For all $s \in (\underline{s}, \overline{s})$,

$$\mathbf{E}[|B_n|] \lesssim 2^{-K} \cdot n$$

and for all $k \le K$ and all $\delta > 0$

$$\mathbf{E}^*\left[\left(\sup_{s \in (\underline{s}, \overline{s})} \left|\mathcal{S}_n\left(\pi_{j_k^s,k}, \pi_{j_{k-1}^s,k-1}\right)\right|\right)^{r_0}\right] \vee \mathbf{E}^*\left[\left(\sup_{s \in (\underline{s}, \overline{s})} \left|\mathcal{S}_n\left(\pi_{j_K^s+1,K}, \pi_{j_K^s,K}\right)\right|\right)^{r_0}\right] \lesssim n^{r_0(\kappa_0+\delta/2)},$$

where κ_0 and r_0 are defined in (3).

In both (i) and (ii), $\lesssim$ means inequality up to a multiplicative constant that depends on δ, A, T, $\underline{s}$, κ_0, the coefficients (a_i), and the distribution of ε, and is independent of k, K, and n.

Proof of (i): By Lemma F.2 and the definition of $\mathcal{S}_n$,

$$\begin{aligned}
&\left|\mathcal{S}_n(s, \pi_{j_K^s,K}) - \mathcal{S}_n(\pi_{j_K^s+1,K}, \pi_{j_K^s,K})\right| = \left|\mathcal{S}_n(\pi_{j_K^s+1,K}, s)\right| \\
&\le \sum_{t=1}^{n-T+1} \mathbb{1}\{[X_{t+\ell}]_\ell \in u_n \cdot ((\pi_{j_K^s+1,K} \cdot A) \setminus (s \cdot A))\} + \mathbf{P}[[X_{t+\ell}]_\ell \in u_n \cdot ((\pi_{j_K^s+1,K} \cdot A) \setminus (s \cdot A))] \\
&\quad + \left\|\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, \pi_{j_K^s+1,K}, s, A)\Big|_{\mathbf{y}=\mathbf{0}_T}\right\|_\infty \|[X_{t+\ell}]_\ell\|_1.
\end{aligned}$$

Since $s \le \pi_{j_K^s,K}$ by definition, and therefore $\big((\pi_{j_K^s+1,K} \cdot A) \setminus (s \cdot A)\big) \subset \big((\pi_{j_K^s+1,K} \cdot A) \setminus (\pi_{j_K^s,K} \cdot A)\big)$, this is bounded above by

$$
\begin{aligned}
&\sum_{t=1}^{n-T+1} \mathbb{1}\{[X_{t+\ell}]_\ell \in u_n \cdot ((\pi_{j_K^s+1,K} \cdot A) \setminus (\pi_{j_K^s,K} \cdot A))\} \;+\; \mathbf{P}[[X_{t+\ell}]_\ell \in u_n \cdot ((\pi_{j_K^s+1,K} \cdot A) \setminus (s \cdot A))] \\
&\qquad + \left\| \nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, \pi_{j_K^s+1,K}, s, A)\Big|_{\mathbf{y}=\mathbf{0}_T} \right\|_\infty \|[X_{t+\ell}]_\ell\|_1 \\
&= \mathcal{S}_n(\pi_{j_K^s+1,K}, \pi_{j_K^s,K}) \\
&\qquad + \sum_{t=1}^{n-T+1} \mathbf{P}[[X_{t+\ell}]_\ell \in u_n \cdot ((\pi_{j_K^s+1,K} \cdot A) \setminus (\pi_{j_K^s,K} \cdot A))] \;+\; \mathbf{P}[[X_{t+\ell}]_\ell \in u_n \cdot ((\pi_{j_K^s+1,K} \cdot A) \setminus (s \cdot A))] \\
&\qquad + \left\| \nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, \pi_{j_K^s+1,K}, s, A)\Big|_{\mathbf{y}=\mathbf{0}_T} \right\|_\infty \|[X_{t+\ell}]_\ell\|_1 \\
&\qquad + \left(\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, \pi_{j_K^s+1,K}, \pi_{j_K^s,K}, A)\Big|_{\mathbf{y}=\mathbf{0}_T} \right)^\top [X_{t+\ell}]_\ell
\end{aligned}
$$

which we further bound, using Lemma C.4, as

$$
\begin{aligned}
&\big|\mathcal{S}_n(\pi_{j_K^s+1,K}, \pi_{j_K^s,K})\big| \\
&\qquad + \sum_{t=1}^{n-T+1} \mathbf{P}[[X_{t+\ell}]_\ell \in u_n \cdot ((\pi_{j_K^s+1,K} \cdot A) \setminus (\pi_{j_K^s,K} \cdot A))] \;+\; \mathbf{P}[[X_{t+\ell}]_\ell \in u_n \cdot ((\pi_{j_K^s+1,K} \cdot A) \setminus (s \cdot A))] \\
&\qquad + \left(\left\| \nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, \pi_{j_K^s+1,K}, s, A)\Big|_{\mathbf{y}=\mathbf{0}_T} \right\|_\infty + \left\| \nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, \pi_{j_K^s+1,K}, \pi_{j_K^s,K}, A)\Big|_{\mathbf{y}=\mathbf{0}_T} \right\|_\infty \right) \|[X_{t+\ell}]_\ell\|_1 \\
&\lesssim \big|\mathcal{S}_n(\pi_{j_K^s+1,K}, \pi_{j_K^s,K})\big| + (\pi_{j_K^s,K} - \pi_{j_K^s+1,K}) \cdot \sum_{t=1}^{n-T+1} (1 \;+\; \|[X_{t+\ell}]_\ell\|_1) \\
&= \big|\mathcal{S}_n(\pi_{j_K^s+1,K}, \pi_{j_K^s,K})\big| \;+\; B_n\,.
\end{aligned}
$$

At the last step we used the fact that $\pi_{j,k} - \pi_{j+1,k} = 2^{-k}(\overline{s} - \underline{s})$. Here, the constant in $\lesssim$ depends only on T, $\underline{s}$, γ, the coefficients (a_i), and the distribution of ε. This dependence remains for the rest of the proof. We conclude that

$$
\begin{aligned}
\big|\mathcal{S}_n(s, \pi_{j_K^s,K})\big| \;&\le\; \big|\mathcal{S}_n(s, \pi_{j_K^s,K}) \;-\; \mathcal{S}_n(\pi_{j_K^s+1,K}, \pi_{j_K^s,K})\big| \;+\; \big|\mathcal{S}_n(\pi_{j_K^s+1,K}, \pi_{j_K^s,K})\big| \\
&\lesssim\; \big|\mathcal{S}_n(\pi_{j_K^s+1,K}, \pi_{j_K^s,K})\big| \;+\; B_n\,.
\end{aligned}
$$

This proves the claim.

Proof of (ii): By definition of B_n and stationarity of (X_t),

$$
\mathbf{E}\left[|B_n|\right] \;\le\; 2^{-K} \cdot n \cdot \mathbf{E}\left[(1 + \|[X_\ell]_\ell\|_1)\right] \;\lesssim\; 2^{-K} \cdot n\,.
$$

Since $\mathcal{S}_n\left(\pi_{i,k}, \pi_{i,k}\right) = 0$ for any i, k, and since $\pi_{j_k^s,k}, \pi_{j_{k-1}^s,k-1}$ are at most neighbors in the partition at step k, it follows from Lemma F.1 that

$$
\left(\sup_{s \in (\underline{s}, \overline{s})} \left| \mathcal{S}_n\left(\pi_{j_k^s,k}, \pi_{j_{k-1}^s,k-1}\right) \right| \right)^{r_0} \;\le\; \left(\max_{0 \le i \le 2^k - 1} \left|\mathcal{S}_n\left(\pi_{i+1,k}, \pi_{i,k}\right)\right| \right)^{r_0} \;\le\; \sum_{0 \le i \le 2^k - 1} \left|\mathcal{S}_n\left(\pi_{i+1,k}, \pi_{i,k}\right)\right|^{r_0}.
$$

We get, by Theorem 3.2, that

$$\begin{aligned}
&\mathbf{E}^*\left[\left(\sup_{s\in(\underline{s},\overline{s})}\left|\mathcal{S}_n\left(\pi_{j_k^s,k},\pi_{j_{k-1}^s,k-1}\right)\right|\right)^{r_0}\right] \;\le\; \sum_{0\le i\le 2^k-1}\mathbf{E}\left[\left|\mathcal{S}_n\left(\pi_{i+1,k},\pi_{i,k}\right)\right|^{r_0}\right]\\
&\quad\lesssim\; n^{r_0(\kappa_0+\delta/2)}\sum_{0\le i\le 2^k-1}\left(\pi_{i,k}-\pi_{i+1,k}\right) \;=\; n^{r_0(\kappa_0+\delta/2)}\cdot(\overline{s}-\underline{s}) \;\lesssim\; n^{r_0(\kappa_0+\delta/2)}\,,
\end{aligned}$$

and

$$\begin{aligned}
&\mathbf{E}^*\left[\left(\sup_{s\in(\underline{s},\overline{s})}\left|\mathcal{S}_n\left(\pi_{j_K^s+1,K},\pi_{j_K^s,K}\right)\right|\right)^{r_0}\right] \;\le\; \sum_{0\le i\le 2^K-1}\mathbf{E}\left[\left|\mathcal{S}_n\left(\pi_{i+1,K},\pi_{i,K}\right)\right|^{r_0}\right]\\
&\quad\lesssim\; n^{r_0(\kappa_0+\delta/2)}\sum_{0\le i\le 2^K-1}\left(\pi_{i,K}-\pi_{i+1,K}\right) \;\lesssim\; n^{r_0(\kappa_0+\delta/2)}\,.
\end{aligned}$$

The proof is complete. □

Proof of Theorem 3.4: Let $\delta>0$ and assume $u_n\to\infty$ is such that $u_n = o(n^{(d+1/(\nu\wedge 2)-\kappa_0-\delta)/(\nu+1)})$, where κ_0 is defined in (3). By Lemma F.2 and Lemma F.3,

$$\begin{aligned}
\sup_{s\in(\underline{s},\overline{s})}|\mathcal{S}_n(s,\overline{s})| \;&\le\; \sum_{k=1}^{K}\sup_{s\in(\underline{s},\overline{s})}\left|\mathcal{S}_n\left(\pi_{j_k^s,k},\pi_{j_{k-1}^s,k-1}\right)\right| \;+\; \sup_{s\in(\underline{s},\overline{s})}\left|\mathcal{S}_n\left(\pi_{j_K^s+1,K},\pi_{j_K^s,K}\right)\right| \;+\; |B_n|\\
&\le\; 2\left((K+1)\left(\max_{1\le k\le K}\sup_{s\in(\underline{s},\overline{s})}\left|\mathcal{S}_n\left(\pi_{j_k^s,k},\pi_{j_{k-1}^s,k-1}\right)\right| \;\vee\; \sup_{s\in(\underline{s},\overline{s})}\left|\mathcal{S}_n\left(\pi_{j_K^s+1,K},\pi_{j_K^s,K}\right)\right|\right) \;\vee\; |B_n|\right)\,,
\end{aligned}$$

so that, fixing $\widetilde{\delta}>0$ and using Markov's inequality and Lemma F.3 again,

$$\begin{aligned}
&\mathbf{P}^*\left[\sup_{s\in(\underline{s},\overline{s})}|\mathcal{S}_n(s,\overline{s})| > \widetilde{\delta}\cdot n^{d+1/\alpha}f_{X_0}(u_n)\right]\\
&\le\; \sum_{k=1}^{K}\mathbf{P}^*\left[\sup_{s\in(\underline{s},\overline{s})}\left|\mathcal{S}_n\left(\pi_{j_k^s,k},\pi_{j_{k-1}^s,k-1}\right)\right| > \frac{\widetilde{\delta}\cdot n^{d+1/\alpha}f_{X_0}(u_n)}{2(K+1)}\right]\\
&\quad+\;\mathbf{P}^*\left[\sup_{s\in(\underline{s},\overline{s})}\left|\mathcal{S}_n\left(\pi_{j_K^s+1,K},\pi_{j_K^s,K}\right)\right| > \frac{\widetilde{\delta}\cdot n^{d+1/\alpha}f_{X_0}(u_n)}{2(K+1)}\right] \;+\; \mathbf{P}\left[|B_n| > \frac{\widetilde{\delta}\cdot n^{d+1/\alpha}f_{X_0}(u_n)}{2}\right]\\
&\le\; \left(\widetilde{\delta}\cdot n^{d+1/\alpha}f_{X_0}(u_n)\right)^{-r_0}\\
&\quad\times\left(2^{r_0}(K+1)^{r_0}\left(\sum_{k=1}^{K}\mathbf{E}^*\left[\sup_{s\in(\underline{s},\overline{s})}\left|\mathcal{S}_n\left(\pi_{j_k^s,k},\pi_{j_{k-1}^s,k-1}\right)\right|^{r_0}\right] \;+\; \mathbf{E}^*\left[\sup_{s\in(\underline{s},\overline{s})}\left|\mathcal{S}_n\left(\pi_{j_K^s+1,K},\pi_{j_K^s,K}\right)\right|^{r_0}\right]\right)\right)\\
&\quad+\;2\left(\widetilde{\delta}\cdot n^{d+1/\alpha}f_{X_0}(u_n)\right)^{-1}\mathbf{E}\left[|B_n|\right]\\
&\lesssim\; \left(\widetilde{\delta}\cdot n^{d+1/\alpha}f_{X_0}(u_n)\right)^{-r_0}\cdot\left(K^{r_0+1}\cdot n^{r_0(\kappa_0+\delta/2)}\right) \;+\; \left(\widetilde{\delta}\cdot n^{d+1/\alpha}f_{X_0}(u_n)\right)^{-1}2^{-K}\cdot n\,,
\end{aligned}$$

where the constant in $\lesssim$ depends only on δ, A, T, $\underline{s}$, γ, r_0, the coefficients (a_i), and the distribution of ε, and is independent of K and n. Since $\limsup_{n\to\infty}2^{-K_n}\frac{n^{2-d-1/\alpha}}{f_{X_0}(u_n)}<\infty$,

$$\left(n^{d+1/\alpha}f_{X_0}(u_n)\right)^{-1}\cdot 2^{-K}n \;=\; \left(2^{-K}\cdot\frac{n^{2-d-1/\alpha}}{f_{X_0}(u_n)}\right)\frac{1}{n} \;\to\; 0\,, \tag{33}$$

as $n \to \infty$. Then, for $\delta > 0$ small enough and $u_n = o(n^{(d+1/(\nu\wedge 2)-\kappa_0-\delta)/(\nu+1)})$,

$$\left(n^{d+1/\alpha} f_{X_0}(u_n)\right)^{-r_0} \cdot K^{r_0+1} \cdot n^{r_0(\kappa_0+\delta/2)} \;=\; \frac{n^{r_0(\kappa_0+3\delta/4-d-1/\alpha)}}{(f_{X_0}(u_n))^{r_0}} K^{r_0+1} n^{-r_0\delta/4} \;\to\; 0\,,$$

by Potter bounds (showing that $n^{r_0(\kappa_0+3\delta/4-d-1/\alpha)}/(f_{X_0}(u_n))^{r_0} \to 0$), see Theorem 1.5.6 on p.25 in Bingham et al. (1987) and since K grows only logarithmically. Therefore

$$\frac{n^{-d-1/\alpha}}{f_{X_0}(u_n)} \sup_{s\in(\underline{s},\overline{s})} |\mathcal{S}_n(s,\overline{s})| \;\to_{\mathbf{P}^*}\; 0\,,$$

as $n \to \infty$. By Corollary 3.3, we get

$$\frac{n^{-d-1/\alpha}}{f_{X_0}(u_n)} |\mathcal{S}_n(\overline{s},\infty)| \;\to_{\mathbf{P}}\; 0\,.$$

It then follows from Lemma F.2

$$\frac{n^{-d-1/\alpha}}{f_{X_0}(u_n)} \sup_{s\in(\underline{s},\overline{s})} |\mathcal{S}_n(s,\infty)| \;\to_{\mathbf{P}^*}\; 0\,.$$

This finishes the proof. □

G. Proofs of the results in Section 4

G.1. Auxiliary results

Assumption 7 has in the denominator of the first term $f_{X_0}(s \cdot u_n)$. We now provide a different formulation with $f_{X_0}(u_n)$ in the denominator of the first term and an additional factor $s^{-1-\nu}$ in the second term, which comes from $f_{X_0}(s\cdot u_n)/f_{X_0}(u_n) \to s^{-1-\nu}$ locally uniformly as $n \to \infty$.

Lemma G.1 (Different formulation of Assumption 7)**.** *Let Assumption 3 hold, and let $A \in \mathcal{C}_T$ satisfy Assumption 7 with H_A and N_A. Then, as $n \to \infty$,*

$$\sup_{s\in N_A} \left| \frac{\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y},s,\infty,A)\Big|_{\mathbf{y}=\mathbf{0}_T}}{f_{X_0}(u_n)} \;-\; s^{-1-\nu} H_A(s) \right| \;\to\; 0\,.$$

Proof. Note that

$$\begin{aligned}
&\sup_{s\in N_A} \left| \frac{\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y},s,\infty,A)\Big|_{\mathbf{y}=\mathbf{0}_T}}{f_{X_0}(u_n)} \;-\; s^{-1-\nu} H_A(s) \right| \\
&\le \sup_{s\in N_A} \left| \frac{f_{X_0}(s\cdot u_n)}{f_{X_0}(u_n)} \right| \sup_{s\in N_A} \left| \frac{\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y},s,\infty,A)\Big|_{\mathbf{y}=\mathbf{0}_T}}{f_{X_0}(s\cdot u_n)} \;-\; H_A(s) \right| \\
&\quad + \sup_{s\in N_A} |H_A(s)| \sup_{s\in N_A} \left| \frac{f_{X_0}(s\cdot u_n)}{f_{X_0}(u_n)} \;-\; s^{-1-\nu} \right|.
\end{aligned}$$

By assumption, $N_A \subset (0,\infty)$ is compact, so that by Assumption 3 (note that regular variation is always locally uniform) and the continuity of H_A,

$$\sup_{s\in N_A} \left| \frac{f_{X_0}(s\cdot u_n)}{f_{X_0}(u_n)} \right|,\ \sup_{s\in N_A} |H_A(s)| \;<\; \infty\,,$$

and also by Assumption 7

$$\sup_{s\in N_A}\left|\frac{\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y},s,\infty,A)\Big|_{\mathbf{y}=\mathbf{0}_T}}{f_{X_0}(s\cdot u_n)} - H_A(s)\right|, \sup_{s\in N_A}\left|\frac{f_{X_0}(s\cdot u_n)}{f_{X_0}(u_n)} - s^{-1-\nu}\right| \to 0$$

as $n\to\infty$, so that

$$\sup_{s\in N_A}\left|\frac{\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y},s,\infty,A)\Big|_{\mathbf{y}=\mathbf{0}_T}}{f_{X_0}(u_n)} - s^{-1-\nu}H_A(s)\right| \to 0$$

as $n\to\infty$. $\square$

In the last result we derived a different formulation of Assumption 7. We now use this to derive an adapted uniform reduction principle from Theorem 3.4, where the coefficient vector of the linear approximation and the partial sum of the $[X_{t+\ell}]_\ell$ have already converged.

Theorem G.2 (Uniform reduction principle with weak limit in the linear approximation)**.** *Let Assumptions 1, 2, and 3 hold, assume that $A\in\mathcal{C}_T$ satisfies Assumption 7 with H_A and N_A in the sense of Assumption 7, and let $(\underline{s},\overline{s})\subset N_A$. Also assume that $u_n\to\infty$ is such that $u_n = o(n^{(d+1/(\nu\wedge 2)-\kappa_0-\delta)/(\nu+1)})$ for some $\delta>0$. Then, as $n\to\infty$,*

$$\sup_{s\in(\underline{s},\overline{s})}\left|\frac{n^{-d-1/\alpha}}{f_{X_0}(u_n)}\sum_{t=1}^{n-T+1}\left(\mathbb{1}\{[X_{t+\ell}]_\ell\in u_n\cdot s\cdot A\} - \mathbf{P}\left[[X_{t+\ell}]_\ell\in u_n\cdot s\cdot A\right]\right) - s^{-1-\nu}(H_A(s)^\top\mathbf{1}_T)Z_\alpha\right| \to_{\mathbf{P}^*} 0,$$

on a sufficiently rich probability space.

Proof. By Theorem 3.4, it remains to prove that there is a probability space on which

$$\sup_{s\in(\underline{s},\overline{s})}\left|n^{-d-1/\alpha}\frac{\left(\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y},s,\infty,A)\Big|_{\mathbf{y}=\mathbf{0}_T}\right)^\top}{f_{X_0}(u_n)}\sum_{t=1}^{n-T+1}[X_{t+\ell}]_\ell - s^{-1-\nu}(H_A(s)^\top\mathbf{1}_T)Z_\alpha\right|$$

vanishes in outer probability as $n\to\infty$. By Lemma G.1

$$\sup_{s\in N_A}\left|\frac{\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y},s,\infty,A)\Big|_{\mathbf{y}=\mathbf{0}_T}}{f_{X_0}(u_n)} - s^{-1-\nu}H_A(s)\right| \to 0$$

as $n\to\infty$, so that

$$\sup_{s\in(\underline{s},\overline{s})}\left|n^{-d-1/\alpha}\frac{\left(\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y},s,\infty,A)\Big|_{\mathbf{y}=\mathbf{0}_T}\right)^\top}{f_{X_0}(u_n)}\sum_{t=1}^{n-T+1}[X_{t+\ell}]_\ell - s^{-1-\nu}(H_A(s)^\top\mathbf{1}_T)Z_\alpha\right|$$
$$\le\ o(1)\cdot\left\|n^{-d-1/\alpha}\sum_{t=1}^{n-T+1}[X_{t+\ell}]_\ell\right\|_1 + \sup_{s\in N_A}\left\|s^{-1-\nu}H_A(s)\right\|_\infty\times\left\|n^{-d-1/\alpha}\sum_{t=1}^{n-T+1}[X_{t+\ell}]_\ell - Z_\alpha\mathbf{1}_T\right\|_1 .$$

The first summand of the bound vanishes in probability by Theorem 3.1 and Slutsky's theorem. For the second summand, note that, by Theorem 3.1 and Skorokhod's representation theorem,

$$n^{-d-1/\alpha}\sum_{t=1}^{n-T+1}[X_{t+\ell}]_\ell - Z_\alpha\mathbf{1}_T$$

vanishes in probability as $n \to \infty$ on a sufficiently rich probability space. Since N_A is compact and bounded away from 0, and recalling that H_A, as a uniform limit of continuous functions, is continuous,

$$\sup_{s \in N_A} \left\| s^{-1-\nu} H_A(s) \right\|_\infty \; < \; \infty \,,$$

so that the second summand vanishes in probability on a sufficiently rich probability space. This proves the claimed convergence. □

As pointed out in Section 4.1.2, we now prove a central limit theorem for empirical tail measures with random thresholds and random centering.

Theorem G.3 (Empirical tail measure with random thresholds and random centering)**.** *Let Assumptions 1, 2, and 3 hold, let $A_0, \ldots, A_h$, $h \in \mathbb{N}_0$, be bounded away from 0, and let the threshold sequence $u_n \to \infty$ be such that $u_n = o(n^{(d+1/(\nu\wedge 2)-\kappa_0-\delta)/(\nu+1)})$ for some $\delta > 0$. If the sets $A_0, \ldots, A_h$ are in $\mathcal{C}_T$ and satisfy Assumption 7 with functions $H_{A_0}, \ldots, H_{A_h}$, and if $(Y_n^{(i)})$, for $i \in \{0, \ldots, h\}$, are random sequences that satisfy $Y_n^{(i)}/u_n \to_{\mathbf{P}} 1$ as $n \to \infty$,*

$$\begin{aligned} & n^{1-d-1/\alpha} u_n \left[\widehat{\mathbf{P}}_n(Y_n^{(i)}/u_n, A_i) \; - \; \mathbf{P}_n(s, A_i)\Big|_{s=Y_n^{(i)}/u_n} \right]_{i=0,\ldots,h} \\ & \to_d \; \frac{\nu}{2} Z_\alpha [H_{A_i}(1)^\top \mathbf{1}_T]_{i=0,\ldots,h} \qquad \textit{as } n \to \infty \,. \end{aligned}$$

Proof. We use the Cramér-Wold device. Let $\lambda_0, \ldots, \lambda_h \in \mathbb{R}$, and $N_{A_0}, \ldots, N_{A_h}$ in the sense of Assumption 7. Fix $(\underline{s}, \overline{s}) \subset \bigcap_{i=0}^h N_{A_i}$, such that $\underline{s} < 1 < \overline{s}$. Since, for all $i \in \{0, \ldots, h\}$, $Y_n^{(i)}/u_n \to_{\mathbf{P}} 1$ as $n \to \infty$, we have $Y_n^{(i)}/u_n \in (\underline{s}, \overline{s})$ for all $i \in \{0, \ldots, h\}$ with probability converging to 1 as $n \to \infty$. Then, for any $\delta > 0$, as $n \to \infty$, noting that we have $n/(n-T+1) \in (1,2)$,

$$\begin{aligned} \mathbf{P} \Bigg[\Bigg| \sum_{i=0}^h \lambda_i \Bigg(& n^{1-d-1/\alpha} \frac{\mathbf{P}[|X_0| > u_n]}{f_{X_0}(u_n)} \Bigg(\widehat{\mathbf{P}}_n(Y_n^{(i)}/u_n, A_i) \; - \; \mathbf{P}_n(s, A_i) \Bigg|_{s=Y_n^{(i)}/u_n} \Bigg) \\ & - \frac{n}{n-T+1} n^{-d-1/\alpha} \sum_{t=1}^{n-T+1} \frac{\Big(\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s, \infty, A_i)\Big|_{(\mathbf{y},s)=(\mathbf{0}_T, Y_n^{(i)}/u_n)} \Big)^\top [X_{t+\ell}]_\ell}{f_{X_0}(u_n)} \Bigg) \Bigg| \; > \; \delta \Bigg] \\ \le \; \sum_{i=0}^h \mathbf{P}^* \Bigg[|\lambda_i| \sup_{s \in (\underline{s}, \overline{s})} \Bigg| & \frac{n^{-d-1/\alpha}}{f_{X_0}(u_n)} \sum_{t=1}^{n-T+1} \Bigg(\mathbb{1}\{[X_{t+\ell}]_\ell \in u_n \cdot s \cdot A_i\} \; - \; \mathbf{P}[[X_\ell]_\ell \in u_n \cdot s \cdot A_i] \\ & - \Big(\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s, \infty, A_i)\Big|_{\mathbf{y}=\mathbf{0}_T} \Big)^\top [X_{t+\ell}]_\ell \Bigg) \Bigg| \; > \; \frac{\delta}{2(h+1)} \Bigg] \\ & + \; o(1) \\ \to \; 0 & \end{aligned}$$

by Theorem 3.4. Thus, by Slutsky's theorem,

$$\sum_{i=0}^h \lambda_i \left(n^{1-d-1/\alpha} \frac{\mathbf{P}[|X_0| > u_n]}{f_{X_0}(u_n)} \left(\widehat{\mathbf{P}}_n(Y_n^{(i)}/u_n, A_i) \; - \; \mathbf{P}_n(s, A_i)\Big|_{s=Y_n^{(i)}/u_n} \right) \right)$$

and

$$\sum_{i=0}^{h} \lambda_i \frac{\left(\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s, \infty, A_i)\Big|_{(\mathbf{y},s)=(\mathbf{0}_T, Y_n^{(i)}/u_n)}\right)^{\top}}{f_{X_0}(u_n)} \left(n^{-d-1/\alpha} \sum_{t=1}^{n-T+1} [X_{t+\ell}]_\ell\right)$$

have the same weak limiting behavior. By Theorem 3.1,

$$n^{-d-1/\alpha} \sum_{t=1}^{n-T+1} [X_{t+\ell}]_\ell \;\to\; Z_\alpha \mathbf{1}_T$$

weakly as $n \to \infty$. On the other hand, by Lemma G.1 and the continuous mapping theorem,

$$\sum_{i=0}^{h} \lambda_i \frac{\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s, \infty, A_i)\Big|_{(\mathbf{y},s)=(\mathbf{0}_T, Y_n^{(i)}/u_n)}}{f_{X_0}(u_n)} \;\to_{\mathbf{P}}\; \sum_{i=0}^{h} \lambda_i H_{A_i}(1)\,.$$

Again, by Slutsky's theorem,

$$\sum_{i=0}^{h} \lambda_i \frac{\left(\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, s, \infty, A_i)\Big|_{(\mathbf{y},s)=(\mathbf{0}_T, Y_n^{(i)}/u_n)}\right)^{\top}}{f_{X_0}(u_n)} \left(n^{-d-1/\alpha} \sum_{t=1}^{n-T+1} [X_{t+\ell}]_\ell\right) \;\to_d\; \sum_{i=0}^{h} \lambda_i \left(H_{A_i}(1)^{\top} \mathbf{1}_T\right) Z_\alpha\,.$$

Finally, note that, by Karamata's theorem (see Bingham et al., 1987, Proposition 1.5.10 on p.27) and symmetry of ε (and hence of X_0)

$$\frac{\mathbf{P}[|X_0| > u_n]}{f_{X_0}(u_n)} \;\sim\; u_n \frac{2}{\nu}\,.$$

This finishes the proof. □

G.2. Proofs of the main results

Proof of Theorem 4.2: We use the Cramér-Wold device. Let $\lambda_0, \ldots, \lambda_h \in \mathbb{R}$, and note that for $n \geq 2T$ we have $n/(n-T+1) \in (1,2)$. Then, for any $\delta > 0$, as $n \to \infty$

$$\begin{aligned}
&\mathbf{P}\Bigg[\Bigg|\sum_{i=0}^{h} \lambda_i \Bigg(n^{1-d-1/\alpha} \frac{\mathbf{P}[|X_0| > u_n]}{f_{X_0}(u_n)} \left(\widehat{\mathbf{P}}_n(1, A_i) \;-\; \mathbf{P}_n(1, A_i)\right) \\
&\qquad - \frac{n}{n-T+1} n^{-d-1/\alpha} \sum_{t=1}^{n-T+1} \frac{\left(\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, 1, \infty, A_i)\Big|_{\mathbf{y}=\mathbf{0}_T}\right)^{\top} [X_{t+\ell}]_\ell}{f_{X_0}(u_n)}\Bigg)\Bigg| \;>\; \delta\Bigg] \\
&\leq \sum_{i=0}^{h} \mathbf{P}\Bigg[|\lambda_i| \left|\frac{n^{-d-1/\alpha}}{f_{X_0}(u_n)} \sum_{t=1}^{n-T+1} \Bigg(\mathbb{1}\{[X_{t+\ell}]_\ell \in u_n \cdot A_i\} \;-\; \mathbf{P}[[X_\ell]_\ell \in u_n \cdot A_i]\right. \\
&\qquad\qquad \left. - \left(\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, 1, \infty, A_i)\Big|_{\mathbf{y}=\mathbf{0}_T}\right)^{\top} [X_{t+\ell}]_\ell\Bigg)\right| \;>\; \frac{\delta}{2(h+1)}\Bigg] \\
&\to\; 0
\end{aligned}$$

by Corollary 3.3. Thus, by Slutsky's theorem and $n/(n-T+1) \to 1$,

$$\sum_{i=0}^{h} \lambda_i \left(n^{1-d-1/\alpha} \frac{\mathbf{P}[|X_0| > u_n]}{f_{X_0}(u_n)} \left(\widehat{\mathbf{P}}_n(1, A_i) \; - \; \mathbf{P}_n(1, A_i) \right) \right)$$

has the same weak limiting behavior as

$$\sum_{i=0}^{h} \lambda_i \frac{\left(\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, 1, \infty, A_i) \Big|_{\mathbf{y}=\mathbf{0}_T} \right)^{\top}}{f_{X_0}(u_n)} \left(n^{-d-1/\alpha} \sum_{t=1}^{n-T+1} [X_{t+\ell}]_\ell \right) .$$

Again, by Karamata's theorem and symmetry of X_0,

$$\frac{\mathbf{P}[|X_0| > u_n]}{f_{X_0}(u_n)} \sim u_n \frac{2}{\nu} ,$$

so that we make this replacement in the final result. By Theorem 3.1,

$$n^{-d-1/\alpha} \sum_{t=1}^{n-T+1} [X_{t+\ell}]_\ell \;\to\; Z_\alpha \mathbf{1}_T$$

weakly as $n \to \infty$. On the other hand, by Assumption 4

$$\sum_{i=0}^{h} \lambda_i \frac{\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, 1, \infty, A_i) \Big|_{\mathbf{y}=\mathbf{0}_T}}{f_{X_0}(u_n)} \;\to_{\mathbf{P}}\; \sum_{i=0}^{h} \lambda_i H_{A_i}(1) ,$$

so that, by Slutsky's theorem,

$$\sum_{i=0}^{h} \lambda_i \frac{\left(\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y}, 1, \infty, A_i) \Big|_{\mathbf{y}=\mathbf{0}_T} \right)^{\top}}{f_{X_0}(u_n)} \left(n^{-d-1/\alpha} \sum_{t=1}^{n-T+1} [X_{t+\ell}]_\ell \right) \;\to\; \sum_{i=0}^{h} \lambda_i \left(H_{A_i}(1)^{\top} \mathbf{1}_T \right) Z_\alpha ,$$

which finishes the proof. □

An overview of the proof of Theorem 4.8, coming next, is given in Section 4.1.2.

Proof of Theorem 4.8: Replacing random centering as in Theorem G.3 by deterministic centering adds a correction to the limit. To account for this correction, we set $r_n = n^{1-d-1/\alpha} u_n$ and we make the decomposition

$$\begin{aligned}
& r_n \left(\widehat{\mathbf{P}}_n(X_{n-k:n}/u_n, A_i) \; - \; \mathbf{P}_n(1, A_i) \right) \\
&= r_n \left(\widehat{\mathbf{P}}_n(X_{n-k:n}/u_n, A_i) \; - \; \mathbf{P}_n(s, A_i) \Big|_{s=X_{n-k:n}/u_n} \right) \\
&\quad + r_n \left(\mathbf{P}_n(s, A_i) \Big|_{s=X_{n-k:n}/u_n} \; - \; \mathbf{P}_n(1, A_i) \right) \\
&= r_n \left(\widehat{\mathbf{P}}_n(X_{n-k:n}/u_n, A_i) \; - \; \mathbf{P}_n(s, A_i) \Big|_{s=X_{n-k:n}/u_n} \right) \\
&\quad + \frac{\partial}{\partial s} \mathbf{P}_n(s, A_i) \Big|_{s=1} \cdot r_n \left(\frac{X_{n-k:n}}{u_n} \; - \; 1 \right) + r_n \int_1^{X_{n-k:n}/u_n} \left(\frac{\partial}{\partial s} \mathbf{P}_n(s, A_i) - \frac{\partial}{\partial s} \mathbf{P}_n(s, A_i) \Big|_{s=1} \right) \mathrm{d}s .
\end{aligned}$$

Note that $\frac{\partial}{\partial s}\mathbf{P}_n(s,A)$ exists and is continuous due to Lemma B.3(ii); see also Remark 4.4. Now

$$\begin{aligned}
&\left| r_n \int_1^{X_{n-k:n}/u_n} \left(\frac{\partial}{\partial s}\mathbf{P}_n(s,A_i) - \frac{\partial}{\partial s}\mathbf{P}_n(s,A_i)\Big|_{s=1} \right) \mathrm{d}s \right| \\
&\le r_n \left| \frac{X_{n-k:n}}{u_n} \; - \; 1 \right| \times \sup_{s\in[1\wedge X_{n-k:n}/u_n, 1\vee X_{n-k:n}/u_n]} \left| \frac{\partial}{\partial s}\mathbf{P}_n(s,A_i) - \frac{\partial}{\partial s}\mathbf{P}_n(s,A_i)\Big|_{s=1} \right| \\
&= r_n \left| \frac{X_{n-k:n}}{u_n} \; - \; 1 \right| \times o_{\mathbf{P}}(1)
\end{aligned}$$

since $X_{n-k:n}/u_n \to_{\mathbf{P}} 1$ as $n \to \infty$ and the (deterministic) integrand converges uniformly to a continuous limit on N_{A_i}, see Remark 4.5. Consequently

$$\begin{aligned}
&r_n \left(\widehat{\mathbf{P}}_n(X_{n-k:n}/u_n, A_i) \; - \; \mathbf{P}_n(1,A_i) \right) \\
&\quad = \; r_n \left(\widehat{\mathbf{P}}_n(X_{n-k:n}/u_n, A_i) \; - \; \mathbf{P}_n(s,A_i)\Big|_{s=X_{n-k:n}/u_n} \right) \\
&\qquad + \left(\frac{\partial}{\partial s}\mathbf{P}_n(s,A_i)\Big|_{s=1} + o_{\mathbf{P}}(1) \right) \cdot r_n \left(\frac{X_{n-k:n}}{u_n} \; - \; 1 \right) .
\end{aligned} \tag{34}$$

We now study the joint convergence of the two terms in the right-hand side of (34). For this, we remark that by the properties of quantile functions, for $s^* \in \mathbb{R}$, the inequalities

$$r_n \left(\frac{X_{n-k:n}}{u_n} \; - \; 1 \right) \; \le \; s^* \tag{35}$$

and

$$r_n \left(\frac{1}{n} \sum_{t=1}^n \frac{\mathbb{1}\{X_t > u_n(1+s^*/r_n)\}}{\mathbf{P}[X_0 > u_n(1+s^*/r_n)]} \; - \; 1 \right) \; \le \; r_n \left(\frac{\mathbf{P}[X_0 > u_n]}{\mathbf{P}[X_0 > (1+s^*/r_n)u_n]} \; - \; 1 \right)$$

are equivalent. Set

$$\xi(u_n) \; = \; \frac{u_n \cdot f_{X_0}(u_n)}{\mathbf{P}[X_0 > u_n]} .$$

As shown in the proof of Lemma D.1 in Scheffel et al. (2026) if, $u_n = o\left(n^{(d+1/(\nu\wedge 2)-\kappa_0-\delta)/(\nu+1)}\right)$ for some $\delta > 0$, it holds for $s^* \in \mathbb{R}$

$$r_n \left(\frac{\mathbf{P}[X_0 > u_n]}{\mathbf{P}[X_0 > (1+s^*/r_n)u_n]} \; - \; 1 \right) \; = \; s^* \cdot \xi(u_n)\,(1+o(1))$$

and

$$\frac{r_n/\xi(u_n)}{n} \sum_{t=1}^n \left(\frac{\mathbb{1}\{X_t > u_n\}}{\mathbf{P}[X_0 > u_n]} \; - \; \frac{\mathbb{1}\{X_t > u_n(1+s^*/r_n)\}}{\mathbf{P}[X_0 > u_n(1+s^*/r_n)]} \right) \; = \; o_{\mathbf{P}}(1) .$$

Recall that $\xi(u_n) \to \nu$ as $n \to \infty$ by Karamata's theorem. Then (35) is equivalent to

$$\frac{r_n}{n} \sum_{t=1}^n \left(\frac{\mathbb{1}\{X_t > u_n\}}{\mathbf{P}[X_0 > u_n]} \; - \; 1 \right) \le s^* \cdot \nu\,(1+o(1)) + o_{\mathbf{P}}(1).$$

Also note that

$$\begin{aligned}
&\frac{r_n}{n} \sum_{t=1}^n \left(\frac{\mathbb{1}\{X_t > u_n\}}{\mathbf{P}[X_0 > u_n]} \; - \; 1 \right) \\
&= \; \frac{r_n}{n} \sum_{t=n-T+2}^n \left(\frac{\mathbb{1}\{X_t > u_n\}}{\mathbf{P}[X_0 > u_n]} - 1 \right) \; + \; 2\frac{n-T+1}{n} \left(\frac{r_n}{n-T+1} \sum_{t=1}^{n-T+1} \left(\frac{\mathbb{1}\{X_t > u_n\}}{\mathbf{P}[|X_0| > u_n]} \; - \; \frac{\mathbf{P}[X_0 > u_n]}{\mathbf{P}[|X_0| > u_n]} \right) \right) \\
&= \; o_{\mathbf{P}}(1) \; + \; 2(1+o(1)) r_n \left(\widehat{\mathbf{P}}_n(1,B) \; - \; \mathbf{P}_n(1,B) \right)
\end{aligned}$$

with $B = (1,\infty)\times\mathbb{R}^{T-1}\in\mathcal{C}_T$. Therefore, (35) is equivalent to

$$r_n\left(\widehat{\mathbf{P}}_n(1,B) - \mathbf{P}_n(1,B)\right) \le s^*\cdot\frac{\nu}{2}\cdot(1+o(1)) + o_{\mathbf{P}}(1). \tag{36}$$

By Equation (36), the joint limit of the first and the second term in (34) can be studied in a unified way using Theorem G.3, at the level of distribution functions. Let $s_0,\ldots,s_h,s^*\in\mathbb{R}$. Write

$$E_n(s_0,\ldots,s_h) := \bigcap_{i=0}^{h}\left\{r_n\left(\widehat{\mathbf{P}}_n(X_{n-k:n}/u_n,A_i) - \mathbf{P}_n(s,A_i)\Big|_{s=X_{n-k:n}/u_n}\right) \le s_i\right\}$$

and, by Equation (36)

$$\begin{aligned}&\mathbf{P}\left[E_n(s_0,\ldots,s_h),\quad r_n\left(\frac{X_{n-k:n}}{u_n} - 1\right) \le s^*\right]\\ &= \mathbf{P}\left[E_n(s_0,\ldots,s_h),\quad \frac{r_n}{\nu/2}\left(\widehat{\mathbf{P}}_n(1,B) - \mathbf{P}_n(1,B)\right) \le s^*\cdot(1+o(1))+o_{\mathbf{P}}(1)\right].\end{aligned}$$

We shall now apply Theorem G.3, with h replaced by $h+1$, $A_{h+1}=B$, $Y_n^{(i)}=X_{n-k:n}$ for $i\in\{0,\ldots,h\}$ and $Y_n^{(h+1)}=u_n$. For this, note that the set B satisfies Assumption 7, because

$$\frac{\nabla_{\mathbf{y}}G_{\infty,n}(\mathbf{y},s,\infty,B)\Big|_{\mathbf{y}=\mathbf{0}_T}}{f_{X_0}(s\cdot u_n)} = \frac{\nabla_{\mathbf{y}}\mathbf{P}[X_0>s\cdot u_n-y_0]\Big|_{\mathbf{y}=\mathbf{0}_T}}{f_{X_0}(s\cdot u_n)} = \mathbf{e}_1 = H_B(s) = H_B(1).$$

In particular $H_B(1)^\top\mathbf{1}_T=1$. Also note that $X_{n-k:n}/u_n\to_{\mathbf{P}}1$ by Lemma D.1 in Scheffel et al. (2026), so that $X_{n-k:n}$ is a valid random threshold in the application of Theorem G.3. Then this theorem provides that

$$r_n\left[\left[\left(\widehat{\mathbf{P}}_n(X_{n-k:n}/u_n,A_i) - \mathbf{P}_n(s,A_i)\Big|_{s=X_{n-k:n}/u_n}\right)\right]_{i=0,\ldots,h},\quad \frac{1}{\nu/2}\left(\widehat{\mathbf{P}}_n(1,B) - \mathbf{P}_n(1,B)\right)\right],$$

converges in distribution to

$$\frac{\nu}{2}Z_\alpha\left[[H_{A_i}(1)^\top\mathbf{1}_T]_{i=0,\ldots,h},\quad\frac{2}{\nu}\right].$$

Finally, by (34), Remark 4.5, and Slutsky's theorem,

$$r_n\left[\widehat{\mathbf{P}}_n(X_{n-k:n}/u_n,A_i) - \mathbf{P}_n(1,A_i)\right]_{i=0,\ldots,h}$$

converges in distribution to

$$\frac{\nu}{2}Z_\alpha\left[(H_{A_i}(1))^\top\mathbf{1}_T + p_{A_i}(1)\right]_{i=0,\ldots,h}.$$

The proof is complete. □

Proof of Theorem 4.10: We closely follow the argument of the proof of Corollary 3.3 in Davis and Mikosch (2009). Note that

$$\begin{aligned}&\frac{\widehat{\mathbf{P}}_n(s,A_i)}{\widehat{\mathbf{P}}_n(s,A_0)} - \frac{\mathbf{P}_n(1,A_i)}{\mathbf{P}_n(1,A_0)}\\ &= \frac{1}{\mathbf{P}_n(1,A_0)\cdot\widehat{\mathbf{P}}_n(s,A_0)}\left(\mathbf{P}_n(1,A_0)\left(\widehat{\mathbf{P}}_n(s,A_i) - \mathbf{P}_n(1,A_i)\right) - \mathbf{P}_n(1,A_i)\left(\widehat{\mathbf{P}}_n(s,A_0) - \mathbf{P}_n(1,A_0)\right)\right),\end{aligned}$$

so that

$$\left[\frac{\widehat{\mathbf{P}}_n(s,A_i)}{\widehat{\mathbf{P}}_n(s,A_0)} - \frac{\mathbf{P}_n(1,A_i)}{\mathbf{P}_n(1,A_0)}\right]_{i=1,\ldots,h}$$

$$= \frac{1}{\mathbf{P}_n(1,A_0)\cdot\widehat{\mathbf{P}}_n(s,A_0)}\begin{bmatrix}\mathbf{P}_n(1,A_0)\cdot\mathbf{I}_h & [-\mathbf{P}_n(1,A_i)]_{i=1,\ldots,h}\end{bmatrix}\begin{bmatrix}\left[\widehat{\mathbf{P}}_n(s,A_i) - \mathbf{P}_n(1,A_i)\right]_{i=1,\ldots,h}\\ \widehat{\mathbf{P}}_n(s,A_0) - \mathbf{P}_n(1,A_0)\end{bmatrix}.$$

It holds, for $Y_n = u_n$ or $Y_n = X_{n-k:n}$ respectively, that

$$\widehat{\mathbf{P}}_n(Y_n/u_n,A_0) - \mathbf{P}_n(1,A_0) \to_{\mathbf{P}} 0 \qquad \text{as } n\to\infty,$$

by Theorem 4.2 and Theorem 4.8 respectively. Since $\mathbf{P}_n(1,A_0)\to\mu_T(A_0)>0$ as $n\to\infty$,

$$\widehat{\mathbf{P}}_n(Y_n/u_n,A_0)\cdot\mathbf{P}_n(1,A_0) \to_{\mathbf{P}} (\mu_T(A_0))^2$$

for $Y_n = u_n$ or $Y_n = X_{n-k:n}$. Therefore,

$$\frac{1}{\mathbf{P}_n(1,A_0)\cdot\widehat{\mathbf{P}}_n(Y_n/u_n,A_0)}\begin{bmatrix}\mathbf{P}_n(1,A_0)\cdot\mathbf{I}_h & [-\mathbf{P}_n(1,A_i)]_{i=1,\ldots,h}\end{bmatrix}$$
$$\to_{\mathbf{P}} \frac{1}{(\mu_T(A_0))^2}\begin{bmatrix}\mu_T(A_0)\cdot\mathbf{I}_h & [-\mu_T(A_i)]_{i=1,\ldots,h}\end{bmatrix}.$$

By Theorem 4.2

$$n^{1-d-1/\alpha}u_n\begin{bmatrix}\left[\widehat{\mathbf{P}}_n(1,A_i) - \mathbf{P}_n(1,A_i)\right]_{i=1,\ldots,h}\\ \widehat{\mathbf{P}}_n(1,A_0) - \mathbf{P}_n(1,A_0)\end{bmatrix} \to_d \frac{\nu}{2}Z_\alpha\begin{bmatrix}\left[H_{A_i}(1)^\top\mathbf{1}_T\right]_{i=1,\ldots,h}\\ H_{A_0}(1)^\top\mathbf{1}_T\end{bmatrix},$$

and by Theorem 4.8

$$n^{1-d-1/\alpha}q_{X_0}\left(1-\frac{k}{n}\right)\begin{bmatrix}\left[\widehat{\mathbf{P}}_n(X_{n-k:n}/u_n,A_i) - \mathbf{P}_n(1,A_i)\right]_{i=1,\ldots,h}\\ \widehat{\mathbf{P}}_n(X_{n-k:n}/u_n,A_0) - \mathbf{P}_n(1,A_0)\end{bmatrix}$$

converges in distribution to

$$\frac{\nu}{2}Z_\alpha\begin{bmatrix}\left[H_{A_i}(1)^\top\mathbf{1}_T + p_{A_i}(1)\right]_{i=1,\ldots,h}\\ H_{A_0}(1)^\top\mathbf{1}_T + p_{A_0}(1)\end{bmatrix}.$$

The claimed convergence in the two cases follows from Slutsky's theorem. □

H. Proofs of results in Section 5

Proof of Proposition 5.2: **Verifying Assumption 6:** We apply partial integration over each coordinate. First, by Lemma B.1(iv) and dominated convergence,

$$\int_{\mathbb{R}^T}\mathbb{1}\{\mathbf{x}\in u_n\cdot s\cdot A\}\cdot\mathbf{x}^\top\cdot\nabla f_{[X_\ell]_\ell}(\mathbf{x})\,\mathrm{d}\mathbf{x}$$
$$= \sum_{i=1}^T\int_{\mathbb{R}^T}\mathbb{1}\left\{\min_{j\in J_A}(x_{j-1}-u_n\cdot s\cdot\mathfrak{a}_j)\geq 0\right\}x_{i-1}\frac{\partial}{\partial x_{i-1}}f_{[X_\ell]_\ell}(\mathbf{x})\,\mathrm{d}\mathbf{x}$$
$$= \sum_{i=1}^T\lim_{M\to\infty}\int_{\mathbb{R}^T}\mathbb{1}\left\{\min_{j\in J_A}(x_{j-1}-u_n\cdot s\cdot\mathfrak{a}_j)\geq 0\,,|x_{i-1}|\leq M\right\}x_{i-1}\frac{\partial}{\partial x_{i-1}}f_{[X_\ell]_\ell}(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$

Fix $i \notin J_A$. By partial integration over the variable x_{i-1},

$$\begin{aligned}
&\lim_{M\to\infty}\int_{\mathbb{R}^T} \mathbb{1}\left\{\min_{j\in J_A}(x_{j-1}-u_n\cdot s\cdot \mathfrak{a}_j)\geq 0\,, |x_{i-1}|\leq M\right\} x_{i-1}\frac{\partial}{\partial x_{i-1}} f_{[X_\ell]_\ell}(\mathbf{x})\,\mathrm{d}\mathbf{x}\\
&= \lim_{M\to\infty}\int_{\mathbb{R}^{T-1}} \mathbb{1}\left\{\min_{j\in J_A}(x_{j-1}-u_n\cdot s\cdot \mathfrak{a}_j)\geq 0\right\} M f_{[X_\ell]_\ell}(\mathbf{x}_{\ell\neq i-1}, x_{i-1}=M)\,\mathrm{d}\mathbf{x}_{\ell\neq i-1}\\
&\quad + \lim_{M\to\infty}\int_{\mathbb{R}^{T-1}} \mathbb{1}\left\{\min_{j\in J_A}(x_{j-1}-u_n\cdot s\cdot \mathfrak{a}_j)\geq 0\right\} M f_{[X_\ell]_\ell}(\mathbf{x}_{\ell\neq i-1}, x_{i-1}=-M)\,\mathrm{d}\mathbf{x}_{\ell\neq i-1}\\
&\quad - \mathbf{P}[[X_\ell]_\ell \in u_n\cdot s\cdot A].
\end{aligned}$$

Now

$$\begin{aligned}
&\int_{\mathbb{R}^{T-1}} \mathbb{1}\left\{\min_{j\in J_A}(x_{j-1}-u_n\cdot s\cdot \mathfrak{a}_j)\geq 0\right\} M f_{[X_\ell]_\ell}(\mathbf{x}_{\ell\neq i-1}, x_{i-1}=\pm M)\,\mathrm{d}\mathbf{x}_{\ell\neq i-1}\\
&\leq \int_{\mathbb{R}^{T-1}} M f_{[X_\ell]_\ell}(\mathbf{x}_{\ell\neq i-1}, x_{i-1}=\pm M)\,\mathrm{d}\mathbf{x}_{\ell\neq i-1} = M f_{X_{i-1}}(\pm M) \;=\; M f_{X_0}(\pm M)
\end{aligned}$$

converges to 0 as $M\to\infty$ by Assumption 3. Therefore, for all $i\notin J_A$,

$$\lim_{M\to\infty}\int_{\mathbb{R}^T} \mathbb{1}\left\{\min_{j\in J_A}(x_{j-1}-u_n\cdot s\cdot \mathfrak{a}_j)\geq 0\,, |x_{i-1}|\leq M\right\} x_{i-1}\frac{\partial}{\partial x_{i-1}} f_{[X_\ell]_\ell}(\mathbf{x})\,\mathrm{d}\mathbf{x} = -\mathbf{P}[[X_\ell]_\ell \in u_n\cdot s\cdot A].$$

For $i\in J_A$, partial integration analogously yields

$$\begin{aligned}
&\lim_{M\to\infty}\int_{\mathbb{R}^T} \mathbb{1}\left\{\min_{j\in J_A}(x_{j-1}-u_n\cdot s\cdot \mathfrak{a}_j)\geq 0\,, |x_{i-1}|\leq M\right\} x_{i-1}\frac{\partial}{\partial x_{i-1}} f_{[X_\ell]_\ell}(\mathbf{x})\,\mathrm{d}\mathbf{x}\\
&= \lim_{M\to\infty}\int_{\mathbb{R}^{T-1}} \mathbb{1}\left\{\min_{j\in J_A\setminus\{i\}}(x_{j-1}-u_n\cdot s\cdot \mathfrak{a}_j)\geq 0\right\} M f_{[X_\ell]_\ell}(\mathbf{x}_{\ell\neq i-1}, x_{i-1}=M)\,\mathrm{d}\mathbf{x}_{\ell\neq i-1}\\
&- \int_{\mathbb{R}^{T-1}} \mathbb{1}\left\{\min_{j\in J_A\setminus\{i\}}(x_{j-1}-u_n\cdot s\cdot \mathfrak{a}_j)\geq 0\right\} u_n\cdot s\cdot \mathfrak{a}_i\, f_{[X_\ell]_\ell}(\mathbf{x}_{\ell\neq i-1}, x_{i-1}=u_n\cdot s\cdot \mathfrak{a}_i)\,\mathrm{d}\mathbf{x}_{\ell\neq i-1}\\
&- \mathbf{P}[[X_\ell]_\ell \in u_n\cdot s\cdot A]\\
&= -s\cdot u_n\cdot \mathfrak{a}_i\cdot f_{X_0}(s\cdot u_n\cdot \mathfrak{a}_i)\cdot \mathbf{P}[[X_\ell]_\ell \in u_n\cdot s\cdot A \mid X_{i-1}=u_n\cdot s\cdot \mathfrak{a}_i] \;-\; \mathbf{P}[[X_\ell]_\ell \in u_n\cdot s\cdot A]\,.
\end{aligned}$$

Consequently

$$\begin{aligned}
&\int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x}\in u_n\cdot s\cdot A\}\cdot \mathbf{x}^\top\cdot \nabla f_{[X_\ell]_\ell}(\mathbf{x})\,\mathrm{d}\mathbf{x}\\
&= -\sum_{i\in J_A} s\cdot u_n\cdot \mathfrak{a}_i\cdot f_{X_0}(s\cdot u_n\cdot \mathfrak{a}_i)\cdot \mathbf{P}[[X_\ell]_\ell \in u_n\cdot s\cdot A \mid X_{i-1}=u_n\cdot s\cdot \mathfrak{a}_i]\\
&\quad - T\cdot \mathbf{P}[[X_\ell]_\ell \in u_n\cdot s\cdot A]\,.
\end{aligned}$$

By regular variation of f_{X_0} (Assumption 3), symmetry of ε and X_0, and Karamata's theorem,

$$\frac{s\cdot u_n\cdot \mathfrak{a}_i\cdot f_{X_0}(s\cdot u_n\cdot \mathfrak{a}_i)}{\mathbf{P}[|X_0|>u_n]} \;=\; \frac{1}{2}\frac{s\cdot u_n\cdot \mathfrak{a}_i\cdot f_{X_0}(s\cdot u_n\cdot \mathfrak{a}_i)}{\mathbf{P}[X_0>u_n\cdot s\cdot \mathfrak{a}_i]}\cdot\frac{\mathbf{P}[X_0>u_n\cdot s\cdot \mathfrak{a}_i]}{\mathbf{P}[X_0>u_n]} \;\to\; \frac{\nu}{2}\cdot (s\cdot \mathfrak{a}_i)^{-\nu}$$

and

$$\frac{\mathbf{P}[[X_\ell]_\ell \in u_n\cdot s\cdot A]}{\mathbf{P}[|X_0|>u_n]} \;\to\; \mu_T(s\cdot A) \;=\; s^{-\nu}\mu_T(A)\,,$$

both locally uniformly in s, so that

$$\begin{aligned}
&\frac{1}{\mathbf{P}[|X_0|>u_n]}\int_{\mathbb{R}^T} \mathbb{1}\{\mathbf{x}\in u_n\cdot s\cdot A\}\cdot \mathbf{x}^\top\cdot \nabla f_{[X_\ell]_\ell}(\mathbf{x})\,\mathrm{d}\mathbf{x}\\
&\to\; s^{-\nu}\left(-\frac{\nu}{2}\sum_{i\in J_A}\mathfrak{a}_i^{-\nu}p_{i,A} \;-\; T\mu_T(A)\right)\,,
\end{aligned}$$

locally uniformly in s. It follows that

$$p_A(s) \;=\; p_A(1) \;=\; -\sum_{i\in J_A} \mathfrak{a}_i^{-\nu} p_{i,A}\,.$$

Verifying Assumption 7: Rewrite $G_{\infty,n}$ as

$$G_{\infty,n}(\mathbf{y},s,\infty,A) \;=\; \mathbf{P}\left[\min_{j\in J_A}\left(X_{j-1}+y_{j-1}-u_n\cdot s\cdot \mathfrak{a}_j\right) \;\geq\; 0\right],$$

so that

$$\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y},s,\infty,A)\Big|_{\mathbf{y}=\mathbf{0}_T} \;=\; \sum_{j\in J_A} f_{X_0}(s\cdot u_n\cdot \mathfrak{a}_j)\cdot \mathbf{P}\left[[X_\ell]_\ell \in s\cdot u_n\cdot A \mid X_{j-1}=s\cdot u_n\cdot \mathfrak{a}_j\right]\cdot \mathbf{e}_j\,.$$

Therefore

$$\frac{\nabla_{\mathbf{y}} G_{\infty,n}(\mathbf{y},s,\infty,A)\Big|_{\mathbf{y}=\mathbf{0}_T}}{f_{X_0}(s\cdot u_n)} \;=\; \sum_{j\in J_A} \frac{f_{X_0}(s\cdot u_n\cdot \mathfrak{a}_j)}{f_{X_0}(s\cdot u_n)}\cdot \mathbf{P}\left[[X_\ell]_\ell \in s\cdot u_n\cdot A \mid X_{j-1}=s\cdot u_n\cdot \mathfrak{a}_j\right]\cdot \mathbf{e}_j.$$

By local uniformity of regular variation and Assumption 8, this converges uniformly in $s\in N_A$ to

$$H_A(s) \;=\; H_A(1) \;=\; \sum_{j\in J_A} \mathfrak{a}_j^{-(\nu+1)}\cdot p_{j,A}\cdot \mathbf{e}_j\,.$$

This is the announced result. □

Proof of Theorem 5.3. Let $(c_i)_{i\in\mathbb{N}_0}\subset[0,\infty)$ satisfy $0<\sum_{i=0}^\infty c_i^\nu<\infty$. Then, under the assumptions of unimodality of ε, it holds that $f_{\sum_{i=0}^\infty c_i\varepsilon_{-i}}$ is symmetric and regularly varying at $+\infty$ with index $-1-\nu$; see Remark 2.2. Throughout the proof we use that for sufficiently large u and $(c_i)_{i\in\mathbb{N}_0}\subset[0,\infty)$ such that $\sum_{i=0}^\infty c_i^\nu<\infty$ it follows from Karamata's theorem and Equation (15.3.5) in Kulik and Soulier (2020) that

$$\begin{aligned} f_{\sum_{i=0}^\infty c_i\varepsilon_{-i}}(u) \;&=\; \frac{u f_{\sum_{i=0}^\infty c_i\varepsilon_{-i}}(u)}{\nu\mathbf{P}[\sum_{i=0}^\infty c_i\varepsilon_{-i}>u]}\cdot\frac{\nu\mathbf{P}[\varepsilon>u]}{u f_\varepsilon(u)}\cdot\frac{\mathbf{P}[\sum_{i=0}^\infty c_i\varepsilon_{-i}>u]}{\mathbf{P}[\varepsilon>u]} f_\varepsilon(u) \;\sim\; \frac{\mathbf{P}[\sum_{i=0}^\infty c_i\varepsilon_{-i}>u]}{\mathbf{P}[\varepsilon>u]} f_\varepsilon(u)\\ &\sim\; \sum_{i=0}^\infty c_i^\nu f_\varepsilon(u) \qquad \text{as} \qquad u\to\infty\,. \end{aligned} \tag{37}$$

We first show by mathematical induction that, for all $k\in\mathbb{N}_0$,

$$\frac{1}{f_{X_0}(u)}\int_{u\mathfrak{a}_1-y_1}^\infty f_{X_{0,k},X_{1,k+1}}(u-y_0,x)\,\mathrm{d}x \;\to\; 0$$

and

$$\frac{1}{f_{X_0}(u)}\int_{u-y_0}^\infty f_{X_{0,k},X_{1,k+1}}(x,u\mathfrak{a}_1-y_1)\,\mathrm{d}x \;\to\; \mathfrak{a}_1^{-1-\nu}\frac{\sum_{i=1}^{k+1} a_i^\nu}{\sum_{i=0}^\infty a_i^\nu}\,,$$

both uniformly in $\mathbf{y}\in N_{\mathbf{y}}$ as $u\to\infty$, where $N_{\mathbf{y}}$ is an arbitrary compact subset of $\mathbb{R}^2$, which we fix from now on. Then, in a second step, we show that we can take the limit $k\to\infty$.

Analysis for $k\in\mathbb{N}_0$: The arguments for both terms are mostly the same, so we introduce some notation to unify the proof for both cases. Let $\sigma:\{0,1\}\to\{0,1\}$ be a permutation and write

$$X_{\sigma(0),k+\sigma(0)} \;=\; \sum_{i=0}^{k+\sigma(0)} a_i\varepsilon_{\sigma(0)-i} \;=\; \sum_{i=-\sigma(0)}^{k} a_{\sigma(0)+i}\varepsilon_{-i}\,.$$

Our goal is then to show that, for all $k \in \mathbb{N}_0$,

$$\begin{aligned}&\frac{1}{f_\varepsilon(u)} \int_{u\mathfrak{a}_{\sigma(0)}-y_{\sigma(0)}}^{\infty} f_{X_{\sigma(0),k+\sigma(0)},X_{\sigma(1),k+\sigma(1)}}(x, u\mathfrak{a}_{\sigma(1)} - y_{\sigma(1)})\,\mathrm{d}x \\ &\to \begin{cases} \mathfrak{a}_1^{-1-\nu} \cdot \sum_{i=1}^{k+1} a_i^\nu\,, & \sigma(0) = 0\,, \\ 0\,, & \sigma(0) = 1\,, \end{cases}\end{aligned} \tag{38}$$

uniformly in $\mathbf{y} \in N_{\mathbf{y}}$ as $u \to \infty$.

Base case $k = 0$**:** Since $a_{-i} = 0$ for $i \in \mathbb{N}$, and $a_0 = 1$,

$$\begin{bmatrix} X_{\sigma(0),k+\sigma(0)} \\ X_{\sigma(1),k+\sigma(1)} \end{bmatrix} = \begin{bmatrix} a_{\sigma(0)-1} & a_{\sigma(0)} \\ a_{\sigma(1)-1} & a_{\sigma(1)} \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_0 \end{bmatrix},$$

where

$$\left| \det \begin{bmatrix} a_{\sigma(0)-1} & a_{\sigma(0)} \\ a_{\sigma(1)-1} & a_{\sigma(1)} \end{bmatrix} \right| = \left| a_{\sigma(0)-1}a_{\sigma(1)} - a_{\sigma(1)-1}a_{\sigma(0)} \right| = 1\,.$$

Note also that, applying the inverse transformation,

$$\pm \begin{bmatrix} a_{\sigma(1)} & -a_{\sigma(0)} \\ -a_{\sigma(1)-1} & a_{\sigma(0)-1} \end{bmatrix} \begin{bmatrix} x \\ u\mathfrak{a}_{\sigma(1)} - y_{\sigma(1)} \end{bmatrix} = \pm \begin{bmatrix} a_{\sigma(1)} \cdot x - a_{\sigma(0)} \cdot (u\mathfrak{a}_{\sigma(1)} - y_{\sigma(1)}) \\ -a_{\sigma(1)-1} \cdot x + a_{\sigma(0)-1} \cdot (u\mathfrak{a}_{\sigma(1)} - y_{\sigma(1)}) \end{bmatrix}.$$

Since the determinant is ± 1, it holds by the symmetry of ε,

$$\begin{aligned}&f_{X_{\sigma(0),k+\sigma(0)},X_{\sigma(1),k+\sigma(1)}}(x, u\mathfrak{a}_{\sigma(1)} - y_{\sigma(1)}) \\ &= f_\varepsilon(a_{\sigma(1)} \cdot x - a_{\sigma(0)} \cdot (u\mathfrak{a}_{\sigma(1)} - y_{\sigma(1)}))f_\varepsilon(-a_{\sigma(1)-1} \cdot x + a_{\sigma(0)-1} \cdot (u\mathfrak{a}_{\sigma(1)} - y_{\sigma(1)}))\,.\end{aligned}$$

Also note that $a_{\sigma(1)} \in \{1, a_1\}$, so that making the change of variables

$$x \mapsto \frac{x + a_{\sigma(0)} \cdot (u\mathfrak{a}_{\sigma(1)} - y_{\sigma(1)})}{a_{\sigma(1)}},$$

the right-hand side becomes

$$f_\varepsilon(x)f_\varepsilon\left(-\frac{a_{\sigma(1)-1}}{a_{\sigma(1)}} \cdot x + \left(a_{\sigma(0)-1} - \frac{a_{\sigma(1)-1}a_{\sigma(0)}}{a_{\sigma(1)}}\right) \cdot (u\mathfrak{a}_{\sigma(1)} - y_{\sigma(1)})\right).$$

The lower bound in (38) for the change of variables in the integral is

$$a_{\sigma(1)}\left(u\mathfrak{a}_{\sigma(0)} - y_{\sigma(0)}\right) - a_{\sigma(0)}\left(u\mathfrak{a}_{\sigma(1)} - y_{\sigma(1)}\right) = u \cdot (a_{\sigma(1)}\mathfrak{a}_{\sigma(0)} - a_{\sigma(0)}\mathfrak{a}_{\sigma(1)}) + O_{N_{\mathbf{y}}}(1)\,,$$

where $O_{N_{\mathbf{y}}}(1)$ denotes a quantity that is bounded in $\mathbf{y} \in N_{\mathbf{y}}$. Then,

$$\begin{aligned}&\frac{1}{f_\varepsilon(u)} \int_{u\mathfrak{a}_{\sigma(0)}-y_{\sigma(0)}}^{\infty} f_{X_{\sigma(0),k+\sigma(0)},X_{\sigma(1),k+\sigma(1)}}(x, u\mathfrak{a}_{\sigma(1)} - y_{\sigma(1)})\,\mathrm{d}x \\ &= \frac{1}{a_{\sigma(1)}} \frac{1}{f_\varepsilon(u)} \int_{u\cdot(a_{\sigma(1)}\mathfrak{a}_{\sigma(0)} - a_{\sigma(0)}\mathfrak{a}_{\sigma(1)}) + O_{N_{\mathbf{y}}}(1)}^{\infty} f_\varepsilon(x) \\ &\quad \times f_\varepsilon\left(-\frac{a_{\sigma(1)-1}}{a_{\sigma(1)}} \cdot x + \left(a_{\sigma(0)-1} - \frac{a_{\sigma(1)-1}a_{\sigma(0)}}{a_{\sigma(1)}}\right) \cdot u\mathfrak{a}_{\sigma(1)} + O_{N_{\mathbf{y}}}(1)\right) \mathrm{d}x\,.\end{aligned}$$

Since $\mathfrak{a}_1 \geq 1$, we have $a_0\mathfrak{a}_1 \geq a_0 > a_1$, the lower bound of the integration range satisfies

$$u(a_{\sigma(1)}\mathfrak{a}_{\sigma(0)} - a_{\sigma(0)}\mathfrak{a}_{\sigma(1)}) \ + \ O_{N_{\mathbf{y}}}(1) \ = \ \begin{cases} u(a_1 - \mathfrak{a}_1) \ + \ O_{N_{\mathbf{y}}}(1)\,, & \sigma(0) = 0\,, \\ u(\mathfrak{a}_1 - a_1) \ + \ O_{N_{\mathbf{y}}}(1)\,, & \sigma(0) = 1\,. \end{cases} \tag{39}$$

Thus, the argument of the second density term satisfies on the integration range in (39) that

$$-\frac{a_{\sigma(1)-1}}{a_{\sigma(1)}} \cdot x \ + \ \left(a_{\sigma(0)-1} - \frac{a_{\sigma(1)-1}a_{\sigma(0)}}{a_{\sigma(1)}}\right) \cdot u\mathfrak{a}_{\sigma(1)} \ + \ O_{N_{\mathbf{y}}}(1) \ = \ \begin{cases} -\dfrac{x + u\mathfrak{a}_1}{a_1} \ + \ O_{N_{\mathbf{y}}}(1)\,, & \sigma(0) = 0\,, \\ u \ + \ O_{N_{\mathbf{y}}}(1)\,, & \sigma(0) = 1\,, \end{cases}$$

$$\begin{cases} \leq \ -u \ + \ O_{N_{\mathbf{y}}}(1)\,, & \sigma(0) = 0\,, \\ \geq \ u \ + \ O_{N_{\mathbf{y}}}(1)\,, & \sigma(0) = 1\,. \end{cases}$$

For $\sigma(0) = 0$, note that, by monotonicity of f_ε (see Remark 2.2) and local uniformity of regular variation,

$$\frac{1}{f_\varepsilon(u)} f_\varepsilon\left(-\frac{x + u\mathfrak{a}_1}{a_1} \ + \ O_{N_{\mathbf{y}}}(1)\right) \ \leq \ \frac{f_\varepsilon(-u + O_{N_{\mathbf{y}}}(1))}{f_\varepsilon(u)} \ < \ \infty$$

uniformly as $u \to \infty$. Thus, it follows from dominated convergence and the uniform convergence of the lower bound $u(a_1 - \mathfrak{a}_1) \ + \ O_{N_{\mathbf{y}}}(1)$ of the integration range in Equation (39) to $-\infty$ that

$$\begin{aligned}
&\frac{1}{a_{\sigma(1)}} \frac{1}{f_\varepsilon(u)} \int_{u\cdot(a_{\sigma(1)}\mathfrak{a}_{\sigma(0)} - a_{\sigma(0)}\mathfrak{a}_{\sigma(1)}) \ + \ O_{N_{\mathbf{y}}}(1)}^{\infty} f_\varepsilon\,(x) \\
&\qquad \times f_\varepsilon\left(-\frac{a_{\sigma(1)-1}}{a_{\sigma(1)}} \cdot x \ + \ \left(a_{\sigma(0)-1} - \frac{a_{\sigma(1)-1}a_{\sigma(0)}}{a_{\sigma(1)}}\right) \cdot u\mathfrak{a}_{\sigma(1)} \ + \ O_{N_{\mathbf{y}}}(1)\right) \mathrm{d}x \\
&= \ \frac{1}{a_1} \int_{u(a_1 - \mathfrak{a}_1) \ + \ O_{N_{\mathbf{y}}}(1)}^{\infty} f_\varepsilon\,(x) \cdot \frac{1}{f_\varepsilon(u)} f_\varepsilon\left(-\frac{x + u\mathfrak{a}_1}{a_1} \ + \ O_{N_{\mathbf{y}}}(1)\right) \mathrm{d}x \\
&\to \ \frac{1}{a_1}(a_1/\mathfrak{a}_1)^{1+\nu} = a_1^\nu \mathfrak{a}_1^{-1-\nu}
\end{aligned}$$

uniformly in $\mathbf{y} \in N_{\mathbf{y}}$ as $u \to \infty$ due to local uniformity of regular variation.

For $\sigma(0) = 1$, we use uniform convergence of the lower bound $u(\mathfrak{a}_1 - a_1) \ + \ O_{N_{\mathbf{y}}}(1)$ of the integration range in Equation (39) to ∞ and local uniformity of regular variation to see that

$$\begin{aligned}
&\frac{1}{a_{\sigma(1)}} \frac{1}{f_\varepsilon(u)} \int_{u\cdot(a_{\sigma(1)}\mathfrak{a}_{\sigma(0)} - a_{\sigma(0)}\mathfrak{a}_{\sigma(1)}) \ + \ O_{N_{\mathbf{y}}}(1)}^{\infty} f_\varepsilon\,(x) \\
&\qquad \times f_\varepsilon\left(-\frac{a_{\sigma(1)-1}}{a_{\sigma(1)}} \cdot x \ + \ \left(a_{\sigma(0)-1} - \frac{a_{\sigma(1)-1}a_{\sigma(0)}}{a_{\sigma(1)}}\right) \cdot u\mathfrak{a}_{\sigma(1)} \ + \ O_{N_{\mathbf{y}}}(1)\right) \mathrm{d}x \\
&= \ \frac{f_\varepsilon(u + O_{N_{\mathbf{y}}}(1))}{f_\varepsilon(u)} \int_{u(\mathfrak{a}_1 - a_1) \ + \ O_{N_{\mathbf{y}}}(1)}^{\infty} f_\varepsilon\,(x)\,\mathrm{d}x \ \to \ 0\,,
\end{aligned}$$

uniformly in $\mathbf{y} \in N_{\mathbf{y}}$ as $u \to \infty$. This proves (38) for $k = 0$.

Induction statement: Assume that, for some $k \in \mathbb{N}_0$, it holds (38).

Induction step: Throughout, we use the disintegration formula

$$\begin{aligned}
&\int_{u\mathfrak{a}_{\sigma(0)} - y_{\sigma(0)}}^{\infty} f_{X_{\sigma(0),k+1+\sigma(0)}, X_{\sigma(1),k+1+\sigma(1)}}(x, u\mathfrak{a}_{\sigma(1)} - y_{\sigma(1)})\,\mathrm{d}x \\
&= \ \int_{\mathbb{R}} \left(\int_{u\mathfrak{a}_{\sigma(0)} - y_{\sigma(0)}}^{\infty} f_{X_{\sigma(0),k+\sigma(0)}, X_{\sigma(1),k+\sigma(1)}}(x - a_{k+1+\sigma(0)}z, u\mathfrak{a}_{\sigma(1)} - a_{k+1+\sigma(1)}z - y_{\sigma(1)})\,\mathrm{d}x\right) f_\varepsilon(z)\,\mathrm{d}z \\
&= \ \int_{\mathbb{R}} \left(\int_{u\mathfrak{a}_{\sigma(0)} - a_{k+1+\sigma(0)}z - y_{\sigma(0)}}^{\infty} f_{X_{\sigma(0),k+\sigma(0)}, X_{\sigma(1),k+\sigma(1)}}(x, u\mathfrak{a}_{\sigma(1)} - a_{k+1+\sigma(1)}z \ - y_{\sigma(1)})\,\mathrm{d}x\right) f_\varepsilon(z)\,\mathrm{d}z\,,
\end{aligned}$$

and fix $M > 0$. We split the analysis of the integral into several cases for z, namely, we write

$$\begin{aligned}
&\frac{1}{f_\varepsilon(u)} \int_{u\mathfrak{a}_{\sigma(0)} - y_{\sigma(0)}}^{\infty} f_{X_{\sigma(0),k+1+\sigma(0)}, X_{\sigma(1),k+1+\sigma(1)}}(x, u\mathfrak{a}_{\sigma(1)} - y_{\sigma(1)})\,\mathrm{d}x \\
&= \sum_{j=1}^{5} \int_{D_j} \left(\frac{1}{f_\varepsilon(u)} \int_{u\mathfrak{a}_{\sigma(0)} - a_{k+1+\sigma(0)} z - y_{\sigma(0)}}^{\infty} f_{X_{\sigma(0),k+\sigma(0)}, X_{\sigma(1),k+\sigma(1)}}(x, u\mathfrak{a}_{\sigma(1)} - a_{k+1+\sigma(1)} z \; - y_{\sigma(1)})\,\mathrm{d}x \right) f_\varepsilon(z)\,\mathrm{d}z\,,
\end{aligned} \tag{40}$$

with

$$\begin{aligned}
D_1 &= \{|z| \le M\}\,, \\
D_2 &= \{|u\mathfrak{a}_{\sigma(1)} - a_{k+1+\sigma(1)} z - y_{\sigma(1)}| \le M \quad \text{and} \quad |z| > M\}\,, \\
D_3 &= \left\{|z| \in (M, \delta u] \quad \text{and} \quad |u\mathfrak{a}_{\sigma(1)} - a_{k+1+\sigma(1)} z - y_{\sigma(1)}| > M\right\}\,, \\
D_4 &= \left\{|z| > \delta u \quad \text{and} \quad |u\mathfrak{a}_{\sigma(1)} - a_{k+1+\sigma(1)} z - y_{\sigma(1)}| \in (M, \delta u]\right\}\,, \\
D_5 &= \left\{|z| \wedge |u\mathfrak{a}_{\sigma(1)} - a_{k+1+\sigma(1)} z - y_{\sigma(1)}| > \delta u\right\}\,,
\end{aligned}$$

for fixed $\delta \in (0, 1/4)$, where the sets form a disjoint partition of $\mathbb{R}$.

Analysis of D_1: If $|z|$ is bounded by M, then it follows from the induction statement that the integral with domain D_1 in (40) converges to

$$\mathbf{P}[|\varepsilon| \le M] \cdot \begin{cases} \mathfrak{a}_1^{-1-\nu} \sum_{i=1}^{k+1} a_i^\nu\,, & \sigma(0) = 0\,, \\ 0\,, & \sigma(0) = 1\,, \end{cases}$$

uniformly in $\mathbf{y} \in N_{\mathbf{y}}$ as $u \to \infty$.

Analysis of D_2: Note that $a_{k+1+\sigma(1)} \in \{a_{k+1}, a_{k+2}\}$ and $a_{k+1} > a_{k+2} > 0$. We apply the change of variables

$$z \;\mapsto\; \frac{u\mathfrak{a}_{\sigma(1)} - y_{\sigma(1)} - z}{a_{k+1+\sigma(1)}}$$

so that D_2 is transformed to

$$|z| \;\le\; M \qquad \text{and} \qquad \left| \frac{u\mathfrak{a}_{\sigma(1)} - y_{\sigma(1)} - z}{a_{k+1+\sigma(1)}} \right| \;>\; M\,.$$

After the change of variables, the lower bound of the integration range in x becomes

$$u\left(\mathfrak{a}_{\sigma(0)} - \mathfrak{a}_{\sigma(1)} \frac{a_{k+1+\sigma(0)}}{a_{k+1+\sigma(1)}} \right) \;+\; z\frac{a_{k+1+\sigma(0)}}{a_{k+1+\sigma(1)}} \;+\; O_{N_{\mathbf{y}}}(1)\,.$$

The integral then becomes

$$\begin{aligned}
\frac{1}{a_{k+1+\sigma(1)}} \frac{1}{f_\varepsilon(u)} \int_{-M}^{M} &\left(\int_{u\left(\mathfrak{a}_{\sigma(0)} - \mathfrak{a}_{\sigma(1)} \frac{a_{k+1+\sigma(0)}}{a_{k+1+\sigma(1)}}\right) \,+\, z\frac{a_{k+1+\sigma(0)}}{a_{k+1+\sigma(1)}} \,+\, O_{N_{\mathbf{y}}}(1)}^{\infty} f_{X_{\sigma(0),k+\sigma(0)}, X_{\sigma(1),k+\sigma(1)}}(x, z)\,\mathrm{d}x \right) \\
&f_\varepsilon\left(\frac{u\mathfrak{a}_{\sigma(1)} + O_{N_{\mathbf{y}}}(1) - z}{a_{k+1+\sigma(1)}} \right) \mathbb{1}\left\{ \left| \frac{u\mathfrak{a}_{\sigma(1)} - y_{\sigma(1)} - z}{a_{k+1+\sigma(1)}} \right| \;>\; M \right\} \mathrm{d}z\,.
\end{aligned}$$

Since $a_{k+1} > a_{k+2}$ and $\mathfrak{a}_1 \ge 1$,

$$\mathfrak{a}_{\sigma(0)} - \mathfrak{a}_{\sigma(1)} \frac{a_{k+1+\sigma(0)}}{a_{k+1+\sigma(1)}} \;=\; \begin{cases} 1 - \mathfrak{a}_1 \frac{a_{k+1}}{a_{k+2}} < 0\,, & \sigma(0) = 0\,, \\ \mathfrak{a}_1 - \frac{a_{k+2}}{a_{k+1}} > 0\,, & \sigma(0) = 1\,, \end{cases}$$

so that

$$u\left(\mathfrak{a}_{\sigma(0)} - \mathfrak{a}_{\sigma(1)}\frac{a_{k+1+\sigma(0)}}{a_{k+1+\sigma(1)}}\right) \;+\; z\frac{a_{k+1+\sigma(0)}}{a_{k+1+\sigma(1)}} \;+\; O_{N_{\mathbf{y}}}(1) \;\to\; \begin{cases} -\infty\,, & \sigma(0)=0\,,\\ +\infty\,, & \sigma(0)=1\,,\end{cases}$$

uniformly in $\mathbf{y}\in N_{\mathbf{y}}$ and $|z|\le M$ as $u\to\infty$. In the case of $\sigma(0)=0$, we have

$$\frac{1}{f_\varepsilon(u)} f_\varepsilon\left(\frac{u\mathfrak{a}_{\sigma(1)} + O_{N_{\mathbf{y}}}(1) - z}{a_{k+1+\sigma(1)}}\right)\mathbb{1}\left\{\left|\frac{u\mathfrak{a}_{\sigma(1)} - y_{\sigma(1)} - z}{a_{k+1+\sigma(1)}}\right| \;>\; M\right\} \;<\; \infty \tag{41}$$

uniformly as $u\to\infty$. Since

$$\begin{aligned}&\int_{u\left(\mathfrak{a}_{\sigma(0)}-\mathfrak{a}_{\sigma(1)}\frac{a_{k+1+\sigma(0)}}{a_{k+1+\sigma(1)}}\right)\,+\,z\frac{a_{k+1+\sigma(0)}}{a_{k+1+\sigma(1)}}\,+\,O_{N_{\mathbf{y}}}(1)}^{\infty} f_{X_{\sigma(0),k+\sigma(0)},X_{\sigma(1),k+\sigma(1)}}(x,z)\,\mathrm{d}x\\ &\to\; \int_{\mathbb{R}} f_{X_{\sigma(0),k+\sigma(0)},X_{\sigma(1),k+\sigma(1)}}(x,z)\,\mathrm{d}x \;=\; f_{X_{\sigma(1),k+\sigma(1)}}(z)\,,\end{aligned}$$

it follows from dominated convergence and the pointwise convergence of (41) to $(\mathfrak{a}_1/a_{k+2})^{-1-\nu}$ that the integral with domain D_2 in (40) converges to

$$\mathbf{P}[|X_{\sigma(1),k+\sigma(1)} + o_{N_{\mathbf{y}}}(1)|\le M]\cdot\mathfrak{a}_1^{-1-\nu}a_{k+2}^{\nu} \qquad \text{as } u\to\infty.$$

If $\sigma(0)=1$, then

$$\int_{u\left(\mathfrak{a}_{\sigma(0)}-\mathfrak{a}_{\sigma(1)}\frac{a_{k+1+\sigma(0)}}{a_{k+1+\sigma(1)}}\right)\,+\,z\frac{a_{k+1+\sigma(0)}}{a_{k+1+\sigma(1)}}\,+\,O_{N_{\mathbf{y}}}(1)}^{\infty} f_{X_{\sigma(0),k+\sigma(0)},X_{\sigma(1),k+\sigma(1)}}(x,z)\,\mathrm{d}x \;\to\; 0\,,$$

so that the integral with domain D_2 in (40) converges to 0 as $u\to\infty$.

Analysis of D_3: On this domain one has, for u large enough, regardless of $\mathbf{y}\in N_{\mathbf{y}}$,

$$u\mathfrak{a}_{\sigma(1)} - a_{k+1+\sigma(1)}z - y_{\sigma(1)} \;>\; u\mathfrak{a}_{\sigma(1)}/2\,.$$

Therefore, the integral with domain D_3 in (40) is bounded above by

$$\frac{1}{f_\varepsilon(u)}\sup_{x>u\mathfrak{a}_{\sigma(1)}/2} f_{X_{\sigma(1),k+\sigma(1)}}(x)\cdot\mathbf{P}[|\varepsilon|>M] \;\lesssim\; \mathbf{P}[|\varepsilon|>M]\,,$$

using (37) and the regular variation property of f_ε. Here, the multiplicative constant in $\lesssim$ depends only on $\mathfrak{a}_0,\mathfrak{a}_1$ and is independent of k and u.

Analysis of D_4: One has $|a_{k+1+\sigma(1)}z - (u\mathfrak{a}_{\sigma(1)} - y_{\sigma(1)})|\le\delta u$, which, for sufficiently large u, results in

$$z \;\ge\; \frac{\mathfrak{a}_{\sigma(1)}-2\delta}{a_{k+1+\sigma(1)}}u \;\ge\; \frac{1}{2a_{k+1+\sigma(1)}}u$$

for all $\mathbf{y}\in N_{\mathbf{y}}$, so that the integral with domain D_4 in (40) is bounded above by

$$\frac{1}{a_{k+1+\sigma(1)}}\frac{1}{f_\varepsilon(u)}\sup_{|z|\ge u/(2a_{k+1+\sigma(1)})} f_\varepsilon(|z|)\cdot\mathbf{P}[|X_{\sigma(1),k+\sigma(1)}|>M] \;\lesssim\; \mathbf{P}[|X_{\sigma(1),k+\sigma(1)}|>M]\,,$$

where the multiplicative constant in $\lesssim$ depends only on (a_i) and is independent of k and u.

Analysis of D_5: The integral with domain D_5 in (40) is bounded above by

$$\frac{1}{f_\varepsilon(u)}\sup_{|x|>\delta u} f_{X_{\sigma(1),k+\sigma(1)}}(x)\mathbf{P}[|\varepsilon|>\delta u] \;\lesssim\; \mathbf{P}[|\varepsilon|>\delta u] \;\to\; 0\,,$$

using (37) and the regular variation property of f_ε, uniformly in $\mathbf{y} \in N_{\mathbf{y}}$ as $u \to \infty$.

Conclusion of analysis for $k \in \mathbb{N}_0$: Taking $M \to \infty$ yields (38) for $k+1$, and by mathematical induction, the claim holds for all $k \in \mathbb{N}_0$.

Analysis for $k \to \infty$: Throughout, we use the independence of

$$(X_{\sigma(0),k+\sigma(0)}, X_{\sigma(1),k+\sigma(1)}) \qquad \text{and} \qquad (X_{\sigma(0)} - X_{\sigma(0),k+\sigma(0)}, X_{\sigma(1)} - X_{\sigma(1),k+\sigma(1)})$$

to obtain the disintegration formula

$$\begin{aligned}
&\frac{1}{f_\varepsilon(u)} \int_{u\mathfrak{a}_{\sigma(0)}}^{\infty} f_{X_{\sigma(0)},X_{\sigma(1)}}(x, u\mathfrak{a}_{\sigma(1)})\,\mathrm{d}x \\
&= \frac{1}{f_\varepsilon(u)} \int_{\mathbb{R}^2} \left(\int_{u\mathfrak{a}_{\sigma(0)}}^{\infty} f_{X_{\sigma(0),k+\sigma(0)},X_{\sigma(1),k+\sigma(1)}}(x - r_0, u\mathfrak{a}_{\sigma(1)} - r_1)\,\mathrm{d}x \right) \\
&\qquad\qquad f_{X_{\sigma(0)}-X_{\sigma(0),k+\sigma(0)},X_{\sigma(1)}-X_{\sigma(1),k+\sigma(1)}}(r_0, r_1)\,\mathrm{d}(r_0, r_1),
\end{aligned}$$

valid for any $k \in \mathbb{N}_0$. Again, we write

$$\begin{aligned}
&\frac{1}{f_\varepsilon(u)} \int_{u\mathfrak{a}_{\sigma(0)}}^{\infty} f_{X_{\sigma(0)},X_{\sigma(1)}}(x, u\mathfrak{a}_{\sigma(1)})\,\mathrm{d}x \\
&= \sum_{j=1}^{4} \int_{D'_j} \left(\frac{1}{f_\varepsilon(u)} \int_{u\mathfrak{a}_{\sigma(0)}-r_0}^{\infty} f_{X_{\sigma(0),k+\sigma(0)},X_{\sigma(1),k+\sigma(1)}}(x, u\mathfrak{a}_{\sigma(1)} - r_1)\,\mathrm{d}x \right) \\
&\qquad\qquad f_{X_{\sigma(0)}-X_{\sigma(0),k+\sigma(0)},X_{\sigma(1)}-X_{\sigma(1),k+\sigma(1)}}(r_0, r_1)\,\mathrm{d}(r_0, r_1),
\end{aligned} \tag{42}$$

with

$$\begin{aligned}
D'_1 &= \{|r_0| \vee |r_1| \le 1\}, \\
D'_2 &= \{|r_0| \vee |r_1| \in (1, \delta u]\}, \\
D'_3 &= \{|r_1| > \delta u\}, \\
D'_4 &= \{|r_0| > \delta u \quad \text{and} \quad |r_1| \le \delta u\},
\end{aligned}$$

for fixed $\delta \in (0, 1/4)$, where the sets form a disjoint partition of $\mathbb{R}^2$.

Analysis of D'_1: By (38), it holds that the integral with domain D'_1 in (42) converges to

$$\mathbf{P}[|X_0 - X_{0,k}| \vee |X_1 - X_{1,k+1}| \le 1] \cdot \begin{cases} \mathfrak{a}_1^{-1-\nu} \sum_{i=1}^{k+1} a_i^{\nu}, & \sigma(0) = 0, \\ 0, & \sigma(0) = 1, \end{cases}$$

as $u \to \infty$.

Analysis of D'_2: It holds $u\mathfrak{a}_{\sigma(1)} - r_1 \in [(1-\delta)u\mathfrak{a}_{\sigma(1)}, (1+\delta)u\mathfrak{a}_{\sigma(1)}]$, so that the integral with domain D'_2 in (42) is bounded above by

$$\begin{aligned}
&\frac{1}{f_\varepsilon(u)} \sup_{z \in [(1-\delta)u\mathfrak{a}_{\sigma(1)}, (1+\delta)u\mathfrak{a}_{\sigma(1)}]} f_{X_{\sigma(1),k+\sigma(1)}}(z) \cdot (\mathbf{P}[|X_0 - X_{0,k}| > 1] + \mathbf{P}[|X_1 - X_{1,k+1}| > 1]) \\
&\lesssim \mathbf{P}[|X_0 - X_{0,k}| > 1] + \mathbf{P}[|X_1 - X_{1,k+1}| > 1],
\end{aligned}$$

using (37) and the regular variation property of f_ε. Here the multiplicative constant in $\lesssim$ depends only on δ and $\mathfrak{a}_0, \mathfrak{a}_1$ and is independent of k and u.

Analysis of D_3': It holds

$$\begin{aligned}
&\int_{|r_1|>\delta u}\left(\int_{\mathbb{R}}\left(\frac{1}{f_\varepsilon(u)}\int_{u\mathfrak{a}_{\sigma(0)}-r_0}^{\infty} f_{X_{\sigma(0),k+\sigma(0)},X_{\sigma(1),k+\sigma(1)}}(x,u\mathfrak{a}_{\sigma(1)}-r_1)\,\mathrm{d}x\right)\right.\\
&\qquad\left. f_{X_{\sigma(0)}-X_{\sigma(0),k+\sigma(0)},X_{\sigma(1)}-X_{\sigma(1),k+\sigma(1)}}(r_0,r_1)\,\mathrm{d}r_0\right)\mathrm{d}r_1\\
&\le \int_{|r_1|>\delta u}\frac{1}{f_\varepsilon(u)}\left(\int_{\mathbb{R}} f_{X_{\sigma(0),k+\sigma(0)},X_{\sigma(1),k+\sigma(1)}}(x,u\mathfrak{a}_{\sigma(1)}-r_1)\,\mathrm{d}x\right)\\
&\qquad\left(\int_{\mathbb{R}} f_{X_{\sigma(0)}-X_{\sigma(0),k+\sigma(0)},X_{\sigma(1)}-X_{\sigma(1),k+\sigma(1)}}(r_0,r_1)\,\mathrm{d}r_0\right)\mathrm{d}r_1\\
&= \int_{|r_1|>\delta u} f_{X_{\sigma(1),k+\sigma(1)}}(u\mathfrak{a}_{\sigma(1)}-r_1)\frac{f_{X_{\sigma(1)}-X_{\sigma(1),k+\sigma(1)}}(r_1)}{f_\varepsilon(u)}\,\mathrm{d}r_1\,.
\end{aligned}$$

By (37), this is bounded above by

$$\frac{1}{f_\varepsilon(u)}\sup_{|r_1|>\delta u} f_{X_{\sigma(1)}-X_{\sigma(1),k+\sigma(1)}}(|r_1|)\ \lesssim\ \sum_{i=k+1}^{\infty} a_i^\nu\,,$$

where the multiplicative constant in $\lesssim$ depends only on δ and is independent of k and u.
Analysis of D_4': It holds that the integral with domain D_4' in (42) is bounded above by

$$\frac{1}{f_\varepsilon(u)}\sup_{z\in[(1-\delta)u\mathfrak{a}_{\sigma(1)},(1+\delta)u\mathfrak{a}_{\sigma(1)}]} f_{X_{\sigma(1),k+\sigma(1)}}(z)\mathbf{P}[|X_{\sigma(0)}-X_{\sigma(0),k+\sigma(0)}|>\delta u]\ \to\ 0$$

as $u\to\infty$.
Conclusion of the analysis for $k\to\infty$: For each $k\in\mathbb{N}_0$,

$$\begin{aligned}
&\mathbf{P}[|X_0-X_{0,k}|\vee|X_1-X_{1,k+1}|\le 1]\cdot\begin{cases}\mathfrak{a}_1^{-1-\nu}\sum_{i=1}^{k+1}a_i^\nu\,, & \sigma(0)=0\,,\\ 0\,, & \sigma(0)=1\,,\end{cases}\\
&\le\ \liminf_{u\to\infty}\frac{1}{f_\varepsilon(u)}\int_{u\mathfrak{a}_{\sigma(0)}}^{\infty} f_{X_{\sigma(0)},X_{\sigma(1)}}(x,u\mathfrak{a}_{\sigma(1)})\,\mathrm{d}x\ \le\ \limsup_{u\to\infty}\frac{1}{f_\varepsilon(u)}\int_{u\mathfrak{a}_{\sigma(0)}}^{\infty} f_{X_{\sigma(0)},X_{\sigma(1)}}(x,u\mathfrak{a}_{\sigma(1)})\,\mathrm{d}x\\
&\le\ \mathbf{P}[|X_0-X_{0,k}|\vee|X_1-X_{1,k+1}|\le 1]\cdot\begin{cases}\mathfrak{a}_1^{-1-\nu}\sum_{i=1}^{k+1}a_i^\nu\,, & \sigma(0)=0\,,\\ 0\,, & \sigma(0)=1\,,\end{cases}\\
&\quad+\ C\left(\mathbf{P}[|X_0-X_{0,k}|>1]+\mathbf{P}[|X_1-X_{1,k+1}|>1]+\sum_{i=k+1}^{\infty}a_i^\nu\right)\,,
\end{aligned}$$

where $C>0$ is a constant that is independent of k. Letting $k\to\infty$ yields

$$\mathbf{P}[|X_0-X_{0,k}|\vee|X_1-X_{1,k+1}|\le 1]\to 1,\quad \mathbf{P}[|X_0-X_{0,k}|>1]+\mathbf{P}[|X_1-X_{1,k+1}|>1]+\sum_{i=k+1}^{\infty}a_i^\nu\ \to\ 0\,,$$

which proves the result. □

Proof of Theorem 5.5: Note that $k=k(n)$ satisfies $k\to\infty$, $k/n\to 0$ and $\mathfrak{A}(n/k)=O(1/\sqrt{k})$. Let $u_n'=q_X(1-\frac{k}{n})$. Then, from the regular variation of X, $u_n\sim u_n'$, and

$$\begin{aligned}
\frac{u_n'}{u_n}-1 &= \left(\frac{q_X(1-\lfloor n\cdot\mathbf{P}[X>u_n]\rfloor/n)}{q_X(1-\mathbf{P}[X>u_n])}-\left(\frac{n\cdot\mathbf{P}[X>u_n]}{\lfloor n\cdot\mathbf{P}[X>u_n]\rfloor}\right)^{1/\nu}\right)\\
&\quad+\left(\left(\frac{n\cdot\mathbf{P}[X>u_n]}{\lfloor n\cdot\mathbf{P}[X>u_n]\rfloor}\right)^{1/\nu}-1\right)\\
&= o(\mathfrak{A}(1/\mathbb{P}[X>u_n]))+O(1/\lfloor n\cdot\mathbf{P}[X>u_n]\rfloor)=o(1/\sqrt{k})
\end{aligned}$$

by the uniform inequality in Theorem 2.3.9 on p.48 in De Haan and Ferreira (2006). Thus, likewise,

$$\frac{\mathbf{P}[X > \mathfrak{a} \cdot u_n'] - \mathbf{P}[X > \mathfrak{a} \cdot u_n]}{\mathbf{P}[X > u_n]} = o(1/\sqrt{k}).$$

We therefore focus, without loss of generality, on the case when $u_n = q_X(1 - \frac{k}{n})$. Then

$$\begin{aligned}
&\left\{\sqrt{k}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\mathbb{1}\{X_i > \mathfrak{a} \cdot X_{n-k:n}\}}{\mathbf{P}[X > u_n]} \;-\; \frac{\mathbf{P}[X > \mathfrak{a} \cdot u_n]}{\mathbf{P}[X > u_n]}\right)\right) \;\leq\; t\right\} \\
&= \left\{\widehat{\overline{F}}_n\left(\mathfrak{a} \cdot X_{n-k:n}\right) \;-\; \mathbf{P}\left[X > \mathfrak{a} \cdot q_X\left(1 - \frac{k}{n}\right)\right] \;\leq\; \frac{t}{\sqrt{k}}\mathbf{P}\left[X > q_X\left(1 - \frac{k}{n}\right)\right]\right\} \\
&= \left\{\mathbf{P}\left[X \leq \mathfrak{a} \cdot q_X\left(1 - \frac{k}{n}\right)\right] \;-\; \frac{t}{\sqrt{k}}\mathbf{P}\left[X > q_X\left(1 - \frac{k}{n}\right)\right] \;\leq\; \widehat{F}_n\left(\mathfrak{a} \cdot X_{n-k:n}\right)\right\}.
\end{aligned}$$

Using that for any $\tau \in (0,1)$ and $x \in \mathbb{R}$, $\tau \leq \widehat{F}_n(x)$ if and only if $\widehat{q}_n(\tau) \leq x$, where $\widehat{q}_n(\tau) = X_{\lceil n\tau \rceil:n}$ is the empirical quantile function, we get that, for n large enough, the above event is equal to

$$\begin{aligned}
&\left\{\widehat{q}_n\left(\mathbf{P}\left[X \leq \mathfrak{a} \cdot q_X\left(1 - \frac{k}{n}\right)\right] \;-\; \frac{t}{\sqrt{k}}\mathbf{P}\left[X > q_X\left(1 - \frac{k}{n}\right)\right]\right) \;\leq\; \mathfrak{a} \cdot \widehat{q}_n\left(1 - \frac{k}{n}\right)\right\} \\
&= \left\{\frac{\widehat{q}_n\left(\mathbf{P}\left[X \leq \mathfrak{a} \cdot q_X\left(1 - \frac{k}{n}\right)\right] \;-\; \frac{t}{\sqrt{k}}\mathbf{P}\left[X > q_X\left(1 - \frac{k}{n}\right)\right]\right)}{q_X\left(1 - \frac{k}{n}\right)} \;-\; \mathfrak{a} \cdot \frac{\widehat{q}_n\left(1 - \frac{k}{n}\right)}{q_X\left(1 - \frac{k}{n}\right)} \;\leq\; 0\right\}.
\end{aligned}$$

Write

$$\begin{aligned}
\alpha_n(t) &:= \mathbf{P}\left[X \leq \mathfrak{a} \cdot q_X\left(1 - \frac{k}{n}\right)\right] \;-\; \frac{t}{\sqrt{k}}\mathbf{P}\left[X > q_X\left(1 - \frac{k}{n}\right)\right], \\
\beta_n(t) &:= \frac{n}{k}\left(1 - \alpha_n(t)\right) \\
&= \frac{n}{k}\left(\mathbf{P}\left[X > \mathfrak{a} \cdot q_X\left(1 - \frac{k}{n}\right)\right] \;+\; \frac{t}{\sqrt{k}}\mathbf{P}\left[X > q_X\left(1 - \frac{k}{n}\right)\right]\right).
\end{aligned}$$

Since $k/n = \mathbf{P}[X > q_X(1 - \frac{k}{n})]$, it follows from regular variation of X that

$$\beta_n(t) \;=\; \frac{\mathbf{P}[X > \mathfrak{a} \cdot u_n]}{\mathbf{P}[X > u_n]} \;+\; \frac{t}{\sqrt{k}} \;\to\; \mathfrak{a}^{-\nu} \in (0, \infty) \qquad \text{as } n \to \infty.$$

By Theorem 2.4.8 on p.52 in De Haan and Ferreira (2006), there is a sequence of Brownian motions (W_n) such that, uniformly in $s \in (0,1]$, and for arbitrarily small $\varepsilon > 0$,

$$\frac{\widehat{q}_n\left(1 - \frac{ks}{n}\right)}{q_X\left(1 - \frac{k}{n}\right)} \;=\; s^{-1/\nu} \;+\; \frac{1}{\sqrt{k}}\left(\frac{1}{\nu}s^{-1/\nu-1}W_n(s) + \sqrt{k} \cdot \mathfrak{A}\left(\frac{n}{k}\right)s^{-1/\nu}\frac{s^{-\rho} - 1}{\rho} + s^{-1/\nu-1/2-\varepsilon}o_{\mathbf{P}}(1)\right).$$

An inspection of the proof of this result reveals that it holds in fact uniformly on any interval of the form $(0, s_0]$, with $s_0 > 0$ fixed. Then

$$\mathfrak{a} \cdot \frac{\widehat{q}_n\left(1 - \frac{k}{n}\right)}{q_X\left(1 - \frac{k}{n}\right)} \;=\; \mathfrak{a}\left(1 \;+\; \frac{1}{\sqrt{k}}\frac{1}{\nu}W_n(1) \;+\; o_{\mathbf{P}}(1/\sqrt{k})\right) \;=\; \mathfrak{a} \;+\; \frac{1}{\nu}\frac{\mathfrak{a}}{\sqrt{k}}W_n(1) \;+\; o_{\mathbf{P}}(1/\sqrt{k}) \qquad (43)$$

and, by $\alpha_n(t) = 1 - (k/n)\beta_n(t)$ with $\beta_n(t) \to \mathfrak{a}^{-\nu} \in (0, \infty)$, and continuity properties of the Brownian motion,

$$\begin{aligned}
&\frac{\widehat{q}_n\left(\alpha_n(t)\right)}{q_X\left(1 - \frac{k}{n}\right)} \\
&= \left(\beta_n(t)\right)^{-1/\nu} \;+\; \frac{1}{\sqrt{k}}\left(\frac{1}{\nu}\left(\beta_n(t)\right)^{-1/\nu-1}W_n(\beta_n(t)) + \sqrt{k} \cdot \mathfrak{A}\left(\frac{n}{k}\right)\left(\beta_n(t)\right)^{-1/\nu}\frac{\left(\beta_n(t)\right)^{-\rho} - 1}{\rho} + o_{\mathbf{P}}(1)\right) \\
&= \left(\beta_n(t)\right)^{-1/\nu} \;+\; \frac{1}{\sqrt{k}}\left(\frac{1}{\nu}\mathfrak{a}^{\nu+1}W_n(\mathfrak{a}^{-\nu}) + \sqrt{k} \cdot \mathfrak{A}\left(\frac{n}{k}\right)\mathfrak{a}\frac{\mathfrak{a}^{\rho\nu} - 1}{\rho} + o_{\mathbf{P}}(1)\right).
\end{aligned}$$

Besides,

$$
\begin{aligned}
\beta_n(t) &= \frac{\mathbf{P}[X > \mathfrak{a} \cdot u_n]}{\mathbf{P}[X > u_n]} + \frac{t}{\sqrt{k}} \\
&= \mathfrak{A}\left(\frac{n}{k}\right)\left(\frac{\frac{\mathbf{P}[X > \mathfrak{a} \cdot u_n]}{\mathbf{P}[X > u_n]} - \mathfrak{a}^{-\nu}}{\mathfrak{A}(n/k)} - \mathfrak{a}^{-\nu}\frac{\mathfrak{a}^{\rho\nu}-1}{\rho/\nu}\right) + \mathfrak{a}^{-\nu} + \mathfrak{A}\left(\frac{n}{k}\right)\mathfrak{a}^{-\nu}\frac{\mathfrak{a}^{\rho\nu}-1}{\rho/\nu} + \frac{t}{\sqrt{k}} \\
&= \mathfrak{a}^{-\nu} + \mathfrak{A}\left(\frac{n}{k}\right)\mathfrak{a}^{-\nu}\frac{\mathfrak{a}^{\rho\nu}-1}{\rho/\nu} + \frac{t}{\sqrt{k}} + o(1/\sqrt{k}) \\
&= \mathfrak{a}^{-\nu}\left(1 + \frac{1}{\sqrt{k}}\left(\sqrt{k}\cdot\mathfrak{A}\left(\frac{n}{k}\right)\frac{\mathfrak{a}^{\rho\nu}-1}{\rho/\nu} + \mathfrak{a}^{\nu}t + o(1)\right)\right) \\
&= \left(\mathfrak{a}\left(1 + \frac{1}{\sqrt{k}}\left(\sqrt{k}\cdot\mathfrak{A}\left(\frac{n}{k}\right)\frac{\mathfrak{a}^{\rho\nu}-1}{\rho/\nu} + \mathfrak{a}^{\nu}t + o(1)\right)\right)^{-1/\nu}\right)^{-\nu} \\
&= \left(\mathfrak{a}\left(1 - \frac{1}{\nu}\frac{1}{\sqrt{k}}\left(\sqrt{k}\cdot\mathfrak{A}\left(\frac{n}{k}\right)\frac{\mathfrak{a}^{\rho\nu}-1}{\rho/\nu} + \mathfrak{a}^{\nu}t + o(1)\right)\right)\right)^{-\nu},
\end{aligned}
$$

so that

$$
(\beta_n(t))^{-1/\nu} = \mathfrak{a}\left(1 - \frac{1}{\sqrt{k}}\left(\sqrt{k}\cdot\mathfrak{A}\left(\frac{n}{k}\right)\frac{\mathfrak{a}^{\rho\nu}-1}{\rho} + \frac{1}{\nu}\mathfrak{a}^{\nu}t + o(1)\right)\right).
$$

Therefore

$$
\frac{\widehat{q}_n\left(\alpha_n(t)\right)}{q_X\left(1-\frac{k}{n}\right)} = \mathfrak{a} + \frac{1}{\nu}\frac{\mathfrak{a}^{\nu+1}}{\sqrt{k}}W_n(\mathfrak{a}^{-\nu}) - \frac{1}{\nu}\mathfrak{a}^{\nu+1}\frac{t}{\sqrt{k}} + o_{\mathbf{P}}(1/\sqrt{k}). \tag{44}
$$

Combining (43) with (44), we get

$$
\frac{\widehat{q}_n\left(\alpha_n(t)\right)}{q_X\left(1-\frac{k}{n}\right)} - \mathfrak{a}\frac{\widehat{q}_n\left(1-\frac{k}{n}\right)}{q_X\left(1-\frac{k}{n}\right)} = \mathfrak{a}\frac{1}{\nu}\frac{1}{\sqrt{k}}\left(\mathfrak{a}^{\nu}W_n(\mathfrak{a}^{-\nu}) - W_n(1) - \mathfrak{a}^{\nu}t + o_{\mathbf{P}}(1)\right).
$$

Therefore

$$
\begin{aligned}
&\mathbf{P}\left[\sqrt{k}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\mathbb{1}\{X_i > \mathfrak{a}\cdot X_{n-k:n}\}}{\mathbf{P}[X > u_n]} - \frac{\mathbf{P}[X > \mathfrak{a}\cdot u_n]}{\mathbf{P}[X > u_n]}\right)\right) \le t\right] \\
&= \mathbf{P}\left[\frac{\widehat{q}_n\left(\alpha_n(t)\right)}{q_X\left(1-\frac{k}{n}\right)} - \mathfrak{a}\frac{\widehat{q}_n\left(1-\frac{k}{n}\right)}{q_X\left(1-\frac{k}{n}\right)} \le 0\right] \\
&= \mathbf{P}\left[\mathfrak{a}^{\nu}W_n(\mathfrak{a}^{-\nu}) - W_n(1) + o_{\mathbf{P}}(1) \le \mathfrak{a}^{\nu}t\right] \\
&\to \mathbf{P}\left[W(\mathfrak{a}^{-\nu}) - \mathfrak{a}^{-\nu}W(1) \le t\right] \qquad \text{as } n\to\infty,
\end{aligned}
$$

where $(W(t))_{t\geq 0}$ is a Brownian motion. We have that

$$
\begin{aligned}
&\mathrm{Var}\left[W(\mathfrak{a}^{-\nu}) - \mathfrak{a}^{-\nu}W(1)\right] \\
&= \mathrm{Var}[W(\mathfrak{a}^{-\nu})] + \mathfrak{a}^{-2\nu}\mathrm{Var}[W(1)] - 2\mathfrak{a}^{-\nu}\mathrm{Cov}[W(\mathfrak{a}^{-\nu}), W(1)] \\
&= \mathfrak{a}^{-\nu} + \mathfrak{a}^{-2\nu} - 2\mathfrak{a}^{-\nu}\left(\mathfrak{a}^{-\nu}\wedge 1\right) = \mathfrak{a}^{-\nu}\left(1 + \mathfrak{a}^{-\nu} - 2(\mathfrak{a}^{-\nu}\wedge 1)\right) \\
&= \mathfrak{a}^{-\nu}\left|\frac{1}{\mathfrak{a}^{\nu}} - 1\right|,
\end{aligned}
$$

which proves the claimed limiting distribution. □

Chapter 7

Maxima of stationary systems of randomly time-changed Lévy particles

Maxima of stationary systems of randomly time-changed Lévy particles

Ioan Scheffel[a]

[a] Institute for Stochastics and Applications, University of Stuttgart, D-70563 Stuttgart, Germany

Abstract

In this work, we consider maxima of systems of randomly time-changed Lévy particles. We give a general construction to obtain infinite-dimensional classes $\{Z^\alpha\}$ of stationary max-infinitely divisible (max-id) processes. These classes are indexed by admissible mass functions α, which induce state-dependent time changes of the underlying Lévy particles. This gives a generalization of the well-known (Lévy–)Brown–Resnick process Z^1. In contrast to $\alpha \equiv 1$, the variability of non-constant mass functions α changes the dependence structure of the max-id process and goes beyond the max-stable setting while preserving stationarity. We then explore the extent of the so-called max-domain of attraction (MDA) of a given (Lévy–)Brown–Resnick process Z^1, by studying convergence of rescaled maxima of independent copies of Z^α to Z^1. Thus, our work combines potential theory for Markov processes and extreme value theory to yield a novel, infinite-dimensional, and interpretable class $\{Z^\alpha\}$ of stationary processes in the MDA of a given (Lévy–)Brown–Resnick process Z^1. So far, results on the extent of such domains have been scarce in the literature.

MSC 2020 subject classifications: 60G70, 60G10, 60G51, 60G55, 60J55

Keywords: extreme value theory, max-infinitely divisible process, max-domain of attraction

1 Introduction

1.1 Background and literature review

Given a Markov process $(X_t)_{t\geq 0}$ and a collection of points $\{U^{(i)}\}_{i\in\mathbb{N}}$, we consider the collection of particles obtained by independently starting in each point $U^{(i)}$ an i.i.d. copy $(X_t^{(i)})_{t\geq 0}$ of the Markov process $(X_t)_{t\geq 0}$, that is,

$$\left\{ \left(X_t^{(i)}(U^{(i)}) \right)_{t\geq 0} \right\}_{i\in\mathbb{N}} .$$

We call this a Markovian particle system. For a single particle $(X_t(U))_{t\geq 0}$, stationarity holds if and only if there exists an invariant probability measure μ of the process $(X_t)_{t\geq 0}$, by setting $U \sim \mu$. Extreme value analysis naturally shifts the viewpoint from a single particle to a system of infinitely many particles, where the starting points $\{U^{(i)}\}_{i\in\mathbb{N}}$ form a so-called Poisson point process (PPP). For a distribution function F with density F', it is, for example, possible to choose the PPP $\{U^{(i)}\}_{i\in\mathbb{N}}$ such that $F(x) = \mathbf{P}\left[\max_{i\in\mathbb{N}} U^{(i)} \leq x\right]$, for $x \in \mathbb{R}$. If the arising Markovian particle system is stationary, then the maximum process

$$Z_t \;=\; \max_{i\in\mathbb{N}} X_t^{(i)}(U^{(i)}) \,, \qquad t \geq 0 \,, \tag{1}$$

is stationary as well, with marginal distribution function F. Such a process (regardless of stationarity) is called max-infinitely divisible (max-id), in analogy to the notion of infinite divisibility for sums rather than maxima. It follows from [Bro70] that the Markovian particle system (and the max-id process) is stationary if and only if $\mu(\mathrm{d}x) = F'(x)/F(x)\mathbb{1}\{F(x) > 0\}\mathrm{d}x$ is invariant for $(X_t)_{t\geq 0}$ in the sense that

$$\int_{\mathbb{R}} \mathbf{P}[X_t(x) \in A]\, \mu(\mathrm{d}x) \;=\; \mu(A) \qquad \text{for all Borel sets } A \text{ and all } t \geq 0\,. \tag{2}$$

In this setting the invariant measure μ is necessarily infinite, so that, unlike in the single-particle case discussed above, it cannot be normalized to a probability measure.

A possible choice for F is the Gumbel distribution function

$$F(x) \;=\; \exp(-\exp(-x))\,, \qquad x \in \mathbb{R}\,,$$

with $\mu(\mathrm{d}x) = \exp(-x)\mathrm{d}x$. Any invariant process–measure pair $((X_t)_{t\geq 0}, \exp(-x)\mathrm{d}x)$ in the sense of (2) then yields a stationary max-id process with Gumbel margins, as in (1). Brown and Resnick [BR77] were the first to construct such a process, based on the observation that this specific choice of μ is invariant for the drifted Brownian motion $(B_t - t/2)_{t\geq 0}$. The resulting process ζ_{BR}, called Brown–Resnick process, is stationary with Gumbel margins:

$$\zeta_{\mathrm{BR}}(t) \;:=\; \max_{i\in\mathbb{N}} \left(B_t^{(i)} - t/2 \;+\; U^{(i)}\right)\,, \qquad t \geq 0\,. \tag{3}$$

Even though the single particles $B_t^{(i)} - t/2 + U^{(i)}$ are non-stationary (they converge to $-\infty$ for $t \to \infty$), the particle system $\{(B_t^{(i)} - t/2 + U^{(i)})_{t\geq 0}\}_{i\in\mathbb{N}}$, and the max-id process ζ_{BR} are stationary. Brown and Resnick also showed that ζ_{BR} is not only max-id, but even max-stable. In the context of this work, with the convention that the margins of the max-stable process ζ are Gumbel distributed and the process is indexed over $[0, \infty)$, this means that, for all $n \in \mathbb{N}$,

$$\left(\max_{k=1,\dots,n} \zeta^{(k)}(t) \;-\; \log(n)\right)_{t\geq 0} \;=\; (\zeta(t))_{t\geq 0} \qquad \text{in finite-dimensional distributions,}$$

where $\zeta^{(k)}$, for $k \in \{1, \dots, n\}$, are i.i.d. copies of ζ. De Haan showed in [Haa84] that all max-stable processes (up to marginal transformations) are max-id, similar in structure to (3).

Max-stable processes obtained from systems of Lévy particles beyond the drifted Brownian motion have been studied in [Sto08], where they show stationarity and Gumbel margins of the process

$$\zeta_L(t) \;:=\; \max_{i\in\mathbb{N}} \left(L_t^{(i)} \;+\; U^{(i)}\right)\,, \qquad t \geq 0\,,$$

if the Lévy process $(L_t)_{t\geq 0}$ satisfies $\mathbf{E}[\exp(L_1)] = 1$. We refer to such processes as Lévy–Brown–Resnick processes. It is then easy to see that the Lévy process $(L_t)_{t\geq 0}$ has invariant measure $\exp(-x)\mathrm{d}x$, so the method of [Bro70] applies; see also [EK15] for a generalization of Lévy–Brown–Resnick processes to the case where $(L_t)_{t\geq 0}$ satisfies $\mathbf{E}[\exp(L_1)] \neq 1$.

We note that Kabluchko et al. showed in their seminal paper [KSd09] that the stationarity of ζ_{BR}, among more general max-stable processes, can be derived using entirely different techniques that apply, for example, to Gaussian fields; see also [Kab10] and [MS13] for a complete characterization of uni- and multivariate stationary Gaussian particle systems, respectively.

1.2 Motivation and contribution of this work

All specific examples of stationary max-id processes in the literature, constructed from non-stationary single particles, are (up to marginal transformations) max-stable. The goal of this work is to generalize this construction by allowing for random time change of the underlying Lévy particles. Starting from a Lévy–Brown–Resnick process ζ_L, we construct infinite-dimensional classes $\{Z^\alpha\}$ of stationary max-id processes. These classes are indexed by admissible mass functions α that induce state-dependent time changes of the underlying Lévy particles. In contrast to $\alpha \equiv 1$, which yields the Lévy–Brown–Resnick process $\zeta_L = Z^1$, non-constant mass functions α allow for variability in the dependence structure of the max-id process that goes beyond the max-stable setting. We then use this variability in α to explore the extent of the so-called maximum domain of attraction (MDA) of a given Lévy–Brown–Resnick process ζ_L. To be more precise, we study, for all admissible α, the convergence

$$\left(\max_{k=1,\dots,n} Z_t^{\alpha,(k)} \; - \; \log(n)\right)_{t\geq 0} \;\to\; (\zeta_L(t))_{t\geq 0} \qquad \text{weakly as } n \to \infty,$$

where $(Z_t^{\alpha,(k)})_{t\geq 0}$, for $k \in \{1,\dots,n\}$, are i.i.d. copies of $(Z_t^\alpha)_{t\geq 0}$. More generally, we combine potential theory for Markov processes with extreme value theory to yield a novel, infinite-dimensional, and interpretable class of stationary processes in the MDA of a given Lévy–Brown–Resnick process ζ_L. So far, results on the extent of such domains have been scarce in the literature.

1.3 Structure of this work

In Section 2 we recall the mechanism behind the method of [Bro70]. In Section 3 we perturb the process–measure pair of a Lévy–Brown–Resnick process $\zeta_L = Z^1$ by random time change induced by admissible mass functions α. In Section 4 we give a general construction of stationary max-id process $(Z_t^\alpha)_{t\geq 0}$. Finally, in Section 5 we show that the extremal behavior of $(Z_t^\alpha)_{t\geq 0}$ links back to Z^1 via $(Z_t^\alpha)_{t\geq 0} \in \mathrm{MDA}(Z^1)$.

1.4 Notation

Let $I \subset [0,\infty)$ be an interval and define $\mathbb{D}(I)$ to be the set of càdlàg functions on I endowed with the Skorokhod topology, and let $\mathbb{C}(I)$ be the continuous functions endowed with the uniform topology. The term Markov process is used exclusively to refer to real-valued, time-homogeneous Markov processes with paths in $\mathbb{D}([0,\infty))$. For a Markov process $(X_t)_{t\geq 0}$ we denote $X_t(x)$ to be the process at time $t \geq 0$, started in $x \in \mathbb{R}$ at time 0. We denote $\mathcal{B}(\mathcal{X})$ to be the Borel σ-algebra of the topological space $\mathcal{X}$.

2 Stationary particle systems

To showcase the method of [Bro70], we first introduce Poisson point processes (PPP) and explain how to construct particle systems by attaching to each point an independent copy of a Markov process. Finally we recap a theorem of [Bro70] which states that the resulting particle system is stationary if and only if the intensity measure of the PPP is invariant for the Markov process.

Definition 2.1 (Poisson point process)**.** Let $(\mathcal{X}, \mathcal{C}, \mu)$ be a measure space satisfying $\{x\} \in \mathcal{C}$ for all $x \in \mathcal{X}$ and $\mu(\mathcal{X}) = \infty$. A random collection $\{U^{(i)}\}_{i\in\mathbb{N}}$ of points in $\mathcal{X}$ is called

Poisson point process having intensity measure μ if for every $m \in \mathbb{N}$ and corresponding disjoint sets $C_1, \ldots, C_m \in \mathcal{C}$ and non-negative (including $+\infty$) integers $r_1, \ldots, r_m$ it holds that

$$\mathbf{P}\left[\sum_{i\in\mathbb{N}} \mathbb{1}\{U^{(i)} \in C_j\} \;=\; r_j \qquad \text{for all } j \in \{1,\ldots,m\}\right] \;=\; \prod_{j=1}^{m} p(\mu(C_j), r_j)\,,$$

where

$$p(\lambda, k) \;=\; \begin{cases} \frac{\lambda^k \exp(-\lambda)}{k!} & \text{if} \quad \lambda < \infty \quad \text{and} \quad k < \infty\,, \\ 1 & \text{if} \quad \lambda = \infty \quad \text{and} \quad k = \infty \qquad \text{or} \qquad \lambda = 0 \quad \text{and} \quad k = 0\,, \\ 0 & \text{elsewhere.} \end{cases}$$

Remark 2.2. A collection $\{U^{(i)}\}_{i\in\mathbb{N}}$, in contrast to a set $\{U^{(i)} \mid i \in \mathbb{N}\}$, can contain the same element more than once. The existence of a PPP with prescribed intensity measure is guaranteed if, for example, the intensity measure is σ-finite.

Let $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ be a σ-finite measure space. Given a PPP $\{U^{(i)}\}_{i\in\mathbb{N}}$ on the real line with intensity measure μ, we can attach to each point $U^{(i)}$ an i.i.d. copy $(X_t^{(i)})_{t\geq 0}$ of a Markov process $(X_t)_{t\geq 0}$, independent of $\{U^{(i)}\}_{i\in\mathbb{N}}$, to obtain the random collection

$$\left\{\left(X_t^{(i)}(U^{(i)})\right)_{t\geq 0}\right\}_{i\in\mathbb{N}} \qquad \text{of sample paths in} \qquad \mathbb{D}([0,\infty))\,. \tag{4}$$

We refer to this collection as the particle system associated with the process–measure pair $((X_t)_{t\geq 0}, \mu)$, and denote it by $\mathfrak{P}((X_t)_{t\geq 0}, \mu)$. For a finite sequence of times $0 \leq t_1 < \cdots < t_\ell < \infty$, $\ell \in \mathbb{N}$, we define a finite-dimensional particle system by

$$\mathfrak{P}_{t_1,\ldots,t_\ell}((X_t)_{t\geq 0}, \mu) \;:=\; \left\{\left(X_{t_1}^{(i)}(U^{(i)}), \ldots, X_{t_\ell}^{(i)}(U^{(i)})\right)\right\}_{i\in\mathbb{N}} \qquad \text{with points in} \qquad \mathbb{R}^\ell\,. \tag{5}$$

We say that the full particle system $\mathfrak{P}((X_t)_{t\geq 0}, \mu)$ is stationary if, for all such finite collections of time points and all $h \geq 0$, it holds that

$$\mathfrak{P}_{t_1,\ldots,t_\ell}((X_t)_{t\geq 0}, \mu) \qquad \text{has the same distribution as} \qquad \mathfrak{P}_{t_1+h,\ldots,t_\ell+h}((X_t)_{t\geq 0}, \mu)\,. \tag{6}$$

It will be useful to have more general transformations of the particle system. To this end, let $(E, \mathcal{E})$ be a measurable space and g be a (deterministic) measurable function from $(\mathbb{D}([0,\infty)), \mathcal{B}(\mathbb{D}([0,\infty))))$ to $(E, \mathcal{E})$. We define the g–transformed particle system by

$$\mathfrak{P}_g\left((X_t)_{t\geq 0}, \mu\right) \;=\; \left\{g\left(X_t^{(i)}(U^{(i)})\right)_{t\geq 0}\right\}_{i\in\mathbb{N}} \qquad \text{with points in} \qquad E\,.$$

The following theorem is a direct consequence of [Bro70, Theorem 1], see also Section 3 therein.

Theorem 2.3. *Let $(X_t)_{t\geq 0}$ be a Markov process, $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ a σ-finite measure space with $\mu(\mathbb{R}) = \infty$, and g a measurable function from $(\mathbb{D}([0,\infty)), \mathcal{B}(\mathbb{D}([0,\infty))))$ to a measurable space $(E, \mathcal{E})$. Then the following holds:*

1. *The particle system $\mathfrak{P}_g((X_t)_{t\geq 0}, \mu)$ defines a PPP on $(E, \mathcal{E})$ with intensity measure*

$$\mu_g(A) \;=\; \int_{\mathbb{R}} \mathbf{P}\left[g\left((X_t(x))_{t\geq 0}\right) \in A\right] \mu(\mathrm{d}x) \qquad \textit{for } A \in \mathcal{E}\,.$$

2. *The particle system $\mathfrak{P}((X_t)_{t\geq 0}, \mu)$ is stationary if and only if the process–measure pair $((X_t)_{t\geq 0}, \mu)$ is invariant.*

3 Perturbation by random time change

Given the invariant process–measure pair $((L_t)_{t\geq 0}, \exp(-x)\mathrm{d}x)$ of a Lévy–Brown–Resnick process, we construct a new pair with different invariant intensity and dependence structure. Our approach is tailored to invariant measures having infinite mass and is based on random time change. The main result of this section, which is also used subsequently in the paper, is Theorem 3.7. Readers interested only in the main result may skip the remainder of the section and return to it later for the precise mathematical statements.

We first introduce the necessary definitions and assumptions. Then, we establish existence of time-changed processes and provide a useful representation of random clocks.

Definition 3.1 (Random clock)**.** Given a Lévy process $(L_t)_{t\geq 0}$ and a bounded function $\alpha : \mathbb{R} \to (0, \infty)$, we define

$$A_x(t) \;=\; \int_0^t \alpha(L_r + x)\,\mathrm{d}r \qquad \text{and} \qquad T_x(t) \;=\; \inf\{s \geq 0 \mid A_x(s) > t\}\,, \qquad t \geq 0\,, x \in \mathbb{R}\,.$$

We call $(T_x(t))_{t\geq 0}$ the random clock, $(A_x(t))_{t\geq 0}$ the inverse random clock, and α the mass function.

Note that A_x, for all $x \in \mathbb{R}$, is a strictly increasing, continuous random function, with inverse T_x, which we denote as $T_x = (A_x)^{-1}$.

Assumption 1 (Assumption on mass function)**.** The mass function α is continuous and satisfies

$$0 \;<\; \underline{\alpha} \;:=\; \inf_{z\in\mathbb{R}} \alpha(z) \;\leq\; \sup_{z\in\mathbb{R}} \alpha(z) \;=:\; \overline{\alpha} \;<\; \infty \qquad \text{and} \qquad \lim_{z\to\infty} \alpha(z) \;=\; 1\,. \tag{7}$$

Remark 3.2 (Assumption 1)**.** By virtue of the first part we obtain existence and infinite lifetime of the time-changed process that we are going to consider. The second part of the assumption is relevant only in Section 5, where convergence plays a key role.

Lemma 3.3 (Existence of time-changed process)**.** *Let $(L_t)_{t\geq 0}$ be a Lévy process and α a mass function satisfying Assumption 1. Then the time-changed process*

$$X_t(x) \;=\; L_{T_x(t)} \;+\; x\,, \qquad t \geq 0, x \in \mathbb{R}\,,$$

is a Markov process.

Proof. For the statement that the time-changed process is a Markov process with paths in $\mathbb{D}([0, \xi))$, where ξ is the (possibly random) lifetime of the process, see, for example, [BG68, (V.2.11)]. To finish the proof we show that the lifetime ξ is infinite, which follows from $T_x(t) \to \infty$ for $t \to \infty$ and all $x \in \mathbb{R}$. Since $T_x = A_x^{-1}$, this is equivalent to $A_x(t) \to \infty$ for $t \to \infty$ and all $x \in \mathbb{R}$. But this follows immediately from the lower bound on α in Assumption 1 by taking the limit $t \to \infty$ in

$$0 \;<\; t \cdot \underline{\alpha} \;\leq\; \int_0^t \alpha(L_r + x)\,\mathrm{d}r \;=\; A_x(t)\,.$$

□

Remark 3.4 (Random time change by continuous additive functionals)**.** The process $(X_t)_{t\geq 0}$ in Lemma 3.3 is an example of random time change by a continuous additive functional (CAF). For background on potential theory of CAFs, see [BG68, Chapter IV.3].

Let λ^+ be the Lebesgue measure on $\mathcal{B}([0,\infty))$, and denote the push-forward measure of λ^+ by A_x as $\lambda^+ \circ A_x^{-1}$. When precision of notation is not needed, we will use the equivalent notation $\mathrm{d}r$ instead of $\lambda^+(\mathrm{d}r)$.

Lemma 3.5 (Integral representation of random clock)**.** *Let $(L_t)_{t\geq 0}$ be a Lévy process, let α be a mass function satisfying Assumption 1, let A_x and T_x be the (inverse) random clock as in Definition 3.1, and let $(X_t)_{t\geq 0}$ be the time-changed process in the sense of Lemma 3.3. Then λ^+ is absolutely continuous with respect to $\lambda^+ \circ A_x^{-1}$ with density $r \mapsto \alpha(X_r(x))$ and*

$$T_x(t) \;=\; \int_0^t \frac{1}{\alpha(X_r(x))}\,\mathrm{d}r\,.$$

Proof. By the definition of the push-forward measure, the monotonicity of A_x, and $A_x^{-1} = T_x$, it holds for all $t \geq 0$

$$\begin{aligned}
\lambda^+([0,t]) \;=\; t \;=\; A_x(T_x(t)) \;&=\; \int_{[0,T_x(t)]} \alpha(L_r + x)\,\lambda^+(\mathrm{d}r)\\
&=\; \int_{[0,t]} \alpha(L_{T_x(r)} + x)\,(\lambda^+ \circ A_x^{-1})(\mathrm{d}r)\\
&=\; \int_{[0,t]} \alpha(X_r(x))\,(\lambda^+ \circ A_x^{-1})(\mathrm{d}r)\,.
\end{aligned}$$

Therefore λ^+ is absolutely continuous with respect to $\lambda^+ \circ A_x^{-1}$ with density $r \mapsto \alpha(X_r(x))$. It follows that

$$T_x(t) \;=\; \lambda^+[[0,T_x(t)]] \;=\; \left(\lambda^+ \circ A_x^{-1}\right)[[0,t]] \;=\; \int_{[0,t]} \frac{1}{\alpha(X_r(x))}\,\mathrm{d}r\,.$$

□

Assumption 2 (Assumption on Lévy process)**.** The process $(L_t)_{t\geq 0}$ is a Lévy process satisfying $\mathbf{E}[\exp(L_1)] = 1$.

Remark 3.6 (Assumption 2)**.** This assumption leads to a Lévy–Brown–Resnick process, as defined in the introduction. In the sequel we will frequently use that, for Lévy processes, $\mathbf{E}[\exp(L_1)] = 1$ implies $\mathbf{E}[\exp(L_t)] = 1$ for all $t \geq 0$. It is easy to see that the process–measure pair $((L_t)_{t\geq 0}, \exp(-x)\mathrm{d}x)$ is invariant. Indeed,

$$\begin{aligned}
&\int_{\mathbb{R}} \mathbf{P}[L_t + x \in B]\exp(-x)\,\mathrm{d}x\\
&=\; \mathbf{E}\left[\int_{\mathbb{R}} \mathbb{1}\{L_t + x \in B\}\exp(-x)\,\mathrm{d}x\right]\\
&=\; \mathbf{E}\left[\int_{\mathbb{R}} \mathbb{1}\{x \in B\}\exp(-(x - L_t))\,\mathrm{d}x\right] \;=\; \mathbf{E}\left[\exp(L_t)\right]\cdot\int_B \exp(-x)\,\mathrm{d}x \;=\; \int_B \exp(-x)\,\mathrm{d}x\,.
\end{aligned}$$

Theorem 3.7 (Perturbed invariant process–measure pair)**.** *Let α be a mass function satisfying Assumption 1, let $(L_t)_{t\geq 0}$ be a Lévy process satisfying Assumption 2, and let $(X_t)_{t\geq 0}$ be the time-changed process in the sense of Lemma 3.3. Then the measure $\alpha(x)\exp(-x)\mathrm{d}x$ is invariant for $(X_t)_{t\geq 0}$.*

Proof. Since $(X_t)_{t\geq 0}$ is a Markov process by Lemma 3.3, it follows from [ADR69, Proposition 2.1] that the invariance of the measure $\alpha(x)\exp(-x)\,\mathrm{d}x$ for $(X_t)_{t\geq 0}$ is equivalent to

$$\int_{\mathbb{R}}\left(\int_0^\infty \exp(-t)\cdot\mathbf{P}[X_t(x)\in B]\,\mathrm{d}t\right)\alpha(x)\exp(-x)\,\mathrm{d}x \;=\; \int_B \alpha(x)\exp(-x)\,\mathrm{d}x\,,$$

for all $B\in\mathcal{B}(\mathbb{R})$. To show this, note that $X_t(x)=L_{T_x(t)}+x$, and that

$$\begin{aligned}
&\int_0^\infty \exp(-t)\mathbb{1}\{L_{T_x(t)}+x\in B\}\,\lambda^+(\mathrm{d}t)\\
&\quad=\int_0^\infty \exp(-t)\mathbb{1}\{L_{T_x(t)}+x\in B\}\alpha(L_{T_x(t)}+x)\,(\lambda^+\circ A_x^{-1})(\mathrm{d}t)\\
&\quad=\int_0^\infty \exp(-A_x(t))\mathbb{1}\{L_t+x\in B\}\alpha(L_t+x)\,\lambda^+(\mathrm{d}t)\,,
\end{aligned}$$

where for the first equality we use that λ^+ is absolutely continuous with respect to $\lambda^+\circ A_x^{-1}$, with density $r\mapsto\alpha(X_r(x))=\alpha(L_{T_x(r)}+x)$ (see Lemma 3.5), and for the second equality we use the properties of push-forward measures and $A_x(0)=0$ and $\lim_{t\to\infty}A_x(t)=\infty$. Therefore, by Fubini-Tonelli and change of variables, it holds

$$\begin{aligned}
&\int_{\mathbb{R}}\exp(-x)\alpha(x)\left(\int_0^\infty \exp(-t)\cdot\mathbf{P}[X_t(x)\in B]\,\mathrm{d}t\right)\mathrm{d}x\\
&=\mathbf{E}\left[\int_{\mathbb{R}}\exp(-x)\alpha(x)\left(\int_0^\infty \exp(-t)\mathbb{1}\{L_{T_x(t)}+x\in B\}\,\mathrm{d}t\right)\mathrm{d}x\right]\\
&=\mathbf{E}\left[\int_{\mathbb{R}}\exp(-x)\alpha(x)\left(\int_0^\infty \exp(-A_x(t))\mathbb{1}\{L_t+x\in B\}\alpha(L_t+x)\,\mathrm{d}t\right)\mathrm{d}x\right]\\
&=\mathbf{E}\left[\int_0^\infty\left(\int_{\mathbb{R}}\exp(-x)\alpha(x)\exp\left(-\int_0^t\alpha(L_{t-s}+x)\,\mathrm{d}s\right)\mathbb{1}\{L_t+x\in B\}\alpha(L_t+x)\,\mathrm{d}x\right)\mathrm{d}t\right]\\
&=\mathbf{E}\left[\int_0^\infty\left(\int_{\mathbb{R}}\exp(-(x-L_t))\alpha(x-L_t)\exp\left(-\int_0^t\alpha(x-(L_t-L_{t-s}))\,\mathrm{d}s\right)\mathbb{1}\{x\in B\}\alpha(x)\,\mathrm{d}x\right)\mathrm{d}t\right]\\
&=\int_B\alpha(x)\exp(-x)\left(\int_0^\infty\mathbf{E}\left[\exp(L_t)\alpha(x-L_t)\exp\left(-\int_0^t\alpha(x-(L_t-L_{t-s}))\,\mathrm{d}s\right)\right]\mathrm{d}t\right)\mathrm{d}x\,,
\end{aligned}$$

where for the third equality we use that $\int_0^t\alpha(L_r+x)\,\mathrm{d}r = \int_0^t\alpha(L_{t-s}+x)\,\mathrm{d}s$, and for the fourth equality we make the change of variables $x\mapsto L_t+x$. For $t\geq 0$, define the process $(Y_s^t)_{s\in[0,t]}$ by $Y_s^t=L_t-L_{t-s}$. Then $(Y_s^t)_{s\in[0,t]}$ has the same finite-dimensional distributions as $(L_s)_{s\in[0,t]}$. Note that $L_t=L_t-L_{t-t}=Y_t^t$, so that

$$\begin{aligned}
&\int_0^\infty\mathbf{E}\left[\exp(L_t)\alpha(x-L_t)\exp\left(-\int_0^t\alpha(x-(L_t-L_{t-s}))\,\mathrm{d}s\right)\right]\mathrm{d}t\\
&\quad=\int_0^\infty\mathbf{E}\left[\exp(Y_t^t)\alpha(x-Y_t^t)\exp\left(-\int_0^t\alpha(x-(Y_s^t))\,\mathrm{d}s\right)\right]\mathrm{d}t\\
&\quad=\int_0^\infty\mathbf{E}\left[\exp(L_t)\alpha(x-L_t)\exp\left(-\int_0^t\alpha(x-L_s)\,\mathrm{d}s\right)\right]\mathrm{d}t\,.
\end{aligned}$$

Now, let $\widehat{A}_x$ and $\widehat{T}_x$ as in Definition 3.1, with the same mass function α as before, but with Lévy process $(-L_t)_{t\geq 0}$, that is,

$$\widehat{A}_x(t)\;=\;\int_0^t\alpha(x-L_r)\,\mathrm{d}r\qquad\text{and}\qquad\widehat{T}_x(t)\;=\;\left(\widehat{A}_x\right)^{-1}(t)\,.$$

It is easy to see that $\widehat{A}_x(0) = 0$ and $\lim_{t\to\infty} \widehat{A}_x(t) = \infty$, so that by the properties of push-forward measures and Lemma 3.5 it follows that

$$\begin{aligned}
&\int_0^\infty \exp(L_t)\alpha(x-L_t)\exp\left(-\widehat{A}_x(t)\right)\,\lambda^+(\mathrm{d}t) \\
&= \int_0^\infty \exp\left(L_{\widehat{T}_x(t)}\right)\alpha\left(x - L_{\widehat{T}_x(t)}\right)\exp(-t)\,\left(\lambda^+\circ\left(\widehat{A}_x\right)^{-1}\right)(\mathrm{d}t) \\
&= \int_0^\infty \exp\left(L_{\widehat{T}_x(t)}\right)\exp(-t)\,\lambda^+(\mathrm{d}t)\,.
\end{aligned}$$

Next, we apply optional stopping to $\exp(L_{\widehat{T}_x(t)})$, whence it follows

$$\mathbf{E}[\exp(L_{\widehat{T}_x(t)})] = \mathbf{E}[\exp(L_0)] = 1\,,$$

and therefore

$$\int_0^\infty \mathbf{E}\left[\exp(L_{\widehat{T}_x(t)})\right]\exp(-t)\,\mathrm{d}t \;=\; \int_0^\infty \exp(-t)\,\mathrm{d}t \;=\; 1\,,$$

which finishes the proof. To this end, we have to show that $\widehat{T}_x(t)$ is a bounded stopping time and that $(\exp(L_t))_{t\geq 0}$ is a martingale. Note that $\widehat{T}_x(t)$, by Lemma 3.5, is a stopping time with

$$\widehat{T}_x(t) \;=\; \int_0^t \frac{1}{\alpha(x - L_{\widehat{T}_x(r)})}\,\mathrm{d}r \;\leq\; t\cdot\frac{1}{\underline{\alpha}}\,,$$

and that for $0 \leq s \leq t$ it holds that

$$\mathbf{E}[\exp(L_t)\mid L_s] \;=\; \mathbf{E}[\exp(L_t - L_s)\exp(L_s)\mid L_s] \;=\; \mathbf{E}[\exp(L_{t-s})]\exp(L_s) \;=\; \exp(L_s)\,,$$

where the second equality follows from the Lévy property of $(L_t)_{t\geq 0}$ and the properties of conditional expectation, and the last equality follows from $\mathbf{E}[\exp(L_{t-s})] = 1$ for all $0 \leq s \leq t$. Therefore $(\exp(L_t))_{t\geq 0}$ is a martingale, and we can apply optional stopping to conclude the proof. $\square$

Remark 3.8. Another possibility of perturbing an invariant process–measure pair $((L_t)_{t\geq 0}, \exp(-x)\mathrm{d}x)$ of a Lévy–Brown–Resnick process is via the theory of Feller processes; see [BSW13, Chapter 4] for an overview of transformations of Feller processes. It is beyond the scope of this work to explore this direction.

4 Stationary max-id processes

In the special case where the intensity measure μ of an invariant process–measure pair $((X_t)_{t\geq 0}, \mu)$ has finite mass on sets bounded away from $-\infty$, the maximum over the corresponding particle system $\mathfrak{P}_t((X_t)_{t\geq 0}, \mu)$ is almost surely finite at all times $t \geq 0$. We call processes obtained in this way max-infinitely divisible (max-id), in the sense that they are componentwise maxima over PPP with points in $\mathbb{D}([0,\infty))$. More precisely, let α be a mass function satisfying Assumption 1, let $(L_t)_{t\geq 0}$ be a Lévy process satisfying Assumption 2, let $(X_t)_{t\geq 0}$ be the time-changed process in the sense of Lemma 3.3, and consider

$$Z_t \;=\; \max \mathfrak{P}_t((X_t)_{t\geq 0}, \alpha(x)\exp(-x)\mathrm{d}x) \;=\; \max_{i\in\mathbb{N}} X_t^{(i)}(U^{(i)})\,,$$

where $\{U^{(i)}\}_{i\in\mathbb{N}}$ is a PPP on the real line with intensity measure $\alpha(x)\exp(-x)\,\mathrm{d}x$, and $(X_t^{(i)})_{t\geq 0}$ are i.i.d. copies of $(X_t)_{t\geq 0}$.

Remark 4.1 (PPP from jump times of Poisson process)**.** Such a PPP $\{U^{(i)}\}_{i\in\mathbb{N}}$ can be obtained by a monotone transformation of the jump times of a unit Poisson process on the real line. To be more precise,

$$U^{(i)} \;=\; F^{\leftarrow}(\exp(-\Gamma_i))\,, \qquad (\Gamma_i)_{i\in\mathbb{N}} \text{ jump times of unit Poisson process}\,,$$

where $F^{\leftarrow}$ is the generalized inverse of F, that is $F^{\leftarrow}(q) = \inf\{x \in \mathbb{R} \mid F(x) \geq q\}$. Therefore, the natural ordering of the points is preserved, where in our case, we can write without loss of generality $U^{(1)} = \max\{U^{(i)}\}_{i\in\mathbb{N}}$, so that

$$\begin{aligned}\mathbf{P}[U^{(1)} \leq x] \;&=\; \mathbf{P}[F^{\leftarrow}(\exp(-\Gamma_1)) \leq x] \;=\; \mathbf{P}[\Gamma_1 \geq -\log(F(x))] \;=\; \exp(-(-\log(F(x))))\\ &=\; F(x)\,.\end{aligned}$$

After a technical lemma on exceedance probabilities for random time-changed processes, the main result of this section (Theorem 4.3) shows that the process $(Z_t)_{t\geq 0}$ has paths in $\mathbb{D}([0,\infty))$ and is stationary.

Lemma 4.2 (Bound on exceedance probability)**.** *Let α be a mass function satisfying Assumption 1, let $(L_t)_{t\geq 0}$ be a Lévy process satisfying Assumption 2, and let $(X_t)_{t\geq 0}$ be the time-changed process in the sense of Lemma 3.3. Then, for all $t \geq 0$, there exists $v > 0$ and $C_{v,t} > 0$ such that for all $u, x \in \mathbb{R}$ satisfying $u - x > v$ it holds that*

$$\mathbf{P}\left[\sup_{s\in[0,t]} X_s(x) \;\geq\; u\right] \;\leq\; C_{v,t}\cdot \mathbf{P}\left[L_{t\cdot\overline{\alpha}} \geq u - x - v\right]\,.$$

Proof. Denote

$$\tau(u,x) \;:=\; \inf\{t\geq 0 \mid L_t + x \geq u\} \qquad\text{and}\qquad \widetilde{\tau}(u,x) \;:=\; \inf\{t \geq 0 \mid L_{T_x(t)} + x \geq u\}\,,$$

that is, the first hitting times of the level $u \in \mathbb{R}$ of $(L_t)_{t\geq 0}$ and $(X_t)_{t\geq 0}$ respectively. Clearly it holds that

$$T_x(\tau(u,x)) \;=\; \widetilde{\tau}(u,x)\,.$$

Since $A_x(t) = (T_x)^{-1}(t)$ is monotone it holds that

$$\widetilde{\tau}(u,x) \;\leq\; t \qquad\text{is equivalent to}\qquad \tau(u,x) \;\leq\; A_x(t)\,.$$

Then by Assumption 1

$$A_x(t) \;=\; \int_0^t \alpha(L_r + x)\,\mathrm{d}r \;\leq\; t\cdot\overline{\alpha}\,. \tag{8}$$

Now choose $v > 0$ large enough, such that

$$0 \;<\; \mathbf{P}\left[\inf_{s\in[0,t\cdot\overline{\alpha}]} L_s \;\geq\; -v\right]\,. \tag{9}$$

It follows from (8) that

$$\begin{aligned}&\mathbf{P}\left[\sup_{s\in[0,t]} X_s(x) \;\geq\; u\right]\\ &= \mathbf{P}\left[\sup_{s\in[0,t]} L_{T_x(s)} + x \;\geq\; u\right] \;=\; \mathbf{P}[\widetilde{\tau}(u,x) \;\leq\; t] \;=\; \mathbf{P}[\tau(u,x) \;\leq\; A_x(t)]\\ &\leq \mathbf{P}[\tau(u,x) \;\leq\; t\cdot\overline{\alpha}]\\ &= \mathbf{P}\left[\sup_{s\in[0,t\cdot\overline{\alpha}]} L_s + x \;\geq\; u\right]\,.\end{aligned} \tag{10}$$

By [Wil87, (2.1)] it holds for $u, x \in \mathbb{R}$ satisfying $u - x > v$ that

$$\mathbf{P}\left[\sup_{s\in[0,t\cdot\overline{\alpha}]} L_s + x \;\geq\; u\right] \;\leq\; \frac{\mathbf{P}\left[L_{t\cdot\overline{\alpha}} \;\geq\; u-x-v\right]}{\mathbf{P}\left[\inf_{s\in[0,t\cdot\overline{\alpha}]} L_s \;\geq\; -v\right]} \;=\; C_{v,t}\cdot\mathbf{P}\left[L_{t\cdot\overline{\alpha}} \;\geq\; u-x-v\right], \tag{11}$$

where, by (9),

$$C_{v,t} \;:=\; \frac{1}{\mathbf{P}\left[\inf_{s\in[0,t\cdot\overline{\alpha}]} L_s \;\geq\; -v\right]} \;<\; \infty\,.$$

Putting together (10) and (11) finishes the proof. □

Theorem 4.3 (Existence and stationarity of max-id processes)**.** *Let α be a mass function satisfying Assumption 1, let $(L_t)_{t\geq0}$ be a Lévy process satisfying Assumption 2, and let $(X_t)_{t\geq0}$ be the time-changed process in the sense of Lemma 3.3, and consider the process $(Z_t)_{t\geq0}$ defined by*

$$Z_t \;=\; \max\mathfrak{P}_t((X_t)_{t\geq0}, \alpha(x)\exp(-x)\mathrm{d}x)\,.$$

Then the process $(Z_t)_{t\geq0}$ has paths in $\mathbb{D}([0,\infty))$ and is stationary. If, furthermore, $(L_t)_{t\geq0}$ has continuous paths, then $(Z_t)_{t\geq0}$ has paths in $\mathbb{C}([0,\infty))$.

Proof. We adapt the proof of [EK15, Proposition 1.14] to our setting. To this end, it is sufficient to prove that $(Z_s)_{s\in[0,t]}$ has paths in $\mathbb{D}([0,t])$ for all $t>0$. Fix $t>0$ and $u\in\mathbb{R}$, and consider

$$I_u \;:=\; \left\{ i\in\mathbb{N} \;\middle|\; \sup_{s\in[0,t]} X_s^{(i)}(U^{(i)}) \;\geq\; u \right\}.$$

This set contains the indices of particles that somewhere in the compact interval $[0,t]$ have values above the threshold u and hence are the only particles that, in case that $\inf_{s\in[0,t]} Z_s \geq u$, could contribute to the maximum. Applying Theorem 2.3 we get that the cardinality of the set I_u has Poisson distribution, which, by an application of Lemma 4.2, has finite rate:

$$\begin{aligned}\lambda_u \;&:=\; \int_{\mathbb{R}} \mathbf{P}\left[\sup_{s\in[0,t]} X_s(x) \;\geq\; u\right]\alpha(x)\exp(-x)\,\mathrm{d}x \\ &\leq\; \overline{\alpha}\cdot\left(\int_{-\infty}^{u-v}\mathbf{P}\left[\sup_{s\in[0,t]} X_s(x) \;\geq\; u\right]\exp(-x)\,\mathrm{d}x \;+\; \exp(-(u-v))\right),\end{aligned}$$

where we chose $v>0$ as in Lemma 4.2. Applying Lemma 4.2 gives

$$\begin{aligned}&\int_{-\infty}^{u-v}\mathbf{P}\left[\sup_{s\in[0,t]} X_s(x) \;\geq\; u\right]\exp(-x)\,\mathrm{d}x \\ &\leq\; C_{v,t}\cdot\int_{-\infty}^{u-v}\mathbf{P}[L_{t\cdot\overline{\alpha}} \geq u-x-v]\exp(-x)\,\mathrm{d}x \\ &=\; C_{v,t}\cdot\exp(v-u)\cdot\int_{-\infty}^{u-v}\mathbf{P}[\exp(L_{t\cdot\overline{\alpha}}) \geq \exp(u-x-v)]\exp(u-x-v)\,\mathrm{d}x \\ &=\; C_{v,t}\cdot\exp(v-u)\cdot\int_{1}^{\infty}\mathbf{P}[\exp(L_{t\cdot\overline{\alpha}}) \geq y]\,\mathrm{d}y \\ &\leq\; C_{v,t}\cdot\exp(v-u)\cdot\mathbf{E}[\exp(L_{t\cdot\overline{\alpha}})] \;=\; C_{v,t}\cdot\exp(v-u)\,.\end{aligned}$$

It follows $\lambda_u < \infty$, and therefore $|I_u| < \infty$ almost surely. Denote $J_u := \{\inf_{s\in[0,t]} Z_s \geq u\}$, and note that $\mathbf{P}[\cup_{u\in\mathbb{Z}} J_u] = 1$. If J_u for some $u \in \mathbb{Z}$, it therefore holds

$$Z_s \;=\; \max_{i\in I_u} X_s^{(i)}(U^{(i)}) \qquad \text{for all } s \in [0,t]\,.$$

Since $(X_s)_{s\in[0,t]}$ has paths in $\mathbb{D}([0,t])$, and $\mathbb{D}([0,t])$ is invariant under taking the maximum over finitely many functions, we get that $(Z_s)_{s\in[0,t]}$ has paths in $\mathbb{D}([0,t])$. If $(L_t)_{t\geq 0} \in \mathbb{C}([0,\infty))$, then $(X_t)_{t\geq 0} \in \mathbb{C}([0,\infty))$ and therefore $(Z_t)_{t\geq 0} \in \mathbb{C}([0,\infty))$ for the same reason. The stationarity of $(Z_t)_{t\geq 0}$ follows then immediately from $(Z_t)_{t\geq 0}$ being a well-defined process, the invariance of the process–measure pair $((X_t)_{t\geq 0}, \alpha(x)\exp(-x)\,\mathrm{d}x)$ by Theorem 3.7, and Theorem 2.3(ii). This finishes the proof. □

5 Max-domain of attraction

In this final section we study the extremal properties of the max-id processes constructed in Section 4 through max-domains of attraction. Let α be a mass function satisfying Assumption 1, let $(L_t)_{t\geq 0}$ be a Lévy process satisfying Assumption 2, let $(X_t)_{t\geq 0}$ be the time-changed process in the sense of Lemma 3.3, and let $(Z_t)_{t\geq 0}$ be the max-id process in the sense of Section 4. Define, for $n \in \mathbb{N}$, the rescaled process $(\zeta_n(t))_{t\geq 0}$, by

$$\zeta_n(t) \;:=\; \max_{k=1,\ldots,n} Z_t^{(k)} - \log(n)\,,$$

where $((Z_t^{(k)})_{t\geq 0}), k = 1, \ldots, n$, are i.i.d. copies of $(Z_t)_{t\geq 0}$, and define the process $(\zeta_L(t))_{t\geq 0}$ by

$$\zeta_L(t) \;:=\; \max \mathfrak{P}_t((L_t)_{t\geq 0}, \exp(-x)\mathrm{dx})\,,$$

where $(L_t)_{t\geq 0}$ is the Lévy process that was time-changed to obtain $(X_t)_{t\geq 0}$. We show that $(Z_t)_{t\geq 0}$ is in the max-domain of attraction of the max-stable process $(\zeta_L(t))_{t\geq 0}$ in the sense that $(\zeta_n(t))_{t\geq 0}$ converges weakly to $(\zeta_L(t))_{t\geq 0}$.

Remark 5.1. The weakest notion of convergence is convergence in finite-dimensional distributions (f.d.d), while in certain cases this can be strengthened to convergence in the space of càdlàg functions $\mathbb{D}([0,\infty))$, endowed with the Skorokhod topology.

The main result of this section is Theorem 5.3, where we show that $(\zeta_n(t))_{t\geq 0}$ converges weakly to $(\zeta_L(t))_{t\geq 0}$, that is, $(Z_t)_{t\geq 0}$ is in the MDA of ζ_L. The following auxiliary result isolates required convergence statements that we will use in the proof of Theorem 5.3.

Lemma 5.2. *Let α be a mass function satisfying Assumption 1, let $(L_t)_{t\geq 0}$ be a Lévy process satisfying Assumption 2, let $(X_t)_{t\geq 0}$ be the time-changed process in the sense of Lemma 3.3, and write*

$$\eta(x) \;:=\; \alpha(x)\exp(-x) \qquad \textit{and} \qquad \eta_n(x) \;:=\; n \cdot \eta(x + \log(n))\,.$$

As $n \to \infty$, we have

1.

$$\begin{aligned} \eta_n(x) \;&\to\; \exp(-x) \qquad \forall x \in \mathbb{R}\,, \\ \int_z^\infty \eta_n(x)\,\mathrm{d}x \;&\to\; \exp(-z) \qquad \forall z \in \mathbb{R}\,, \end{aligned} \tag{12}$$

2.

$$X_t^n(x) \ := \ X_t(x+\log(n)) - \log(n) \ \to \ L_t \qquad \textit{weakly in} \qquad \mathbb{D}([0,\infty)) \,. \tag{13}$$

Proof. The first part follows from Assumption 1 and dominated convergence. To show the second part, we apply [EK86, Theorem 1.5]. To this end, note that

$$X_t^n(x) \ = \ X_t(x+\log(n)) \ - \ \log(n) \ = \ L_{T_{x+\log(n)}(t)} + x \,,$$

where

$$T_{x+\log(n)}(t) \ = \ \int_0^t \frac{1}{\alpha\big(X_s(x+\log(n))\big)}\,\mathrm{d}s \ = \ \int_0^t \frac{1}{\alpha\left(X_s^n(x)+\log(n)\right)}\,\mathrm{d}s \,,$$

and by Assumption 1

$$\lim_{n\to\infty}\sup_{z\in K}\left|\frac{1}{\alpha(z+\log(n))} \ - \ 1\right| \ = \ 0 \,,$$

for all compact sets $K \subset \mathbb{R}$. Since α is continuous by Assumption 1, this allows us to apply [EK86, Theorem 1.5], which finishes the proof. $\square$

Theorem 5.3 (max-domain of attraction of the constructed max-id process)**.** *Let α be a mass function satisfying Assumption 1, let $(L_t)_{t\geq 0}$ be a Lévy process satisfying Assumption 2, and let $(X_t)_{t\geq 0}$ be the time-changed process in the sense of Lemma 3.3. Consider the max-id process $(Z_t)_{t\geq 0}$ defined by*

$$Z_t \ = \ \max \mathfrak{P}_t\left((X_t)_{t\geq 0}, \alpha(x)\exp(-x)\mathrm{d}x\right) \,,$$

let

$$\zeta_n(t) \ := \ \max_{k=1,\ldots,n} Z_t^{(k)} - \log(n) \,, \qquad n \in \mathbb{N} \,,$$

where $((Z_t^{(k)})_{t\geq 0})_{k=1,\ldots,n}$ are i.i.d. copies of $(Z_t)_{t\geq 0}$, and let

$$\zeta_L(t) \ := \ \max \mathfrak{P}_t((L_t)_{t\geq 0}, \exp(-x)\mathrm{d}x) \,.$$

Then ζ_n converges weakly to ζ_L in f.d.d. If, furthermore, $(L_t)_{t\geq 0}$ has continuous paths, then convergence holds in $\mathbb{D}([0,\infty))$.

Remark 5.4. The only Lévy processes with continuous paths satisfying Assumption 2 are $(\sqrt{\sigma}(B_t - \sqrt{\sigma}t/2))_{t\geq 0}$, for $\sigma \geq 0$. For $\sigma = 1$ we get the classical Brown–Resnick process ζ_{BR} from the introduction. In view of [LdH01, Theorem 2.4], it is natural that convergence in $\mathbb{D}([0,\infty))$ can only hold when the limit is continuous. A concrete reason to support this is that for the extended modulus of continuity

$$\omega''(f,\delta,K) \ := \ \sup_{\substack{t_1,t_2\in K \\ t_1\leq s\leq t_2\,, |t_1-t_2|<\delta}} |f(s)-f(t_1)| \wedge |f(s)-f(t_2)| \,,$$

we cannot proceed as in (16).

Proof f.d.d. Let $\ell \in \mathbb{N}$, and $0 \le t_1 < \ldots, < t_\ell < \infty$. We want to show that

$$(\zeta_n(t_1), \ldots, , \zeta_n(t_\ell)) \to (\zeta_L(t_1), \ldots, , \zeta_L(t_\ell)) \qquad \text{in distribution as } n \to \infty . \tag{14}$$

To this end, let $x_1, \ldots, x_\ell \in \mathbb{R}$ and let η and η_n be defined as in Lemma 5.2. Note that since ζ_n is the maximum over n independent and identically distributed max-id processes, it holds that

$$\begin{aligned}
&\mathbf{P}\left[\forall_{j\in\{1,\ldots,\ell\}}\ \zeta_n(t_j) \ \le\ x_j\right] \\
&= \mathbf{P}\left[\forall_{j\in\{1,\ldots,\ell\},k\in\{1,\ldots,n\}}\ Z_{t_j}^{(k)} \ \le\ x_j + \log(n)\right] \\
&= \left(\mathbf{P}\left[\forall_{j\in\{1,\ldots,\ell\}}\ \max_{i\in\mathbb{N}} X_{t_j}^{(i)}(U^{(i)}) \ \le\ x_j + \log(n)\right]\right)^n \\
&= \left(\mathbf{P}\left[\sum_{i\in\mathbb{N}} \mathbb{1}\{\exists_{j\in\{1,\ldots,\ell\}}\ X_{t_j}^{(i)}(U^{(i)}) > x_j + \log(n)\} \ =\ 0\right]\right)^n \\
&= \exp\left(-n\int_{\mathbb{R}} \mathbf{P}[\exists_{j\in\{1,\ldots,\ell\}}\ X_{t_j}(x) > x_j + \log(n)]\eta(x)\,\mathrm{d}x\right) ,
\end{aligned}$$

where the last step follows from Theorem 2.3(i). Applying the inclusion-exclusion formula and Tonelli's theorem, we get

$$\begin{aligned}
&\int_{\mathbb{R}} \mathbf{P}[\exists_{j\in\{1,\ldots,\ell\}}\ X_{t_j}(x) > x_j + \log(n)] \cdot \eta(x)\,\mathrm{d}x \\
&= \sum_{l=1}^{\ell} (-1)^{l+1} \sum_{\substack{I\subseteq\{1,\ldots,\ell\} \\ |I|=l}} \int_{\mathbb{R}} \mathbf{P}[\forall_{j\in I}\ X_{t_j}(x) > x_j + \log(n)] \cdot \eta(x)\,\mathrm{d}x .
\end{aligned}$$

By invariance of the process measure pair $((X_t)_{t\ge 0}, \eta)$ (see Theorem 3.7) we get for fixed $I \subseteq \{1, \ldots, \ell\}$ that

$$\begin{aligned}
&n\int_{\mathbb{R}} \mathbf{P}[\forall_{j\in I}\ X_{t_j}(x) > x_j + \log(n)] \cdot \eta(x)\,\mathrm{d}x \\
&= n\int_{x_{\min\{j\in I\}}+\log(n)}^{\infty} \mathbf{P}[\forall_{j\in I}\ X_{t_j - t_{\min\{j\in I\}}}(x) > x_j + \log(n)] \cdot \eta(x)\,\mathrm{d}x \\
&= \int_{x_{\min\{j\in I\}}}^{\infty} \mathbf{P}[\forall_{j\in I}\ X^n_{t_j - t_{\min\{j\in I\}}}(x) > x_j] \cdot \eta_n(x)\,\mathrm{d}x \\
&\to \int_{x_{\min\{j\in I\}}}^{\infty} \mathbf{P}[\forall_{j\in I}\ (L_{t_j - t_{\min\{j\in I\}}} + x) > x_j] \cdot \exp(-x)\,\mathrm{d}x .
\end{aligned}$$

The convergence follows from

$$\mathbf{P}[\forall_{j\in I}\ X^n_{t_j - t_{\min\{j\in I\}}}(x) > x_j] \cdot \eta_n(x) \ \le\ \eta_n(x) \ \le\ \bar{a}\exp(-x) \qquad \text{for all } x \in \mathbb{R} ,$$

Lemma 5.2 and dominated convergence. Using the invariance of $((L_t)_{t\ge 0}, \exp(-x)\mathrm{d}x)$ we reverse the time shift, and reversing the previous steps as well, we find that this expression equals

$$\mathbf{P}\left[\forall_{j\in\{1,\ldots,\ell\}}\ \zeta_L(t_j) \ \le\ x_j\right] .$$

Since what we showed is equivalent to (14), this finishes the proof. □

Proof of convergence in $\mathbb{D}([0,\infty))$. By assumption, $(L_t)_{t\geq 0} \in \mathbb{C}([0,\infty))$. Theorem 4.3 then implies that ζ_L, $(Z_t)_{t\geq 0}$ and ζ_n have paths in $\mathbb{C}([0,\infty))$. Hence, it suffices to prove convergence in $\mathbb{C}([0,\infty))$.

We prove convergence by establishing tightness. First note that $(\zeta_n(0))_{n\in\mathbb{N}}$ converges weakly to a Gumbel distributed random variable and is therefore tight. By [Bil99, Theorem 7.3], showing tightness of ζ_n then amounts to controlling the modulus of continuity ω. Let $K \subset [0,\infty)$ be an arbitrary compact set (possibly not containing 0). For $f \in \mathbb{C}(K)$ and $\delta > 0$ we define the modulus of continuity by

$$\omega\left(f,\delta,K\right) \;:=\; \sup_{\substack{s,t\in K\\ |s-t|<\delta}} \left|f(t)-f(s)\right| \,.$$

Our goal is to show that for all $\varepsilon, a > 0$ there exists some $\delta > 0$ and $N \in \mathbb{N}$ such that

$$\mathbf{P}[\omega(\zeta_n,\delta,K) > a] < \varepsilon \qquad \text{for all } n \;>\; N\,. \tag{15}$$

Together with the f.d.d. convergence, we get convergence in $\mathbb{C}(K)$ for an arbitrary compact set $K \subset [0,\infty)$ and therefore convergence in $\mathbb{C}[0,\infty)$. We adapt parts of the proof of [KSd09, Theorem 17] to our setting. Write

$$Y_{k,n}(t) \;:=\; Z_t^{(k)} - \log(n)\,, \qquad t \;\geq\; 0\,.$$

We structure the proof in several steps.

Decomposition

We decompose the event $\{\omega\left(\zeta_n,\delta,K\right) \;>\; a\}$. Note that, by the properties of maxima and the invariance of the modulus of continuity ω under spatial shifts, we have

$$\begin{aligned}\omega\left(\zeta_n,\delta,K\right) \;&=\; \sup_{\substack{s,t\in K\\ |s-t|<\delta}} \left|\max_{k=1,\dots,n} Z_t^{(k)} - \max_{k=1,\dots,n} Z_s^{(k)}\right| \\ &\leq\; \max_{k=1,\dots,n} \sup_{\substack{s,t\in K\\ |s-t|<\delta}} \left|Z_t^{(k)} - Z_s^{(k)}\right| \;=\; \max_{k=1,\dots,n} \omega(Y_{k,n},\delta,K)\,.\end{aligned} \tag{16}$$

Therefore

$$
\begin{aligned}
&\{\omega(\zeta_n, \delta, K) \;>\; a\} \\
&= \left\{ \omega(\zeta_n, \delta, K) \;>\; a \quad \text{and} \quad \exists_{t \in K}\ \zeta_n(t) \neq \max_{\substack{k=1,\dots,n \\ Y_{k,n}(0) > -c_2}} Y_{k,n}(t) \right\} \\
&\qquad \cup \left\{ \omega(\zeta_n, \delta, K) \;>\; a \quad \text{and} \quad \forall_{t \in K}\ \zeta_n(t) = \max_{\substack{k=1,\dots,n \\ Y_{k,n}(0) > -c_2}} Y_{k,n}(t) \right\} \\
&\subset \left\{ \exists_{t \in K}\ \zeta_n(t) \neq \max_{\substack{k=1,\dots,n \\ Y_{k,n}(0) > -c_2}} Y_{k,n}(t) \right\} \\
&\qquad \cup \left\{ \omega(\zeta_n, \delta, K) \;>\; a \quad \text{and} \quad \forall_{t \in K}\ \zeta_n(t) = \max_{\substack{k=1,\dots,n \\ Y_{k,n}(0) > -c_2}} Y_{k,n}(t) \quad \text{and} \quad \exists_{k \in \{1,\dots,n\}}\ Y_{k,n}(0) > -c_2 \right\} \\
&\subset \left\{ \exists_{t \in K}\ \zeta_n(t) \neq \max_{\substack{k=1,\dots,n \\ Y_{k,n}(0) > -c_2}} Y_{k,n}(t) \right\} \cup \bigcup_{k=1}^{n} \{\omega(Y_{k,n}, \delta, K) \;>\; a \quad \text{and} \quad Y_{k,n}(0) > -c_2\} \\
&=: G_n \cup \bigcup_{k=1}^{n} C_{k,n}\,,
\end{aligned}
\tag{17}
$$

where in the last step we used (16). Next we define and estimate

$$
\begin{aligned}
E_n &:= \left\{ \inf_{t \in K} \zeta_n(t) \;<\; -c_1 \right\} \\
&\subset \bigcap_{k=1}^{n} \left\{ \inf_{t \in K} Y_{k,n}(t) \;<\; -c_1 \right\} \\
&\subset \bigcap_{k=1}^{n} \left\{ \inf_{t \in K} Y_{k,n}(t) \;<\; -c_1 \qquad \text{or} \qquad Y_{k,n}(0) \;<\; -c_1 \right\} \;=:\; \bigcap_{k=1}^{n} A_{k,n}\,.
\end{aligned}
\tag{18}
$$

With E_n defined in this way, we have

$$
\begin{aligned}
G_n &\subset E_n \cup \left\{ \exists_{t^* \in K, k \in \{1,\dots,n\}}\ \zeta_n(t^*) \;=\; Y_{k,n}(t^*)\,, \quad Y_{k,n}(0) \le -c_2 \quad \text{and} \quad \inf_{t \in K} \zeta_n(t) \;\ge\; -c_1 \right\} \\
&\subset E_n \cup \bigcup_{k=1}^{n} \left\{ Y_{k,n}(0) \;\le\; -c_2 \qquad \text{and} \qquad \sup_{t \in K} Y_{k,n}(t) \;\ge\; -c_1 \right\} \;=:\; E_n \cup \bigcup_{k=1}^{n} B_{k,n}\,.
\end{aligned}
\tag{19}
$$

Using (17), (18), and (19), we get

$$
\{\omega(\zeta_n, \delta, K) \;>\; a\} \;\subset\; \bigcap_{k=1}^{n} A_{k,n} \cup \bigcup_{k=1}^{n} B_{k,n} \cup \bigcup_{k=1}^{n} C_{k,n}\,.
$$

Since the events are i.i.d. for $k \in \{1,\dots,n\}$, we have

$$
\mathbf{P}\left[\omega(\zeta_n, \delta, K) \;>\; a\right] \;\le\; (\mathbf{P}[A_{1,n}])^n \;+\; 1 \;-\; \left(\mathbf{P}[B_{1,n}^c]\right)^n \;+\; n\mathbf{P}[C_{1,n}]\,, \tag{20}
$$

where

$$
\begin{aligned}
A_{1,n} &= \left\{ \inf_{t\in K} Y_{1,n}(t) \;<\; -c_1 \qquad \text{or} \qquad Y_{1,n}(0) \;<\; -c_1 \right\}, \\
B_{1,n}^c &= \left\{ Y_{1,n}(0) \;>\; -c_2 \qquad \text{or} \qquad \sup_{t\in K} Y_{1,n}(t) \;<\; -c_1 \right\}, \\
C_{1,n} &= \left\{ \omega\left(Y_{1,n}, \delta, K\right) \;>\; a \quad \text{and} \quad Y_{1,n}(0) > -c_2 \right\}.
\end{aligned}
$$

Analysis of $(\mathbf{P}[A_{1,n}])^n$
The goal is to show that there exists c_1 such that, for $N = N(c_1)$ sufficiently large, we have

$$
(\mathbf{P}[A_{1,n}])^n \;\leq\; \frac{\varepsilon}{4} \qquad \text{for all } n \geq N(c_1)\,. \tag{21}
$$

Observe that

$$
\begin{aligned}
A_{1,n} &= \left\{ \inf_{t\in K} \max_{i\in\mathbb{N}} X_t^{(i)}(U^{(i)}) \;<\; -c_1 + \log(n) \qquad \text{or} \qquad \max_{i\in\mathbb{N}} U^{(i)} \;<\; -c_1 + \log(n) \right\} \\
&\subset \left\{ \max_{i\in\mathbb{N}} \inf_{t\in K} X_t^{(i)}(U^{(i)}) \;<\; -c_1 + \log(n) \qquad \text{or} \qquad \max_{i\in\mathbb{N}} U^{(i)} \;<\; -c_1 + \log(n) \right\} \\
&= \left\{ \forall_{i\in\mathbb{N}} \inf_{t\in K} X_t^{(i)}(U^{(i)}) \;<\; -c_1 + \log(n) \qquad \text{or} \qquad \forall_{i\in\mathbb{N}} U^{(i)} \;<\; -c_1 + \log(n) \right\} \\
&= \bigcap_{i\in\mathbb{N}} \left\{ \inf_{t\in K} X_t^{(i)}(U^{(i)}) \;<\; -c_1 + \log(n) \right\} \cup \bigcap_{i\in\mathbb{N}} \left\{ U^{(i)} \;<\; -c_1 + \log(n) \right\} \\
&\subset \bigcap_{i\in\mathbb{N}} \left\{ \inf_{t\in K} X_t^{(i)}(U^{(i)}) \;<\; -c_1 + \log(n) \qquad \text{or} \qquad U^{(i)} \;<\; -c_1 + \log(n) \right\} \\
&= \bigcap_{i\in\mathbb{N}} \left\{ \inf_{t\in K} X_t^{(i)}(U^{(i)}) \;\geq\; -c_1 + \log(n) \qquad \text{and} \qquad U^{(i)} \;\geq\; -c_1 + \log(n) \right\}^c \\
&= \left\{ \sum_{i\in\mathbb{N}} \mathbb{1} \left\{ \inf_{t\in K} X_t^{(i)}(U^{(i)}) \;\geq\; -c_1 + \log(n) \qquad \text{and} \qquad U^{(i)} \;\geq\; -c_1 + \log(n) \right\} \;=\; 0 \right\},
\end{aligned}
$$

so that

$$
\begin{aligned}
&(\mathbf{P}[A_{1,n}])^n \\
&\leq \left(\mathbf{P}\left[\sum_{i\in\mathbb{N}} \mathbb{1} \left\{ \inf_{t\in K} X_t^{(i)}(U^{(i)}) - \log(n) \;\geq\; -c_1 \qquad \text{and} \qquad U^{(i)} \;\geq\; \log(n) - c_1 \right\} \;=\; 0 \right] \right)^n.
\end{aligned} \tag{22}
$$

To further bound this, note that, by Theorem 2.3 and the change of variables $x \mapsto x + \log(n)$,

$$
\begin{aligned}
&\left(\mathbf{P}\left[\sum_{i\in\mathbb{N}} \mathbb{1} \left\{ \inf_{t\in K} X_t^{(i)}(U^{(i)}) - \log(n) \;\geq\; -c_1 \qquad \text{and} \qquad U^{(i)} \;\geq\; \log(n) - c_1 \right\} \;=\; 0 \right] \right)^n \\
&= \exp\left(-n \int_{\log(n)-c_1}^{\infty} \mathbf{P}\left[\inf_{t\in K} X_t(x) - \log(n) \;\geq\; -c_1 \right] \eta(x)\,\mathrm{d}x \right) \\
&= \exp\left(- \int_{-c_1}^{\infty} \mathbf{P}\left[\inf_{t\in K} X_t^n(x) \;\geq\; -c_1 \right] \eta_n(x)\,\mathrm{d}x \right),
\end{aligned} \tag{23}
$$

where

$$X_t^n(x) \;=\; X_t(x+\log(n)) - \log(n)\,.$$

To apply weak convergence of $(X_t^n)_{t\geq 0}$ to $(L_t)_{t\geq 0}$ (see Lemma 5.2(ii)), we write

$$\begin{aligned}
&\exp\left(-\int_{-c_1}^{\infty} \mathbf{P}\left[\inf_{t\in K} X_t^n(x) \;\geq\; -c_1\right]\eta_n(x)\,\mathrm{d}x\right)\\
&\leq\; \exp\left(-\int_{-c_1}^{\infty} \mathbf{P}\left[\inf_{t\in K} L_t + x \;\geq\; -c_1\right]\exp(-x)\,\mathrm{d}x\right)\\
&\quad+\;\left|\exp\left(-\int_{-c_1}^{\infty} \mathbf{P}\left[\inf_{t\in K} X_t^n(x) \;\geq\; -c_1\right]\eta_n(x)\,\mathrm{d}x\right)\right.\\
&\qquad\left.-\;\exp\left(-\int_{-c_1}^{\infty} \mathbf{P}\left[\inf_{t\in K} L_t + x \;\geq\; -c_1\right]\exp(-x)\,\mathrm{d}x\right)\right|\,.
\end{aligned} \tag{24}$$

For the first term of the upper bound in (24), making the change of variables $x \mapsto x + c_1$ we get

$$\begin{aligned}
&\int_{-c_1}^{\infty} \mathbf{P}\left[\inf_{t\in K} L_t + x \;\geq\; -c_1\right]\exp(-x)\,\mathrm{d}x\\
&\quad=\;\exp(c_1)\int_{0}^{\infty} \mathbf{P}\left[\inf_{t\in K} L_t + x \;\geq\; 0\right]\exp(-x)\,\mathrm{d}x \;>\; 0\,,
\end{aligned}$$

so that, for c_1 large enough, it holds

$$\exp\left(-\int_{-c_1}^{\infty} \mathbf{P}\left[\inf_{t\in K} L_t + x \;\geq\; -c_1\right]\exp(-x)\,\mathrm{d}x\right) \;\leq\; \frac{\varepsilon}{8}\,.$$

For the second term of the upper bound in (24), note that analogously to the convergence proof for f.d.d. we get that

$$\int_{-c_1}^{\infty} \mathbf{P}\left[\inf_{t\in K} X_t^n(x) \;\geq\; -c_1\right]\eta_n(x)\,\mathrm{d}x \;\to\; \int_{-c_1}^{\infty} \mathbf{P}\left[\inf_{t\in K} L_t + x \;\geq\; -c_1\right]\exp(-x)\,\mathrm{d}x$$

for all $c_1 \in \mathbb{R}$ and as $n \to \infty$. We can therefore choose $n \geq N = N(c_1)$ large enough, such that

$$\begin{aligned}
&\left|\exp\left(-\int_{-c_1}^{\infty} \mathbf{P}\left[\inf_{t\in K} X_t^n(x) \;\geq\; -c_1\right]\eta_n(x)\,\mathrm{d}x\right)\right.\\
&\qquad\left.-\;\exp\left(-\int_{-c_1}^{\infty} \mathbf{P}\left[\inf_{t\in K} L_t + x \;\geq\; -c_1\right]\exp(-x)\,\mathrm{d}x\right)\right| \;\leq\; \frac{\varepsilon}{8}\,.
\end{aligned}$$

Combining (22)-(24) yields (21).

Analysis of $1 - \mathbf{P}[B_{1,n}^c]^n$

The goal is to show that there exists $c_2 = c_2(c_1)$ such that, for $N(c_1, c_2)$ sufficiently large, we have

$$1 \;-\; \left(\mathbf{P}[B_{1,n}^c]\right)^n \;\leq\; \frac{\varepsilon}{4} \qquad \text{for all } n \geq N(c_1,c_2)\,. \tag{25}$$

It holds

$$\begin{aligned}
\mathbf{P}[B_{1,n}^c] &= \mathbf{P}\left[Y_{1,n}(0) > -c_2 \quad \text{or} \quad \sup_{t\in K} Y_{1,n}(t) < -c_1\right] \\
&= 1 - \mathbf{P}\left[\max_{i\in\mathbb{N}} U^{(i)} \le -c_2 + \log(n)\right] \\
&\quad + \mathbf{P}\left[\max_{i\in\mathbb{N}} U^{(i)} \le -c_2 + \log(n) \quad \text{and} \quad \max_{i\in\mathbb{N}} \sup_{t\in K} X_t^{(i)}(U^{(i)}) < -c_1 + \log(n)\right],
\end{aligned} \tag{26}$$

where the first probability in (26) satisfies

$$\begin{aligned}
\mathbf{P}\left[\max_{i\in\mathbb{N}} U^{(i)} \le -c_2 + \log(n)\right] &= \mathbf{P}\left[\sum_{i\in\mathbb{N}} \mathbb{1}\left\{U^{(i)} > -c_2 + \log(n)\right\} = 0\right] \\
&= \exp\left(-\int_{-c_2+\log(n)}^{\infty} \alpha(x)\exp(-x)\,\mathrm{d}x\right).
\end{aligned} \tag{27}$$

For the second probability in (26), note that

$$\begin{aligned}
&\left\{\max_{i\in\mathbb{N}} U^{(i)} \le -c_2 + \log(n) \quad \text{and} \quad \max_{i\in\mathbb{N}} \sup_{t\in K} X_t^{(i)}(U^{(i)}) < -c_1 + \log(n)\right\} \\
&= \bigcap_{i\in\mathbb{N}} \left\{U^{(i)} \le -c_2 + \log(n)\right\} \cap \bigcap_{i\in\mathbb{N}} \left\{\sup_{t\in K} X_t^{(i)}(U^{(i)}) < -c_1 + \log(n)\right\} \\
&\subset \bigcap_{i\in\mathbb{N}} \left\{U^{(i)} \le -c_2 + \log(n)\right\} \\
&\qquad \cap \left\{\sup_{t\in K} X_t^{(i)}(U^{(i)}) < -c_1 + \log(n) \quad \text{or} \quad U^{(i)} > -c_2 + \log(n)\right\} \\
&= \bigcap_{i\in\mathbb{N}} \Bigg(\left\{U^{(i)} > -c_2 + \log(n)\right\} \\
&\qquad \cup \left\{\sup_{t\in K} X_t^{(i)}(U^{(i)}) \ge -c_1 + \log(n) \quad \text{and} \quad U^{(i)} \le -c_2 + \log(n)\right\}\Bigg)^c \\
&= \Bigg\{\sum_{i\in\mathbb{N}} \mathbb{1}\Bigg\{U^{(i)} > -c_2 + \log(n) \quad \text{or} \\
&\qquad \left(\sup_{t\in K} X^{(i)}(U^{(i)}) \ge -c_1 + \log(n) \quad \text{and} \quad U^{(i)} \le -c_2 + \log(n)\right)\Bigg\} = 0\Bigg\},
\end{aligned} \tag{28}$$

where the probability of the latter set, by Theorem 2.3, is

$$\begin{aligned}
&\exp\Bigg(-\int_{-c_2+\log(n)}^{\infty} \alpha(x)\cdot\exp(-x)\ \mathrm{d}x \\
&\qquad - \int_{-\infty}^{-c_2+\log(n)} \mathbf{P}[\sup_{t\in K} X_t(x) > -c_1 + \log(n)]\cdot\alpha(x)\cdot\exp(-x)\ \mathrm{d}x\Bigg).
\end{aligned} \tag{29}$$

It follows from (26)-(29) that

$$
\begin{aligned}
&1-\mathbf{P}[B_{1,n}^c]^n \\
&= 1- \\
&\quad\Bigg(1-\exp\left(-\int_{-c_2+\log(n)}^{\infty}\alpha(x)\exp(-x)\,\mathrm{d}x\right) \\
&\qquad + \exp\Bigg(-\int_{-c_2+\log(n)}^{\infty}\alpha(x)\cdot\exp(-x)\,\mathrm{d}x \\
&\qquad\qquad - \int_{-\infty}^{-c_2+\log(n)}\mathbf{P}[\sup_{t\in K}X_t(x)>-c_1+\log(n)]\cdot\alpha(x)\cdot\exp(-x)\,\mathrm{d}x\Bigg)\Bigg)^n \\
&= 1- \\
&\quad\Bigg(1-\exp\left(-\int_{-c_2+\log(n)}^{\infty}\alpha(x)\exp(-x)\,\mathrm{d}x\right) \\
&\qquad\times\left(1\;-\;\exp\left(-\int_{-\infty}^{-c_2+\log(n)}\mathbf{P}[\sup_{t\in K}X_t(x)>-c_1+\log(n)]\cdot\alpha(x)\cdot\exp(-x)\,\mathrm{d}x\right)\right)\Bigg)^n \\
&= 1- \\
&\quad\Bigg(1-\exp\left(-\frac{1}{n}\int_{-c_2}^{\infty}\alpha(x+\log(n))\exp(-x)\,\mathrm{d}x\right) \\
&\quad\times\left(1\;-\;\exp\left(-\frac{1}{n}\int_{-\infty}^{-c_2}\mathbf{P}[\sup_{t\in K}X_t(x+\log(n))>-c_1+\log(n)]\cdot\alpha(x+\log(n))\cdot\exp(-x)\,\mathrm{d}x\right)\right)\Bigg)^n \\
&\leq 1- \\
&\quad\Bigg(1-\exp\left(-\frac{1}{n}\int_{-c_2}^{\infty}\alpha(x+\log(n))\exp(-x)\,\mathrm{d}x\right) \\
&\quad\times\frac{1}{n}\int_{-\infty}^{-c_2}\mathbf{P}\left[\sup_{t\in K}X_t(x+\log(n))>-c_1+\log(n)\right]\cdot\alpha(x+\log(n))\cdot\exp(-x)\,\mathrm{d}x\Bigg)^n,
\end{aligned}
\tag{30}
$$

where in the last inequality we used $1-\exp(-x)\leq x$. It remains to show that, for $c_2=c_2(c_1)$ large enough, we have

$$
\int_{-\infty}^{-c_2}\mathbf{P}\left[\sup_{t\in K}X_t(x+\log(n))\geq -c_1+\log(n)\right]\alpha(x+\log(n))\exp(-x)\,\mathrm{d}x\;\leq\;\frac{\varepsilon}{4}. \tag{31}
$$

Then, for $n\geq N(c_1,c_2)$ sufficiently large,

$$
\begin{aligned}
&-\exp\left(-\frac{1}{n}\int_{-c_2}^{\infty}\alpha(x+\log(n))\exp(-x)\,\mathrm{d}x\right) \\
&\quad\times\frac{1}{n}\int_{-\infty}^{-c_2}\mathbf{P}[\sup_{t\in K}X_t(x+\log(n))>-c_1+\log(n)]\cdot\alpha(x+\log(n))\cdot\exp(-x)\,\mathrm{d}x \\
&\geq\;\frac{-\varepsilon}{4n}\;>\;-1,
\end{aligned}
$$

and applying Bernoulli's inequality together with (30) gives

$$
\begin{aligned}
&1 \;-\; \big(\mathbf{P}[B_{1,n}^c]\big)^n \\
&\quad\le\; 1 \\
&\qquad -\; \bigg(1-\exp\bigg(-\frac{1}{n}\int_{-c_2}^{\infty}\alpha(x+\log(n))\exp(-x)\,\mathrm{d}x\bigg) \\
&\qquad \times\; \frac{1}{n}\int_{-\infty}^{-c_2}\mathbf{P}[\sup_{t\in K}X_t(x+\log(n))>-c_1+\log(n)]\cdot\alpha(x+\log(n))\cdot\exp(-x)\;\mathrm{d}x\bigg)^n \\
&\quad\le\; 1 \;-\; \Big(1-\frac{\varepsilon}{4}\Big) \;=\; \frac{\varepsilon}{4}\,.
\end{aligned}
$$

Let us show (31). Observe that

$$(-c_1+\log(n))-(x+\log(n)) \;=\; -c_1-x \;>\; -c_1+c_2 \qquad \text{for } x\in(-\infty,-c_2)\,.$$

Hence, there exists $c_2=c_2(c_1)$ large enough such that, by Lemma 4.2 and Assumption 1, there exist $v>0$, depending only on K, and a constant $C_{v,K,\alpha}>0$ such that, for $x<-c_2$,

$$
\begin{aligned}
&\mathbf{P}\left[\sup_{t\in K}X_t(x+\log(n))\ge -c_1+\log(n)\right] \\
&\quad\le\; C_{v,K,\alpha}\cdot\mathbf{P}[L_{\overline{K}\cdot\overline{\alpha}}\ge(-c_1+\log(n))-(x+\log(n))-v] \\
&\quad=\; C_{v,K,\alpha}\cdot\mathbf{P}[L_{\overline{K}\cdot\overline{\alpha}}\ge -c_1-x-v]\,,
\end{aligned}
$$

and therefore

$$
\begin{aligned}
&\int_{-\infty}^{-c_2}\mathbf{P}\left[\sup_{t\in K}X_t(x+\log(n))\ge -c_1+\log(n)\right]\alpha(x+\log(n))\exp(-x)\,\mathrm{d}x \\
&\quad\le\; C_{v,K,\alpha}\cdot\overline{\alpha}\int_{-\infty}^{-c_2}\mathbf{P}\left[L_{\overline{K}\cdot\overline{\alpha}}\ge -c_1-x-v\right]\exp(-x)\,\mathrm{d}x \\
&\quad=\; C_{v,K,\alpha}\cdot\overline{\alpha}\cdot\exp(c_1+v)\int_{c_2-c_1-v}^{\infty}\mathbf{P}\left[L_{\overline{K}\sup\alpha}\ge x\right]\exp(x)\,\mathrm{d}x \\
&\quad=\; C_{v,K,\alpha}\cdot\overline{\alpha}\cdot\exp(c_1+v)\cdot\mathbf{E}\left[\exp(L_{\overline{K}\overline{\alpha}})\mathbb{1}\left\{L_{\overline{K}\overline{\alpha}}\ge c_2-c_1-v\right\}\right]\,.
\end{aligned}
$$

Since $\mathbf{E}[\exp(L_t)]=1$ for all $t\ge 0$, we can choose $c_2>c_1$ large enough to obtain (31).

Analysis of $n\mathbf{P}[C_{1,n}]$

Following Remark 4.1, we assume without loss of generality that $U^{(1)}=\max\{U^{(i)}\}_{i\in\mathbb{N}}$. We decompose

$$
\begin{aligned}
C_{1,n} \;&=\; \{\omega(Y_{1,n},\delta,K)>a \quad\text{and}\quad Y_{1,n}(0)>-c_2\} \\
&\subset\; \Big\{\omega((X_t^{(1)}(U^1)_{t\ge0}),\delta,K)>a \;\;\text{and}\;\; U^{(1)}>-c_2+\log(n) \;\;\text{and}\;\; (Z_t)_{t\in K} \;=\; (X_t^{(1)}(U^{(1)}))_{t\in K}\Big\} \\
&\qquad\cup\Big\{U^{(1)}>-c_2+\log(n) \quad\text{and}\quad \exists_{j>1}\,\exists_{t\in K}\,X_t^{(j)}(U^{(j)}) \;>\; X_t^{(1)}(U^{(1)})\Big\} \\
&\subset\; D_{1,n}\;\cup\;D_{2,n}\,,
\end{aligned}
$$

where

$$
\begin{aligned}
D_{1,n} \;&:=\; \Big\{\omega((X_t^{(1)}(U^{(1)})_{t\ge0}),\delta,K)>a \quad\text{and}\quad U^{(1)}>-c_2+\log(n)\Big\}\,, \\
D_{2,n} \;&:=\; \Big\{U^{(1)}>-c_2+\log(n) \quad\text{and}\quad \exists_{j>1}\,\exists_{t\in K}\,X_t^{(j)}(U^{(j)}) \;>\; X_t^{(1)}(U^{(1)})\Big\}\,,
\end{aligned}
$$

so that

$$n\mathbf{P}[C_{1,n}] \;\leq\; n\left(\mathbf{P}[D_{1,n}] \;+\; \mathbf{P}[D_{2,n}]\right) .$$

Analysis of $n\mathbf{P}[D_{1,n}]$
We show that there exists $\delta > 0$ sufficiently small such that, for $n \geq N(c_1, c_2, \delta)$ sufficiently large,

$$n\mathbf{P}[D_{1,n}] \;\leq\; \frac{\varepsilon}{8} . \tag{32}$$

Observe that

$$\mathbf{P}[U^{(1)} \leq x] \;=\; \exp\left(-\int_x^\infty \alpha(s)\exp(-s)\,\mathrm{d}s\right) , \qquad x \in \mathbb{R} , \tag{33}$$

so that

$$\begin{aligned} n \cdot \mathbf{P}[D_{1,n}] \;&=\; n \cdot \int_{-c_2+\log(n)}^\infty \mathbf{P}[\omega((X_t(x))_{t\geq 0}, \delta, K) > a]\,\mathrm{d}\mathbf{P}_{U^{(1)}}(x) \\ &=\; n \cdot \int_{-c_2+\log(n)}^\infty \mathbf{P}[\omega((X_t(x))_{t\geq 0}, \delta, K) > a] \cdot \alpha(x) \cdot \exp(-x) \cdot \mathbf{P}[U^{(1)} \leq x]\,\mathrm{d}x \\ &\leq\; \overline{\alpha} \int_{-c_2}^\infty \mathbf{P}[\omega((X_t(x+\log(n)) - \log(n))_{t\geq 0}, \delta, K) > a] \cdot \exp(-x)\,\mathrm{d}x \\ &=\; \overline{\alpha} \int_{-c_2}^\infty \mathbf{P}[\omega((X_t^n)_{t\geq 0}, \delta, K) > a] \cdot \exp(-x)\,\mathrm{d}x , \end{aligned} \tag{34}$$

where in the last inequality we used a change of variables and that ω is invariant under spatial shift. To apply weak convergence of $(X_t^n)_{t\geq 0}$ to $(L_t)_{t\geq 0}$, we write

$$\begin{aligned} &\int_{-c_2}^\infty \mathbf{P}[\omega((X_t(x+\log(n)) - \log(n))_{t\geq 0}, \delta, K) > a] \cdot \exp(-x)\,\mathrm{d}x \\ &\leq\; \int_{-c_2}^\infty \mathbf{P}[\omega((L_t)_{t\geq 0}, \delta, K) > a] \cdot \exp(-x)\,\mathrm{d}x \\ &\quad +\; \left|\int_{-c_2}^\infty \mathbf{P}[\omega((X_t(x+\log(n)) - \log(n))_{t\geq 0}, \delta, K) > a] \cdot \exp(-x)\,\mathrm{d}x \right. \\ &\qquad \left. -\; \int_{-c_2}^\infty \mathbf{P}[\omega((L_t)_{t\geq 0}, \delta, K) > a] \cdot \exp(-x)\,\mathrm{d}x \right| . \end{aligned} \tag{35}$$

For the first term in the upper bound of (35), note that since $(L_t)_{t\geq 0}$ has continuous paths we can choose $\delta(c_2) > 0$ small enough, such that

$$\mathbf{P}[\omega((L_t)_{t\geq 0}, \delta, K) > a] \;\leq\; \frac{\varepsilon}{16}\exp(-c_2) ,$$

so that

$$\int_{-c_2}^\infty \mathbf{P}[\omega((L_t)_{t\geq 0}, \delta, K) > a] \cdot \exp(-x)\,\mathrm{d}x \;\leq\; \frac{\varepsilon}{16} . \tag{36}$$

For the second term in the upper bound of (35) note that since

$$\int_{-c_2}^\infty \exp(-x)\,\mathrm{d}x \;=\; \exp(c_2) \;<\; \infty ,$$

and, by Lemma 5.2(ii), as $n \to \infty$,

$$\mathbf{P}[\omega((X_t(x+\log(n)) - \log(n))_{t\geq 0}, \delta, K) > a] \;\to\; \mathbf{P}[\omega((L_t)_{t\geq 0}, \delta, K) > a]\,,$$

so that we can choose $N(c_1, c_2, \delta)$ large enough, such that for $n \geq N(c_1, c_2, \delta)$ it holds

$$\begin{aligned}
&\bigg| \int_{-c_2}^{\infty} \mathbf{P}[\omega((X_t(x+\log(n)) - \log(n))_{t\geq 0}, \delta, K) > a] \cdot \exp(-x)\,\mathrm{d}x \\
&\qquad - \int_{-c_2}^{\infty} \mathbf{P}[\omega((L_t)_{t\geq 0}, \delta, K) > a] \cdot \exp(-x)\,\mathrm{d}x \bigg| \\
&\leq \frac{\varepsilon}{16}\,.
\end{aligned} \tag{37}$$

Putting together (34)-(37) yields (32).

Analysis of $n\mathbf{P}[D_{2,n}]$

We show that there exists $N \in \mathbb{N}$ sufficiently larger such that

$$n\mathbf{P}[D_{2,n}] \;\leq\; \frac{\varepsilon}{8} \qquad \text{for all } n \geq N\,. \tag{38}$$

Step 1

By (34) it holds

$$\begin{aligned}
n\mathbf{P}[D_{2,n}] \;&=\; n\mathbf{P}\left[U^{(1)} > -c_2 + \log(n) \quad \text{and} \quad \exists_{j>1}\,\exists_{t\in K}\, X_t^{(j)}(U^{(j)}) \;\geq\; X_t^{(1)}(U^{(1)})\right] \\
&=\; n\int_{-c_2+\log(n)}^{\infty} \mathbf{P}\left[\exists_{j>1}\,\exists_{t\in K}\, X_t^{(j)}(U^{(j)}) \;\geq\; X_t^{(1)}(U^{(1)}) \mid U^{(1)} = x\right] \mathrm{d}\mathbf{P}_{U^{(1)}}(x) \\
&=\; n\int_{-c_2+\log(n)}^{\infty} \mathbf{P}\left[\exists_{j>1}\,\exists_{t\in K}\, X_t^{(j)}(U^{(j)}) \;\geq\; X_t^{(1)}(U^{(1)}) \mid U^{(1)} = x\right] \\
&\qquad\qquad \times \alpha(x) \cdot \exp(-x) \cdot \mathbf{P}[U^{(1)} \leq x]\,\mathrm{d}x \\
&=\; \int_{-c_2}^{\infty} \mathbf{P}\left[\exists_{j>1}\,\exists_{t\in K}\, X_t^{(j)}(U^{(j)}) \;\geq\; X_t^{(1)}(U^{(1)}) \mid U^{(1)} = x + \log(n)\right] \\
&\qquad\qquad \times \alpha(x+\log(n)) \cdot \exp(-x) \cdot \mathbf{P}[U^{(1)} \leq x + \log(n)]\,\mathrm{d}x \\
&\leq\; \overline{\alpha}\exp(c_2) \sup_{x>-c_2} \mathbf{P}\left[\exists_{j>1}\,\exists_{t\in K}\, X_t^{(j)}(U^{(j)}) \;\geq\; X_t^{(1)}(U^{(1)}) \mid U^{(1)} = x + \log(n)\right]\,.
\end{aligned} \tag{39}$$

Step 2

We define, for $g \in \mathbb{C}([0,\infty))$, the transformation $\widetilde{g}$ by

$$\widetilde{g} \;:\; [0,\infty) \times \mathbb{R} \;\to\; [0,\infty)\,, \qquad (t,z) \;\mapsto\; \left(\int_0^{(\cdot)} \alpha(g(r) + z)\,\mathrm{d}r\right)^{-1}(t)\,.$$

Then, conditional on $(L_t^{(1)})_{t\geq 0} = g$, for some $g \in \mathbb{C}([0,\infty))$, we get

$$X_t^{(1)}(z) \;=\; g(\widetilde{g}(t,z)) \;+\; z\,.$$

By independence of $(L_t^{(1)})_{t\geq 0}$ from the remaining variables,

$$
\begin{aligned}
&\mathbf{P}\left[\exists_{j>1}\,\exists_{t\in K}\, X_t^{(j)}(U^{(j)}) \;\geq\; X_t^{(1)}(U^{(1)}) \mid U^{(1)} = x+\log(n)\right] \\
&\leq\; \mathbf{P}\left[\exists_{j>1} \sup_{t\in K} X_t^{(j)}(U^{(j)}) \;\geq\; \inf_{t\in K} X_t^{(1)}(x+\log(n)) \mid U^{(1)} = x+\log(n)\right] \\
&=\; \mathbf{P}\left[\sum_{j>1} \mathbb{1}\{\sup_{t\in K} X_t^{(j)}(U^{(j)}) \;\geq\; \inf_{t\in K} X_t^{(1)}(x+\log(n))\} \;>\; 0 \;\middle|\; U^{(1)} = x+\log(n)\right] \\
&=\; \int_{\mathbb{C}([0,\infty))} \mathrm{d}\mathbf{P}_{(L_t^{(1)})_{t\geq 0}}(g) \\
&\qquad \mathbf{P}\left[\sum_{j>1} \mathbb{1}\{\sup_{t\in K} X_t^{(j)}(U^{(j)}) \;\geq\; \inf_{t\in K} g(\widetilde{g}(t, x+\log(n))) \;+\; x+\log(n)\} \;>\; 0 \;\middle|\; U^{(1)} = x+\log(n)\right] \\
&\leq\; \int_{\mathbb{C}([0,\infty))} \mathbb{1}\left\{\inf_{t\in[0,\overline{K}/\underline{\alpha}]} g(t) > -R\right\} \mathrm{d}\mathbf{P}_{(L_t^{(1)})_{t\geq 0}}(g) \\
&\qquad \mathbf{P}\left[\sum_{j>1} \mathbb{1}\{\sup_{t\in K} X_t^{(j)}(U^{(j)}) \;\geq\; \inf_{t\in K} g(\widetilde{g}(t, x+\log(n))) \;+\; x+\log(n)\} \;>\; 0 \;\middle|\; U^{(1)} = x+\log(n)\right] \\
&\qquad +\; \mathbf{P}\left[\inf_{t\in[0,\overline{K}/\underline{\alpha}]} L_t \leq -R\right],
\end{aligned}
\tag{40}
$$

where

$$
\mathbf{P}\left[\inf_{t\in[0,\overline{K}/\underline{\alpha}]} L_t \leq -R\right] \;<\; \frac{\varepsilon}{16}\exp(-c_2)\,, \tag{41}
$$

for $R > 0$ large enough.

Step 3

Write

$$
S \;:=\; \{\widetilde{g}(t,z) \mid t\in K\,, z\in\mathbb{R}\} \;=\; \bigcup_{z\in\mathbb{R}} \{\widetilde{g}(t,z) \mid t\in K\} \;=:\; \bigcup_{z\in\mathbb{R}} S_z\,,
$$

where

$$
S_z \;=\; (\widetilde{g}(\cdot,z))^{-1}(K) \;\subset\; (\widetilde{g}(\cdot,z))^{-1}([0,\overline{K}]) \;\subset\; [0,\overline{K}/\underline{\alpha}]\,,
$$

and therefore $S\subset[0,\overline{K}/\underline{\alpha}]$, independent of the choice of $g\in\mathbb{C}([0,\infty))$. Hence,

$$
\inf_{z\in\mathbb{R}}\inf_{t\in K} g(\widetilde{g}(t,z)) \;=\; \inf_{t\in S} g(t) \;\geq\; \inf_{t\in[0,\overline{K}/\underline{\alpha}]} g(t)\,,
$$

so that

$$\begin{aligned}
&\int_{\mathbb{C}([0,\infty))} \mathbb{1}\left\{\inf_{t\in[0,\overline{K}/\underline{\alpha}]} g(t) > -R\right\} \mathrm{d}\mathbf{P}_{(L_t^{(1)})_{t\geq 0}}(g) \\
&\quad \mathbf{P}\left[\sum_{j>1} \mathbb{1}\{\sup_{t\in K} X_t^{(j)}(U^{(j)}) \;\geq\; \inf_{t\in K} g(\widetilde{g}(t, x+\log(n))) \;+\; x+\log(n)\} \;>\; 0 \;\middle|\; U^{(1)} = x+\log(n)\right] \\
&\leq\; \mathbf{P}\left[\sum_{j>1} \mathbb{1}\{\sup_{t\in K} X_t^{(j)}(U^{(j)}) \;\geq\; -R \;+\; x+\log(n)\} \;>\; 0 \;\middle|\; U^{(1)} = x+\log(n)\right].
\end{aligned} \tag{42}$$

Step 4
By the Slivniak-Mecke Theorem (see for example [DV08, Proposition 13.1.VII]), $\{U^{(j)}\}_{j>1}$ conditioned on $U^{(1)} = x+\log(n)$ is a PPP with intensity measure $\eta[\cdot\cap(-\infty, x+\log(n))]$, so that $\{\sup_{t\in K} X^{(j)}(U^{(j)})\}_{j>1}$ is a PPP on the real line with intensity

$$\int_{-\infty}^{x+\log(n)} \mathbf{P}\left[\sup_{t\in K} X_t(z) \in A\right] \eta(z)\,\mathrm{d}z \qquad \text{for all } A\in\mathcal{B}(\mathbb{R})\,.$$

It therefore holds

$$\begin{aligned}
&\mathbf{P}\left[\sum_{j>1} \mathbb{1}\{\sup_{t\in K} X_t^{(j)}(U^{(j)}) \;\geq\; -R \;+\; x+\log(n)\} \;>\; 0 \;\middle|\; U^{(1)} = x+\log(n)\right] \\
&=\; 1 \\
&\qquad -\; \mathbf{P}\left[\sum_{j>1} \mathbb{1}\{\sup_{t\in K} X_t^{(j)}(U^{(j)}) \;\geq\; -R \;+\; x+\log(n)\} \;=\; 0 \;\middle|\; U^{(1)} = x+\log(n)\right] \\
&=\; 1 \;-\; \exp\left(-\int_{-\infty}^{x+\log(n)} \mathbf{P}\left[\sup_{t\in K} X_t(z) \;\geq\; -R \;+\; x+\log(n)\right] \eta(z)\,\mathrm{d}z\right) \\
&\leq\; \int_{-\infty}^{x+\log(n)} \mathbf{P}\left[\sup_{t\in K} X_t(z) \;\geq\; -R \;+\; x+\log(n)\right] \eta(z)\,\mathrm{d}z\,.
\end{aligned} \tag{43}$$

Step 5
Observe that

$$\begin{aligned}
&\int_{-\infty}^{x+\log(n)} \mathbf{P}\left[\sup_{t\in K} X_t(z) \;\geq\; -R \;+\; x+\log(n)\right] \eta(z)\,\mathrm{d}z \\
&\leq\; \int_{x+\log(n)-R-v}^{x+\log(n)} \eta(z)\,\mathrm{d}z \\
&\qquad +\; \int_{-\infty}^{x+\log(n)-R-v} \mathbf{P}\left[\sup_{t\in K} X_t(z) \;\geq\; -R \;+\; x+\log(n)\right] \eta(z)\,\mathrm{d}z\,,
\end{aligned} \tag{44}$$

where we choose v as in Lemma 4.2, and note that for $x > -c_2$

$$\begin{aligned}\int_{x+\log(n)-R-v}^{x+\log(n)} \eta(z)\,\mathrm{d}z \;&\le\; \overline{\alpha}\frac{1}{n}\exp(-x)\left(\exp(R+v)-1\right)\\ &\le\; \exp(c_2)\overline{\alpha}\frac{1}{n}\left(\exp(R+v)-1\right)\\ &\le\; \exp(-c_2)\frac{\varepsilon}{16}\,,\end{aligned} \tag{45}$$

for $n \ge N(c_1, c_2, \delta, v, R)$ large enough.
Step 6
By Lemma 4.2, there exists a constant $C_{v,\overline{K}} > 0$ such that

$$\begin{aligned}&\int_{-\infty}^{x+\log(n)-R-v} \mathbf{P}\left[\sup_{t\in K} X_t(z) \;\ge\; -R \;+\; x+\log(n)\right]\eta(z)\,\mathrm{d}z\\ &\le\; \overline{\alpha}\int_{-\infty}^{x+\log(n)-R-v} \mathbf{P}\left[L_{\overline{\alpha}\cdot\overline{K}} \;\ge\; -R \;+\; x+\log(n)-z-v\right]\exp(-z)\,\mathrm{d}z\\ &=\; \overline{\alpha}\exp(R+v)\exp(-x)\frac{1}{n}\int_0^{\infty} \mathbf{P}\left[L_{\overline{\alpha}\cdot\overline{K}} \;\ge\; z\right]\exp(-z)\,\mathrm{d}z\\ &\le\; \overline{\alpha}\exp(R+v)\exp(c_2)\frac{1}{n}\mathbf{E}[\exp(L_{\overline{\alpha}\cdot\overline{K}})]\\ &\le\; \exp(-c_2)\frac{\varepsilon}{16}\,,\end{aligned} \tag{46}$$

for $n \ge N(c_1, c_2, \delta, v, R)$ large enough. Combining (39)-(46) yields

$$n\mathbf{P}[D_{2,n}] \;\le\; \varepsilon\exp(c_2)\exp(-c_2)\left(\frac{1}{16} \;+\; \frac{1}{16}\right) \;=\; \frac{\varepsilon}{8}\,,$$

that is, (38). This finishes the proof. □

Acknowledgments

The author is grateful to Marco Oesting for many fruitful discussions. The author also wants to thank Lukas Trottner for introducing him to [ADR69].

References

[ADR69] Jacques Azéma, Marie Duflo, and Daniel Revuz. Mesure invariante des processus de Markov récurrents. *Séminaire de probabilités*, 3:24–33, 1969.

[BG68] Robert McCallum Blumenthal and Ronald Kay Getoor. *Markov Processes and Potential Theory.* Academic Press, 1968.

[Bil99] Patrick Billingsley. *Convergence of Probability Measures.* Wiley Series in Probability and Statistics. Probability and Statistics Section. Wiley, New York, 2nd edition, 1999.

[BR77] Bruce M. Brown and Sidney I. Resnick. Extreme Values of Independent Stochastic Processes. *Journal of Applied Probability*, 14(4):732–739, 1977.

[Bro70] Mark Brown. A Property of Poisson Processes and its Application to Macroscopic Equilibrium of Particle Systems. *The Annals of Mathematical Statistics*, 41(6):1935–1941, 1970.

[BSW13] Björn Böttcher, René Schilling, and Jian Wang. *Lévy Matters III: Lévy-Type Processes: Construction, Approximation and Sample Path Properties*, volume 2099 of *Lecture Notes in Mathematics*. Springer International Publishing, Cham, 2013.

[DV08] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes*. Probability and Its Applications. Springer, New York, NY, 2008.

[EK86] S. Ethier and T. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley Series in Probability and Statistics. 1986.

[EK15] Sebastian Engelke and Zakhar Kabluchko. Max-stable processes and stationary systems of Lévy particles. *Stochastic Processes and their Applications*, 125(11):4272–4299, November 2015.

[Haa84] L. De Haan. A Spectral Representation for Max-stable Processes. *The Annals of Probability*, 12(4):1194–1204, November 1984.

[Kab10] Zakhar Kabluchko. Stationary systems of Gaussian processes. *The Annals of Applied Probability*, 20(6):2295–2317, December 2010.

[KSd09] Zakhar Kabluchko, Martin Schlather, and Laurens de Haan. Stationary max-stable fields associated to negative definite functions. *The Annals of Probability*, 37(5):2042–2065, September 2009.

[LdH01] Tao Lin and Laurens de Haan. On convergence toward an extreme value distribution in C[0,1]. *The Annals of Probability*, 29(1):467–483, February 2001.

[MS13] Ilya Molchanov and Kaspar Stucki. Stationarity of multivariate particle systems. *Stochastic Processes and their Applications*, 123(6):2272–2285, June 2013.

[Sto08] Stilian A. Stoev. On the ergodicity and mixing of max-stable processes. *Stochastic Processes and their Applications*, 118(9):1679–1705, September 2008.

[Wil87] Eric Willekens. On the supremum of an infinitely divisible process. *Stochastic Processes and their Applications*, 26:173–175, January 1987.