I N S T I T U T   D E   S T A T I S T I Q U E

B I O S T A T I S T I Q U E   E T

S C I E N C E S   A C T U A R I E L L E S

(I S B A)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

2014/22

Statistics for Tail Processes of Markov Chains

DREES, H., SEGERS, J. and M. WARCHOL

Statistics for Tail Processes of Markov Chains

Holger Drees
University of Hamburg
Department of Mathematics
Bundesstrasse 55, 20146 Hamburg, Germany
holger.drees@math.uni-hamburg.de

Johan Segers Michał Warchoł
Université catholique de Louvain
Institut de Statistique, Biostatistique et Sciences Actuarielles
Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium
johan.segers@uclouvain.be, michal.warchol@uclouvain.be

June 2, 2014

Abstract

At high levels, the asymptotic distribution of a stationary, regularly varying Markov chain is conveniently given by its tail process. The latter takes the form of a geometric random walk, the increment distribution depending on the sign of the process at the current state and on the flow of time, either forward or backward. Estimation of the tail process provides a nonparametric approach to analyze extreme values. A duality between the distributions of the forward and backward increments provides additional information that can be exploited in the construction of more efficient estimators. The large-sample distribution of such estimators is derived via empirical process theory for cluster functionals. Their finite-sample performance is evaluated via Monte Carlo simulations involving copula Markov models and solutions to stochastic recurrence equations. The estimators are applied to stock market data to study the absence or presence of symmetries in the succession of large losses and gains.

Keywords: Heavy–tailed Markov chains; Regular variation; Stationary time series; Tail process; Time reversibility.

1 Introduction

If serial dependence at high levels is sufficiently strong, extreme values of a stationary time series may arrive in clusters rather than in isolation. This is the case, for instance, for linear time series with heavy-tailed innovations and for solutions of stochastic recurrence equations. If a particular time series model is to be used for prediction at such high levels, it is important to model these clusters well. Think of tail-related risk measures in finance or of return levels in hydrology: a rapid succession of particularly rainy days may be especially dangerous if the capacity of the system to absorp the water is limited.

To judge the quality of fit of a time series model at extreme levels, it is useful to have a benchmark relying on as little model assumptions as possible. A purely nonparametric approach, however, has the drawback that there may be too few data that are sufficiently large. For the purpose of extrapolation, the empirical measure is inadequate.

A solution is to rely on asymptotic theory describing possible limit distributions for the extremes of a time series. If this family of distributions is not too large, one may hope to be able to fit it to actual data.

For extremes of stationary time series, there are several asymptotic frameworks available, all of them more or less equivalent. For the study of short-range extremal dependence, the tail process (Basrak and Segers, 2009) is a convenient choice. It captures the collection of finite-dimensional limit distributions of the series conditionally on the event that, at a particular time instant, the series is far from the origin. See Section 2 for a review.

The family of tail processes of regularly varying time series is still too large to permit accurate nonparametric estimation. Additional assumptions serve to render the inference problem more manageable. The choice made in this paper is to focus on stationary univariate Markov chains. The joint distribution of such a chain is determined by its bivariate margins, yielding considerable simplifications. Its tail process takes the form of a geometric random walk, the increments depending both on the sign of the process at the current state and on the direction of time, forward or backward. The random walk representation goes back to Smith (1992) and was developed further in Perfekt (1997), Bortot and Coles (2000), and Yun (2000). The formulation in terms of the tail process stems from Segers (2007) and Janßen and Segers (2014); see Section 3 for details.

The tail process of a stationary time series is itself not stationary because of the special role played by the time instant figuring in the conditioning event. Still, its finite-dimensional distributions satisfy a collection of identities regarding the effect of a time shift. These equations can be summarized into the so-called time-change formula; see equation (2.7) below. Apart from being a probabilistic nicety, the time-change formula is useful from a statistical perspective because it provides additional information on the distribution of the tail process. Exploiting this information can lead to more efficient inference.

Our contribution is to propose and study nonparametric estimators for the tail process of a stationary univariate Markov chain. Large-sample theory and Monte Carlo simulations both confirm that efficiency gains are possible when the time-change formula is incorporated into the estimation procedure. The asymptotic distributions of the estimators are described via functional central limit theorems building on the empirical process theory developed in Drees and Rootzén (2010). The finite-sample performance is investigated for solutions of stochastic recurrence equations and for copula Markov models (Chen and Fan, 2006).

The structure of the paper is as follows. The tail process of a stationary regularly varying time series is reviewed in Section 2. The theory is specialized to Markov chains in Section 3. The estimators of the tail process of a regularly Markov chain are described in Section 4, and their properties are worked out in Sections 5 and 6 from the asymptotic and the finite-sample perspectives, respectively. In Section 7, the estimators are applied to analyze time series of daily log returns of Google and UBS stock prices, revealing interesting patterns regarding the succession of large losses and gains. Proofs and calculations are deferred to Section 8.

Some notational conventions: the law of a random object is denoted by $\mathcal{L}(\,\cdot\,)$. Weak convergence is denoted by the arrow $\rightsquigarrow$. The indicator variable of the event E is denoted by $\mathbf{1}(E)$. The set of integers is denoted by $\mathbb{Z}$, while $\mathbb{N} = \{h \in \mathbb{Z} : h \geq 1\}$.

2 Tail processes and spectral tail processes

Throughout, let $(X_t)_{t\in\mathbb{Z}}$ be a strictly stationary univariate time series. Aggregating positive and negative extremes, a straightforward summary of the extremal dependence within $(X_t)_{t\in\mathbb{Z}}$ is given by the sequence of *tail dependence coefficients*

$$\lambda_h = \lim_{u\to\infty} \mathrm{P}\left[|X_h| > u \mid |X_0| > u\right], \qquad h \in \mathbb{Z}, \tag{2.1}$$

provided the limits exist. By stationarity, $\lambda_h = \lambda_{-h}$.

In order to get a more detailed picture of extremal dependence, we might want to distinguish between positive and negative extremes, for instance to allow for separate modelling of large positive and negative returns on a financial asset. Moreover, one may be interested in the exact value of the exceedance over u. These considerations lead us to consider the full conditional distribution of arbitrary long stretches of the time series given that at a particular time instant a large value occurs.

Formally, a strictly stationary time series $(X_t)_{t\in\mathbb{Z}}$ is said to have a *tail process* $(Y_t)_{t\in\mathbb{Z}}$ if for all $s, t \in \mathbb{Z}$ such that $s \le t$

$$\mathcal{L}\left(u^{-1}X_s, \dots, u^{-1}X_t \mid |X_0| > u\right) \rightsquigarrow \mathcal{L}\left(Y_s, \dots, Y_t\right), \qquad u \to \infty, \tag{2.2}$$

with the implicit understanding that the law of $|Y_0|$ is non-degenerate. Existence of the tail dependence coefficients (2.1) is implied by existence of the tail process (2.2), and

$$\lambda_h = \mathrm{P}\left[|Y_h| > 1\right], \qquad h \in \mathbb{Z}.$$

Here, we use the fact that the distribution of Y_h does not have any atom at $\mathbb{R} \setminus \{0\}$, a property which follows from the spectral decomposition in (2.5).

Specializing equation (2.2) to $t = 0$ implies that $\mathrm{P}\left[|X_0| > uy\right] / \mathrm{P}\left[|X_0| > u\right] \to \mathrm{P}\left[|Y_0| > y\right]$ as $u \to \infty$ for all continuity points y of the law of $|Y_0|$. Since the law of $|Y_0|$ was supposed to be non-degenerate, it follows that the function $u \mapsto \mathrm{P}\left[|X_0| > u\right]$ is regularly varying at infinity, i.e., there exists $\alpha > 0$ such that

$$\lim_{u\to\infty} \frac{\mathrm{P}\left[|X_0| > uy\right]}{\mathrm{P}\left[|X_0| > u\right]} = y^{-\alpha}, \qquad y \in (0, \infty). \tag{2.3}$$

In particular, the law of $|Y_0|$ is Pareto(α), i.e., $\mathrm{P}\left[|Y_0| > y\right] = y^{-\alpha}$ for all $y \ge 1$. More generally, according to Theorem 2.1 in Basrak and Segers (2009), the time series $(X_t)_{t\in\mathbb{Z}}$ admits a tail process $(Y_t)_{t\in\mathbb{Z}}$ with non-degenerate $|Y_0|$ if and only if $(X_t)_{t\in\mathbb{Z}}$ is jointly regularly varying with index $\alpha > 0$, i.e., if for all integers $k \le l$ the random vector $(X_k, \dots, X_l)$ is multivariate regularly varying with index α.

Many time series models are jointly regularly varying and hence admit a tail process. Examples include linear processes with heavy-tailed innovations, solutions to stochastic recurrence equations, and models of the ARCH and GARCH families. Sufficient conditions for such models to be regularly varying can be found in Davis et al. (2013).

The *spectral tail process* is a normalized tail process defined by

$$\Theta_t = Y_t / |Y_0|, \qquad t \in \mathbb{Z}.$$

By (2.2) and the continuous mapping theorem, it follows that for all $s, t \in \mathbb{Z}$ such that $s \le t$

$$\mathcal{L}\left(X_s/|X_0|, \dots, X_t/|X_0| \mid |X_0| > u\right) \to \mathcal{L}\left(\Theta_s, \dots, \Theta_t\right), \qquad u \to \infty. \tag{2.4}$$

The difference between (2.2) and (2.4) is that in the latter equation, the variables X_t are normalized by $|X_0|$ rather than by the threshold u. Such auto-normalization allows the tail process to be decomposed into two stochastically independent components, i.e.,

$$Y_t = |Y_0|\, \Theta_t, \qquad t \in \mathbb{Z}. \tag{2.5}$$

Independence of $|Y_0|$ and $(\Theta_t)_{t\in\mathbb{Z}}$ is stated in Basrak and Segers (2009, Theorem 3.1). The random variable $|Y_0|$ characterizes the magnitudes of extremes, whereas $(\Theta_t)_{t\in\mathbb{Z}}$ captures serial dependence of extremes. The spectral tail process at time $t = 0$ yields information on the relative weights of the upper and lower tails of $|X_0|$: since $\Theta_0 = Y_0/|Y_0| = \mathrm{sign}(Y_0)$, we have

$$p = \mathrm{P}\left[\Theta_0 = +1\right] = \lim_{u\to\infty} \frac{\mathrm{P}\left[X_0 > u\right]}{\mathrm{P}\left[|X_0| > u\right]} \tag{2.6}$$

and likewise $1 - p = \mathrm{P}\left[\Theta_0 = -1\right]$.

Even though the underlying process $(X_t)_{t\in\mathbb{Z}}$ is assumed to be stationary, the (spectral) tail process is itself not stationary. Still, stationarity of $(X_t)_{t\in\mathbb{Z}}$ has consequences for the tail process $(Y_t)_{t\in\mathbb{Z}}$. For instance, for $x > 0$ and $h \in \mathbb{Z} \setminus \{0\}$, we have, by stationarity and regular variation,

$$\begin{aligned}
\mathrm{P}\left[|Y_h| > x\right] &= \lim_{u\to\infty} \mathrm{P}\left[|X_h| > xu \mid |X_0| > u\right] \\
&= \lim_{u\to\infty} \mathrm{P}\left[|X_0| > u \mid |X_h| > xu\right] \frac{\mathrm{P}\left[|X_h| > xu\right]}{\mathrm{P}\left[|X_0| > u\right]} \\
&= x^{-\alpha}\, \mathrm{P}\left[|Y_{-h}| > x^{-1}\right].
\end{aligned}$$

Setting $x = 1$ yields $\lambda_h = \mathrm{P}[|Y_h| > 1] = \mathrm{P}[|Y_{-h}| > 1] = \lambda_{-h}$, as we already knew. For general $x > 0$, the equation in the display provides a way to retrieve the distribution of $|Y_h|$ from the one of $|Y_{-h}|$ and vice versa. More generally, it turns out that the distributions of the *forward* tail process $(Y_t)_{t\geq 0}$ and the *backward* tail process $(Y_t)_{t\leq 0}$ mutually determine each other.

The precise connection between the forward and backward (spectral) tail processes is captured by Theorem 3.1 in Basrak and Segers (2009). For all $i, s, t \in \mathbb{Z}$ with $s \leq 0 \leq t$ and for all measurable functions $f : \mathbb{R}^{t-s+1} \to \mathbb{R}$ satisfying $f(y_s, \ldots, y_t) = 0$ whenever $y_0 = 0$, we have, provided the expectations exist,

$$\mathrm{E}\left[f\left(\Theta_{s-i}, \ldots, \Theta_{t-i}\right)\right] = \mathrm{E}\left[f\left(\frac{\Theta_s}{|\Theta_i|}, \ldots, \frac{\Theta_t}{|\Theta_i|}\right) |\Theta_i|^\alpha \, \mathbf{1}\{\Theta_i \neq 0\}\right]. \tag{2.7}$$

In particular, setting $s = 0$, $t = 1$, and $i = 1$ yields, for every measurable function $f : \mathbb{R}^2 \to \mathbb{R}$ such that $f(0, \cdot\,) = 0$ and for which the expectations exist,

$$\mathrm{E}\left[f(\Theta_{-1}, \Theta_0)\right] = \mathrm{E}\left[f\left(\frac{\Theta_0}{|\Theta_1|}, \frac{\Theta_1}{|\Theta_1|}\right) |\Theta_1|^\alpha \, \mathbf{1}(\Theta_1 \neq 0)\right]. \tag{2.8}$$

Setting $s = -1$, $t = 0$ and $i = -1$ yields the same formula, but with the roles of Θ_{-1} and Θ_1 interchanged. The condition that $f(0, \cdot\,) = 0$ cannot be omitted; to treat a general function $g : \mathbb{R}^2 \to \mathbb{R}$, consider the function $f(\theta_0, \theta_1) = g(\theta_0, \theta_1) - g(0, \theta_1)$.

We will refer to (2.7) as the *time-change formula*. By exploiting the time-change formula, we will be able to improve upon the efficiency of estimators of the tail process.

In equations (2.2) and (2.4), the conditioning event is $\{|X_0| > u\}$, making no distinction between positive extremes, $X_0 > u$, and negative extremes, $X_0 < -u$. Nevertheless, such a distinction is embedded in the spectral tail process itself. If $p = \mathrm{P}[\Theta_0 = +1] > 0$, then (2.4) implies that for all $s, t \in \mathbb{Z}$, $s \leq t$

$$\mathcal{L}\left(X_s/X_0, \ldots, X_t/X_0 \mid X_0 > u\right) \to \mathcal{L}\left(\Theta_s, \ldots, \Theta_t \mid \Theta_0 = +1\right), \qquad u \to \infty. \tag{2.9}$$

Similarly, if $1 - p = \mathrm{P}[\Theta_0 = -1] > 0$, then

$$\mathcal{L}\left(X_s/X_0, \ldots, X_t/X_0 \mid X_0 < -u\right) \to \mathcal{L}\left(-\Theta_s, \ldots, -\Theta_t \mid \Theta_0 = -1\right), \qquad u \to \infty. \tag{2.10}$$

Observe that in (2.9) and (2.10), the process X_t has been normalized by X_0 rather than by $|X_0|$. For Markov chains, specializing the above formulas to the time instants $t = 1$ or $t = -1$ will suffice to reconstruct the whole tail process.

A common procedure in multivariate extreme value theory is to standardize the margins. For jointly regularly varying time series, such a standardization is possible too, although some care is needed because of the possible presence of both positive and negative extremes.

Lemma 2.1. *Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary time series, jointly regularly varying with index $\alpha > 0$, and having spectral tail process $(\Theta_t)_{t\in\mathbb{Z}}$. Put $\bar{F}_{|X_0|}(u) = \mathrm{P}[|X_0| > u]$ for $u \geq 0$. Define a stationary time series $(X_t^*)_{t\in\mathbb{Z}}$ by*

$$X_t^* = \frac{\operatorname{sign}(X_t)}{\bar{F}_{|X_0|}(|X_t|)}, \qquad t \in \mathbb{Z}. \tag{2.11}$$

Then $(X_t^*)_{t\in\mathbb{Z}}$ is jointly regularly varying with index 1. Its spectral tail process $(\Theta_t^*)_{t\in\mathbb{Z}}$ is given by

$$\Theta_t^* = \operatorname{sign}(\Theta_t) \, |\Theta_t|^\alpha, \qquad t \in \mathbb{Z}. \tag{2.12}$$

The standardized series $(X_t^*)_{t\in\mathbb{Z}}$ may have a tail process even if the original series $(X_t)_{t\in\mathbb{Z}}$ has none. In that sense, the standardization procedure in (2.11) widens the field of possible applications of tail processes. Furthermore, as the map $y \mapsto \operatorname{sign}(y) \, |y|^\alpha$ is monotone and symmetric, many structural properties of the original spectral tail process $(\Theta_t)_{t\in\mathbb{Z}}$ are inherited by the standardized process $(\Theta_t^*)_{t\in\mathbb{Z}}$ in (2.12) and vice versa.

Remark 2.2 (Related objects). The tail process does not only determine the tail dependence coefficients but it is also related to other insightful extreme value characteristics of time series.

- Under weak conditions, the *extremal index* (Leadbetter, 1983), viewed as the reciprocal of the expected extremal cluster size, can be expressed through the tail process (Basrak and Segers, 2009, Remark 4.7) as

$$\theta := \lim_{r\to\infty} \lim_{u\to\infty} \mathrm{P}\left[\max_{t=1,\dots,r} |X_t| \le u \,\middle|\, |X_0| > u \right] = \mathrm{P}\left[\sup_{t\ge1} |Y_t| \le 1 \right] = \mathrm{E}\left[\sup_{t\ge0} |\Theta_t|^\alpha - \sup_{t\ge1} |\Theta_t|^\alpha \right].$$

- For suitable sets $A, B \subset \{x : |x| \ge 1\}$ and $h \ge 1$, the *extremogram* (Davis and Mikosch, 2009) can be written as

$$\gamma_{AB}(h) := \lim_{n\to\infty} n \operatorname{cov}\left(\mathbf{1}\left(u_n^{-1} X_0 \in A\right), \mathbf{1}\left(u_n^{-1} X_h \in B\right) \right) = \mathrm{P}\left[Y_0 \in A,\, Y_h \in B \right],$$

  where u_n is such that $n\,\mathrm{P}\left[|X_0| > u_n\right] \to 1$ as $n \to \infty$.

- Let $\|\cdot\|$ be a norm on $\mathbb{R}^2$ and for $h \in \mathbb{Z}$ let S_h be the spectral tail measure of (X_0, X_h) that arises as the weak limit in

$$\mathcal{L}\left((X_0, X_h)/\|(X_0, X_h)\| \mid \|(X_0, X_h)\| > u \right) \rightsquigarrow S_h, \qquad u \to \infty.$$

  Assuming that all variables X_t are nonnegative, the *extremal dependence measure* (Larsson and Resnick, 2012) is given by

$$\mathrm{EDM}[X_0, X_h] := \int_{\aleph_+} a_0 a_h \, \mathrm{d}S_h(\mathrm{d}\boldsymbol{a}) = \frac{\mathrm{E}\left[(1/\Gamma_h)\,(\Theta_h/\Gamma_h)\,\Gamma_h^\alpha \right]}{\mathrm{E}\left[\Gamma_h^\alpha\right]},$$

  where $\aleph_+ = \{\boldsymbol{a} = (a_0, a_h) \in [0,\infty)^2 : \|\boldsymbol{a}\| = 1\}$ and $\Gamma_h = \|(1, \Theta_h)\|$.

Remark 2.3 (Asymptotic independence). Some time series models exhibit asymptotic independence of consecutive observations, that is, $\mathrm{P}\left[|X_k| > u \mid |X_0| > u\right] \to 0$ as $\mathrm{P}\left[|X_0| > u\right] \to 0$ for all $k \in \mathbb{Z} \setminus \{0\}$. Well-known examples are non-degenerate Gaussian time series and classical stochastic volatility models. In that case, the spectral tail process is trivial in the sense that $\Theta_k = 0$ for all $k \ne 0$. Hence the only information it conveys about the dependence structure is that $\mathrm{P}\left[|X_0| > ux,\, |X_k| > uy\right]$ is of smaller order than $\mathrm{P}\left[|X_0| > u\right]$ as $u \to \infty$, but it specifies neither the rate of convergence nor a possible non-degenerate limit after suitable standardization. For that reason, Ledford and Tawn (1996) and Ledford and Tawn (2003), focusing on positive extremes, assumed that there exists a normalizing function $b(u) > 0$, different from $\mathrm{P}\left[X_0 > u\right]$, such that $\mathrm{P}\left[X_0/u > x, X_k/u > y\right]/b(u)$ converges to a non-degenerate limit. Recently, Janßen and Drees (2013) determined this limit for certain classes of stochastic volatility models.

In a different approach, closer in spirit to ours, Kulik and Soulier (2013) examined the asymptotic behavior of $\mathcal{L}(X_0/u, X_1/a_1(u), \dots, X_t/a_t(u) \mid X_0 > u)$ for $t \in \mathbb{N}$. The normalizing functions $u \mapsto a_k(u) > 0$, $k \in \mathbb{N}$, increase more slowly than u if consecutive observations of the time series are asymptotically independent. For certain Markovian time series, the authors proved the existence of a non-trivial limit, conveying more information than the usual tail process. However, they did not prove the existence of an analog to the backward tail process. As a consequence, there is no time-change formula either. As we are interested in constructing more efficient estimators for the tail dependence structure employing the time-change formula, we focus on time series with non-degenerate tail processes.

3 Regularly varying Markov chains

For the purpose of statistical inference, the class of spectral tail processes is too large to be really useful: without additional modelling assumptions, it is impossible to estimate all limiting finite-dimensional distributions that appear in (2.2) or (2.4). Therefore, it is reasonable to consider families of spectral tail processes arising under additional constraints on the underlying time series.

One such family was identified in Segers (2007) and Janßen and Segers (2014) in the context of first-order Markov chains. Let $(\Theta_t)_{t\in\mathbb{Z}}$ be the spectral tail process of an α-regularly varying, stationary time series $(X_t)_{t\in\mathbb{Z}}$. Put $p = \mathrm{P}[\Theta_0 = 1]$ as in (2.6). Specializing equations (2.9) and (2.10) to $s = t = 1$ and $s = t = -1$, introduce random variables A_1, B_1, A_{-1}, B_{-1} as follows: if $p > 0$, then, as $u \to \infty$,

$$\mathcal{L}(X_1/X_0 \mid X_0 > u) \rightsquigarrow \mathcal{L}(A_1) = \mathcal{L}(\Theta_1 \mid \Theta_0 = +1), \tag{3.1}$$

$$\mathcal{L}(X_{-1}/X_0 \mid X_0 > u) \rightsquigarrow \mathcal{L}(A_{-1}) = \mathcal{L}(\Theta_{-1} \mid \Theta_0 = +1), \tag{3.2}$$

and if $p < 1$, then

$$\mathcal{L}(X_1/X_0 \mid X_0 < -u) \rightsquigarrow \mathcal{L}(B_1) = \mathcal{L}(-\Theta_1 \mid \Theta_0 = -1), \tag{3.3}$$

$$\mathcal{L}(X_{-1}/X_0 \mid X_0 < -u) \rightsquigarrow \mathcal{L}(B_{-1}) = \mathcal{L}(-\Theta_{-1} \mid \Theta_0 = -1). \tag{3.4}$$

Further, let $\Theta_0, A_1, A_{-1}, A_2, A_{-2}, \ldots, B_1, B_{-1}, B_2, B_{-2}$ be independent random variables such that $\mathcal{L}(A_t) = \mathcal{L}(A_1)$, $\mathcal{L}(A_{-t}) = \mathcal{L}(A_{-1})$, $\mathcal{L}(B_t) = \mathcal{L}(B_1)$, and $\mathcal{L}(B_{-t}) = \mathcal{L}(B_{-1})$ for all $t \in \mathbb{N}$. Then the spectral tail process $(\Theta_t)_{t\in\mathbb{Z}}$ is said to be a *Markov spectral tail chain* if the following holds: the forward spectral tail process is given recursively by

$$\Theta_t = \begin{cases} \Theta_{t-1} A_t & \text{if } \Theta_{t-1} > 0, \\ 0 & \text{if } \Theta_{t-1} = 0, \\ \Theta_{t-1} B_t & \text{if } \Theta_{t-1} < 0, \end{cases} \qquad t \in \mathbb{N}, \tag{3.5}$$

whereas the backward spectral tail process is given by

$$\Theta_{-t} = \begin{cases} \Theta_{-t+1} A_{-t} & \text{if } \Theta_{-t+1} > 0, \\ 0 & \text{if } \Theta_{-t+1} = 0, \\ \Theta_{-t+1} B_{-t} & \text{if } \Theta_{-t+1} < 0, \end{cases} \qquad t \in \mathbb{N}. \tag{3.6}$$

If $p = 1$, then $\Theta_t \geq 0$ almost surely for all $t \in \mathbb{Z}$ and thus the definition of $B_{\pm t}$ is immaterial; similarly if $p = 0$. This can be seen by applying the time-change formula (2.7).

The motivation behind the above definition is that such spectral tail processes typically arise when $(X_t)_{t\in\mathbb{Z}}$ is a stationary, first-order Markov chain; see Theorem 5.2 in Segers (2007) and Corollary 5.1 in Janßen and Segers (2014). The forward and backward spectral tail processes $(\Theta_t)_{t\geq 0}$ and $(\Theta_t)_{t\leq 0}$ are Markovian themselves, and, conditionally on Θ_0, they are independent. Their structure is that of a geometric random walk where the distribution of the increment at time t depends on the sign of the process at time $t-1$. The point zero acts as an absorbing state.

For Markov spectral tail chains, the distribution of the forward part $(\Theta_t)_{t\geq 0}$ is determined by p, A_1, and B_1. Given additionally the index of regular variation $\alpha > 0$, the distributions of A_{-1} and B_{-1} and thus of the backward part $(\Theta_t)_{t\leq 0}$ can be reconstructed from the time-change formula (2.7); see Lemma 4.1 below. It follows that the law of a Markov spectral tail process is determined by $\alpha > 0$, $p \in [0, 1]$, and the laws of A_1 and B_1. This reduction provides a handle on the spectral tail process that can be exploited for statistical inference.

3.1 Copula Markov processes

A particularly convenient feature of above reduction to Markov chains is that to know the quantities α, p, A_1 and B_1, it suffices to know the law of the pair (X_0, X_1). In view of stationarity, this distribution is determined by the univariate marginal distribution function, say G, and a copula C, through the formula

$$\mathrm{P}[X_0 \leq x_0,\, X_1 \leq x_1] = C\big(G(x_0),\, G(x_1)\big), \qquad (x_0, x_1) \in \mathbb{R}^2. \tag{3.7}$$

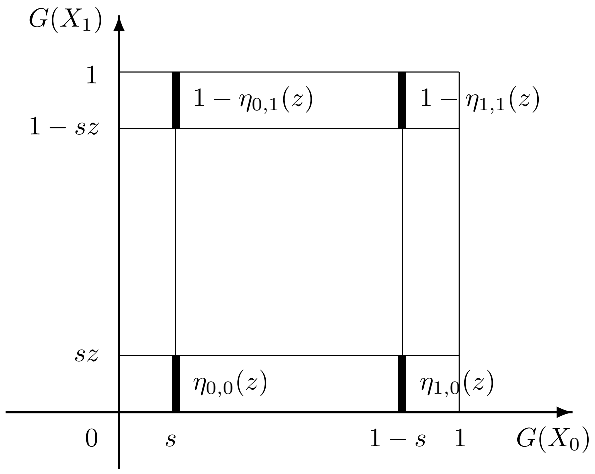

Figure 1: Viewing the functions $\eta_{0,0}$, $\eta_{0,1}$, $\eta_{1,0}$ and $\eta_{1,1}$ in terms of the conditional distribution of $G(X_1)$ conditionally on $G(X_0) = s$ or on $G(X_0) = 1 - s$.

For stationary Markov processes, the joint distribution of $(X_t)_{t\in\mathbb{Z}}$ is determined by the law of (X_0, X_1) and thus by G and C as well. This observation is the motivation for the copula Markov models studied in Chen and Fan (2006) and Chen et al. (2009).

The result below links the distributions of A_1 and B_1 in (3.1) and (3.3) to the margin G and the copula C. Thereby, it provides a direct way of computing the spectral tail chain of copula Markov models. The result is framed in terms of the limit behaviour of $\dot{C}_1(u, v) = \partial C(u, v)/\partial u$ as u tends to 0 or 1. The function $\dot{C}_1$ is related to the conditional distribution of X_1 given X_0; see (8.3) below. For $z \geq 0$, consider the limits (whose existence is an assumption)

$$\lim_{s\searrow 0} \dot{C}_1(1 - s, 1 - sz) = \eta_{1,1}(z), \tag{3.8}$$

$$\lim_{s\searrow 0} \dot{C}_1(1 - s, sz) = \eta_{1,0}(z), \tag{3.9}$$

and

$$\lim_{s\searrow 0} \dot{C}_1(s, 1 - sz) = \eta_{0,1}(z), \tag{3.10}$$

$$\lim_{s\searrow 0} \dot{C}_1(s, sz) = \eta_{0,0}(z), \tag{3.11}$$

covering the four corners of the unit square; see Figure 1.

Proposition 3.1. *Let the distribution of (X_0, X_1) be given by (3.7), where X_0 has a Lebesgue density and satisfies the regular variation condition (2.3) and the tail-balance condition (2.6). Assume that C admits a continuous first-order partial derivative $\dot{C}_1$ on $(0, 1) \times [0, 1]$. If $p > 0$ and if the limits in (3.8) and (3.9) exist and are continuous on $[0, \infty)$, then (3.1) holds and*

$$\mathrm{P}[A_1 \leq x] = \begin{cases} \eta_{1,1}(x^{-\alpha}) & \text{if } x > 0, \\ \eta_{1,0}(\frac{1-p}{p}\,|x|^{-\alpha}) & \text{if } x < 0. \end{cases} \tag{3.12}$$

Similarly, if $p < 1$ and if the limits in (3.10) and (3.11) exist and are continuous on $[0, \infty)$, then (3.3) holds and

$$\mathrm{P}[B_1 \leq x] = \begin{cases} 1 - \eta_{0,0}(x^{-\alpha}) & \text{if } x > 0, \\ 1 - \eta_{0,1}(\frac{p}{1-p}\,|x|^{-\alpha}) & \text{if } x < 0. \end{cases} \tag{3.13}$$

Similarly, the distributions of A_{-1} and B_{-1} in (3.2) and (3.4) can be obtained via the limit behaviour of $\dot{C}_2(u,v) = \partial C(u,v)/\partial v$ as v tends to 0 or 1. Some examples are worked out in Appendix 8.1, whereas the proof of Proposition 3.1 is given in Appendix 8.2.

3.2 Stochastic recurrence equations

The stochastic recurrence equation

$$X_t = C_t X_{t-1} + D_t, \quad t \in \mathbb{Z}, \tag{3.14}$$

received some attention in time series analysis and extreme value theory. We focus on the case where (C_t, D_t), $t \in \mathbb{Z}$, is an i.i.d. $\mathbb{R}^2$–valued sequence. Provided that $-\infty \leq \mathrm{E}[\log|C_1|] < 0$ and $\mathrm{E}\left[\log^+|D_1|\right] < \infty$, where $\log^+ x = \max(\log x, 0)$, there exists a unique strictly stationary causal solution to (3.14) (Basrak et al., 2002, Corollary 2.2).

Results on regular variation of X_0 were first developed by Kesten (1973). His Theorem 5 states that if there exists $\alpha > 0$ such that $\mathrm{E}[|C_1|^\alpha] = 1$, $\mathrm{E}\left[|C_1|^\alpha \log^+|C_1|\right] < \infty$ and $\mathrm{E}[|D_1|^\alpha] < \infty$ and if some other conditions are satisfied, then X_0 is regularly varying; more specifically,

$$\left.\begin{array}{rcl} \Pr(X_0 > x) &=& c_+ x^{-\alpha}(1 + o(1)) \\ \Pr(X_0 \leq -x) &=& c_- x^{-\alpha}(1 + o(1)) \end{array}\right\}, \qquad x \to \infty, \tag{3.15}$$

for constants $c_+,\ c_- \geq 0$ such that $c_+ + c_- > 0$. This result was extended by Goldie (1991, Lemma 2.2 and Theorem 4.1), who gave explicit expressions for c_+ and c_-. Regular variation of X_0 and iteration of (3.14) gives joint regular variation of $(X_t)_{t\in\mathbb{Z}}$ with index α.

The forward spectral tail process $(\Theta_t)_{t\geq 0}$ of $(X_t)_{t\in\mathbb{Z}}$ admits the representation

$$\Theta_t = \Theta_0 \prod_{h=1}^{t} \tilde{C}_h, \qquad t \in \mathbb{N},$$

where $\tilde{C}_t$, $t \in \mathbb{N}$, are i.i.d. random variables with the same distribution as C_1 which are independent of Θ_0 (Janßen and Segers, 2014, Example 6.1). Hence, $(\Theta_t)_{t\geq 0}$ becomes a Markov spectral tail chain satisfying (3.5) with $\mathcal{L}(A_1) = \mathcal{L}(B_1) = \mathcal{L}(C_1)$. Moreover, by Goldie (1991, Theorem 4.1), if $\mathrm{P}[C_1 < 0] > 0$ then necessarily $c_+ = c_-$ and therefore $p = \mathrm{P}[\Theta_0 = +1] = 1/2$. This property also follows from the following result on more general tail spectral processes; its proof is given in Appendix 8.2.

Lemma 3.2. *Let $(\Theta_t)_{t\in\mathbb{Z}}$ be the spectral tail process of an α-regularly varying stationary time series $(X_t)_{t\in\mathbb{Z}}$. Suppose the following two conditions hold:*

(i) $\mathrm{E}[|\Theta_1|^\alpha] = 1$*;*

(ii) Θ_1/Θ_0 *and* Θ_0 *are independent.*

Then Θ_{-1}/Θ_0 and Θ_0 are independent too. Moreover, $\mathrm{P}[\Theta_1/\Theta_0 < 0] > 0$ *implies* $p = \mathrm{P}[\Theta_0 = 1] = 1/2$.

Lemma 3.2 applies to the solution $(X_t)_{t\in\mathbb{Z}}$ of the stochastic recurrence equation (3.14). Above, we had already seen that $\mathcal{L}(A_1) = \mathcal{L}(B_1) = \mathcal{L}(C_1)$ and $\mathrm{E}[|C_1|^\alpha] = 1$. Hence, the spectral tail process $(\Theta_t)_{t\in\mathbb{Z}}$ satisfies the two conditions in Lemma 3.2. We obtain that Θ_{-1}/Θ_0 is independent of Θ_0 too. But since the backward spectral tail process admits the representation in (3.6), we conclude that $\mathcal{L}(A_{-1}) = \mathcal{L}(B_{-1})$. If $p = 0$ or $p = 1$, the previous statements simplify, since A_1 and A_{-1} in (3.1)–(3.2) are defined only if $p > 0$ whereas B_1 and B_{-1} in (3.3)–(3.4) are defined only if $p < 1$. For $p \in \{0, 1\}$, the law of Θ_0 is degenerate at -1 or at 1, respectively, so that the statements on independence are trivially satisfied.

4 Estimating Markov spectral tail processes

In this section we propose estimators for p, A_1 and B_1. In combination with the index of regular variation $\alpha > 0$, this triplet fully determines the law of a Markov spectral tail process as defined in equations (3.5) and (3.6), and of the tail processes $(Y_t)_{t\in\mathbb{Z}}$.

Replacing population distributions by sampling distributions in the left-hand sides of (3.1) and (3.3) yields forward estimators for the laws of A_1 and B_1. However, exploiting the time-change formula (2.7) allows to express the laws of A_1 and B_1 in terms of A_{-1} and B_{-1} (and p and α). These expressions motivate so-called backward estimators for A_1 and B_1. Convex combinations of forward and backward estimators finally produce mixture estimators. For an appropriate choice of the mixture weights, the mixture estimators may be more efficient than both the forward and the backward estimators separately.

In order to estimate $p = \mathrm{P}[\Theta_0 = 1]$, we simply take the empirical version of (2.6), yielding

$$\hat{p}_n = \frac{\sum_{i=1}^n \mathbf{1}(X_i > u_n)}{\sum_{i=1}^n \mathbf{1}(|X_i| > u_n)}. \tag{4.1}$$

For $\hat{p}_n$ to be consistent and asymptotically normal, the threshold sequence u_n should tend to infinity at a certain rate described in detail in condition **(B)** in the next section.

For estimating the cdf $F^{(A_1)}(x)$ of A_1 we propose

$$\hat{F}_n^{(\mathrm{f},A_1)}(x) = \frac{\sum_{i=1}^n \mathbf{1}(X_{i+1}/X_i \le x,\ X_i > u_n)}{\sum_{i=1}^n \mathbf{1}(X_i > u_n)}, \tag{4.2}$$

which we refer to as the *forward estimator* of the cdf of A_1. Similarly, for the forward estimator of the cdf of B_1 we take

$$\hat{F}_n^{(\mathrm{f},B_1)}(x) = \frac{\sum_{i=1}^n \mathbf{1}(X_{i+1}/X_i \le x,\ X_i < -u_n)}{\sum_{i=1}^n \mathbf{1}(X_i < -u_n)}. \tag{4.3}$$

The forward estimators of the cdf's of A_1 and B_1 are empirical versions of the left-hand sides of (3.1) and (3.3), respectively. Note that one can expect consistency of these estimators only if the target distribution functions are continuous in x, because otherwise $\mathrm{P}[X_1/X_0 \le x \mid X_0 > u_n]$ need not converge to $\mathrm{P}[A_1 \le x]$, for instance.

The time-change formula (2.7) yields a different representation of A_1 and B_1, motivating different estimators than the ones above, based on different data points. For ease of reference, we record the relevant formulas in a lemma, whose proof is given in Appendix 8.2.

Lemma 4.1. *Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary time series, jointly regularly varying with index α and spectral tail process $(\Theta_t)_{t\in\mathbb{Z}}$. Let A_1, A_{-1}, B_1, B_{-1} be given as in* (3.1) *to* (3.4)*. If $p = \mathrm{P}[\Theta_0 = 1] > 0$, then*

$$\mathrm{P}[A_1 > x] = \mathrm{E}\left[A_{-1}^{\alpha}\,\mathbf{1}(1/A_{-1} > x)\right], \qquad x \ge 0, \tag{4.4}$$

$$\mathrm{P}[A_1 \le x] = \frac{1-p}{p}\,\mathrm{E}\left[(-B_{-1})^{\alpha}\,\mathbf{1}(1/B_{-1} \le x)\right], \qquad x < 0. \tag{4.5}$$

Similarly, if $p < 1$, then

$$\mathrm{P}[B_1 > x] = \mathrm{E}\left[B_{-1}^{\alpha}\,\mathbf{1}(1/B_{-1} > x)\right], \qquad x \ge 0, \tag{4.6}$$

$$\mathrm{P}[B_1 \le x] = \frac{p}{1-p}\,\mathrm{E}\left[(-A_{-1})^{\alpha}\,\mathbf{1}(1/A_{-1} \le x)\right], \qquad x < 0. \tag{4.7}$$

Formulas (4.4) *to* (4.7) *remain valid when the time instances* 1 *and* -1 *are interchanged.*

Assume for the moment that α is known. Below, we will consider the more realistic situation that α is unknown. Lemma 4.1 suggests the following *backward estimator* of the cdf of A_1:

$$\hat{F}_n^{(\mathrm{b},A_1)}(x) = \begin{cases} 1 - \dfrac{\sum_{i=1}^n \left(\frac{X_{i-1}}{X_i}\right)^\alpha \mathbf{1}\left(X_i/X_{i-1} > x,\ X_i > u_n\right)}{\sum_{i=1}^n \mathbf{1}\left(X_i > u_n\right)} & \text{if } x \geq 0, \\ \dfrac{\sum_{i=1}^n \left(\frac{-X_{i-1}}{X_i}\right)^\alpha \mathbf{1}\left(X_i/X_{i-1} \leq x,\ X_i < -u_n\right)}{\sum_{i=1}^n \mathbf{1}\left(X_i > u_n\right)} & \text{if } x < 0. \end{cases} \tag{4.8}$$

Similarly, we define the backward estimator of the cdf of B_1 as

$$\hat{F}_n^{(\mathrm{b},B_1)}(x) = \begin{cases} 1 - \dfrac{\sum_{i=1}^n \left(\frac{X_{i-1}}{X_i}\right)^\alpha \mathbf{1}\left(X_i/X_{i-1} > x,\ X_i < -u_n\right)}{\sum_{i=1}^n \mathbf{1}\left(X_i < -u_n\right)} & \text{if } x \geq 0, \\ \dfrac{\sum_{i=1}^n \left(\frac{-X_{i-1}}{X_i}\right)^\alpha \mathbf{1}\left(X_i/X_{i-1} \leq x,\ X_i > u_n\right)}{\sum_{i=1}^n \mathbf{1}\left(X_i < -u_n\right)} & \text{if } x < 0. \end{cases} \tag{4.9}$$

For $|x|$ large, the backward estimators usually have a smaller variance than the forward estimators. To see this, note that for negative x with large modulus only very few summands in the numerator of (4.2) do not vanish, because X_{i+1} must be even larger in absolute value than $|x|X_i > |x|u_n$, leading to a large variance of the numerator. In contrast, usually many more non-vanishing terms will be summed up in the numerator of (4.8), while each of them gets a rather low weight $(-X_{i-1}/X_i)^\alpha \leq |x|^{-\alpha}$, leading to a smaller variance. For large positive x one may argue similarly by considering the corresponding estimators of the survival function. Indeed, we show in Remark 5.3 hat, provided $x \geq 1$, the backward estimator of the cdf of A_1 at x has a smaller asymptotic variance than the forward estimator.

For well-chosen weights, convex combinations of the forward and backward estimators can achieve a lower asymptotic variance than each of the estimators individually. Unfortunately, the expression for the asymptotic covariance of the two estimators is intractable; see Corollary 5.2. It remains an open issue how to choose the mixture weights in order to minimize the asymptotic variance.

A pragmatic approach is to give more weight to the forward estimator for small $|x|$ and to give more weight to the backward estimator for large $|x|$. To this end, define weights by

$$\lambda^+(x) = \begin{cases} \left(1 - \hat{F}_n^{(\mathrm{f},A_1)}(x)\right) / \left(1 - \hat{F}_n^{(\mathrm{f},A_1)}(0)\right) & \text{if } \hat{F}_n^{(\mathrm{f},A_1)}(0) \neq 1, \\ 1 & \text{if } \hat{F}_n^{(\mathrm{f},A_1)}(0) = 1; \end{cases}$$

$$\lambda^-(x) = \begin{cases} \hat{F}_n^{(\mathrm{f},A_1)}(x) / \hat{F}_n^{(\mathrm{f},A_1)}(0) & \text{if } \hat{F}_n^{(\mathrm{f},A_1)}(0) \neq 0, \\ 1 & \text{if } \hat{F}_n^{(\mathrm{f},A_1)}(0) = 0. \end{cases}$$

The *mixture estimator* for the cdf of A_1 is defined as

$$\hat{F}_n^{(\mathrm{m},A_1)}(x) = \begin{cases} \lambda^+(x)\hat{F}_n^{(\mathrm{f},A_1)}(x) + \left[1 - \lambda^+(x)\right]\hat{F}_n^{(\mathrm{b},A_1)}(x) & \text{if } x \geq 0, \\ \lambda^-(x)\hat{F}_n^{(\mathrm{f},A_1)}(x) + \left[1 - \lambda^-(x)\right]\hat{F}_n^{(\mathrm{b},A_1)}(x) & \text{if } x < 0. \end{cases} \tag{4.10}$$

The mixture estimator for B_1 is defined by replacing A_1 in (4.10) with B_1.

The backward and the mixture estimators require the value of the index α of regular variation, which is unknown in most applications. There are at least two approaches to deal with this issue:

1. Estimate α separately, for instance, by the Hill–type estimator

$$\hat{\alpha} = \frac{\sum_{i=1}^n \mathbf{1}\left(|X_i| > u_n\right)}{\sum_{i=1}^n \log\left(|X_i|/u_n\right)\mathbf{1}\left(|X_i| > u_n\right)}, \tag{4.11}$$

   and plug in the estimated value of α in (4.8), (4.9) and (4.10).

2. Employ an empirical version of the transformation in Lemma 2.1 to ensure that, after transformation, $\alpha = 1$. The transformation in (2.11) requires the tail function $\bar{F}_{|X_0|}(u) = \mathrm{P}\,[|X_0| > u]$. This function can be estimated, for instance, by

$$\hat{\bar{F}}_{|X_0|,n}(u) = 1 - \frac{1}{n+1} \sum_{j=1}^{n} \mathbf{1}(|X_j| \le u), \tag{4.12}$$

where we divide by $n+1$ rather than by n in order to avoid division by zero later on. The transformed variable

$$\hat{X}^*_{n,i} = \operatorname{sign}(X_i)/\hat{\bar{F}}_{|X_0|,n}(|X_i|)$$

is based on the sign of X_i and the rank of $|X_i|$ among $|X_1|, \ldots, |X_n|$.

In the simulation study in Section 6, the mixture estimator based on the rank-transformed data performs better than the plug-in version. Note, however, that the two approaches are not directly comparable: with the second approach, what we estimate is the tail process $(\Theta^*_t)_{t\in\mathbb{Z}}$ of the transformed series $(X^*_t)_{t\in\mathbb{Z}}$. From (2.12) and (3.1), it follows that, if $(\Theta_t)_{t\in\mathbb{Z}}$ is a Markov spectral tail chain as in (3.5) and (3.6), then so is $(\Theta^*_t)_{t\in\mathbb{Z}}$, with A_t and B_t to be replaced by $A^*_t = \operatorname{sign}(A_t)\,|A_t|^\alpha$ and $B^*_t = \operatorname{sign}(B_t)\,|B_t|^\alpha$, respectively. Combining the above two estimation approaches, one could even recover A_t and B_t via $A_t = \operatorname{sign}(A^*_t)\,|A^*_t|^{1/\alpha}$ and $B_t = \operatorname{sign}(B^*_t)\,|B^*_t|^{1/\alpha}$. Finally, as $(X^*_t)_{t\in\mathbb{Z}}$ may have a tail process if the original time series has none, the second approach is more widely applicable.

5 Large sample theory

In this section we show that, under certain conditions, the standardized estimation error of the forward and the backward estimators jointly converge to a centered Gaussian process. In order not to overload the presentation, we focus on non-negative Markov chains. In that case, the distribution of $\Theta_1 = A_1$ determines the distribution of the forward spectral tail process, and thus, via the time-change formula, together with α, also the one of the backward spectral tail process. We distinguish between the cases where α is known (Section 5.1) and unknown (Section 5.2). In addition, we briefly indicate how the conditions and results must be modified in the real-valued case (Remark 5.7).

5.1 Known index of regular variation

If the index of regular variation, α, is known, all estimators under consideration can be expressed in terms of *generalized tail array sums*, that is, statistics of the form $\sum_{i=1}^n \phi(X_{n,i})$, where

$$X_{n,i} := \frac{(X_{i-1}, X_i, X_{i+1})}{u_n} \mathbf{1}(X_i > u_n).$$

Drees and Rootzén (2010) give conditions under which, after standardization, such statistics converge to a centered Gaussian process, uniformly over appropriate families of functions ϕ. From these results we will deduce a functional central limit theorem for the processes of forward and backward estimators defined in (4.2) and (4.8), respectively.

For $x \ge 0$, define functions $\phi_1, \phi_{2,x}, \phi_{3,x} : [0,\infty)^3 \to [0,\infty)$ by

$$\begin{aligned}
\phi_1(y_{-1}, y_0, y_1) &:= \mathbf{1}(y_0 > 1) \\
\phi_{2,x}(y_{-1}, y_0, y_1) &:= \mathbf{1}(y_1/y_0 > x, y_0 > 1) \\
\phi_{3,x}(y_{-1}, y_0, y_1) &:= (y_{-1}/y_0)^\alpha \mathbf{1}(y_0/y_{-1} > x, y_{-1} > 0, y_0 > 1).
\end{aligned}$$

The forward and backward estimators of the cdf of A_1 can be written as

$$\hat{F}_n^{(\mathrm{f},A_1)}(x) = 1 - \frac{\sum_{i=1}^n \phi_{2,x}(X_{n,i})}{\sum_{i=1}^n \phi_1(X_{n,i})}, \qquad \hat{F}_n^{(\mathrm{b},A_1)}(x) = 1 - \frac{\sum_{i=1}^n \phi_{3,x}(X_{n,i})}{\sum_{i=1}^n \phi_1(X_{n,i})}.$$

Taking up the notation of Drees and Rootzén (2010), we consider the empirical process $\tilde{Z}_n$ defined by

$$\tilde{Z}_n(\psi) := (nv_n)^{-1/2} \sum_{i=1}^{n} \big(\psi(X_{n,i}) - \mathrm{E}\left[\psi(X_{n,i})\right]\big),$$

where ψ is one of ϕ_1, $\phi_{2,x}$ or $\phi_{3,x}$. To establish a functional central limit theorem for $\tilde{Z}_n$, we will need to impose a number of conditions. Let $v_n := \mathrm{P}\left[X_0 > u_n\right]$ and let

$$\beta_{n,k} := \sup_{1 \le l \le n-k-1} \mathrm{E}\left[\sup_{B \in \mathcal{B}^n_{n,l+k+1}} \left|\mathrm{P}\left[B \mid \mathcal{B}^l_{n,1}\right] - \mathrm{P}\left[B\right]\right|\right]$$

denote the β-mixing coefficients. Here $\mathcal{B}^j_{n,i}$ is the σ-field generated by $(X_{n,l})_{i \le l \le j}$. We assume that there exist sequences $l_n, r_n \to \infty$ and some $x_0 \ge 0$ such that the following hold:

(A(x_0)) The cdf $F^{(A_1)}$ of A_1 is continuous on $[x_0, \infty)$.

(B) (i) As $n \to \infty$, we have $l_n \to \infty$, $l_n = o(r_n)$, $r_n = o((nv_n)^{1/2})$, $r_n v_n \to 0$;

(ii) $\beta_{n,l_n} n / r_n \to 0$ as $n \to \infty$ and $\lim_{m\to\infty} \limsup_{n\to\infty} \beta_{n,m} = 0$.

Sufficient conditions to ensure that a Markov chain is β-mixing can be found in Doukhan (1995, Section 2.4). Usually, $\beta_{n,k} = O(\eta^k)$ for some $\eta \in (0,1)$ and one may choose $l_n = O(\log n)$.

(C) For all $k \in \{0, \ldots, r_n\}$ there exists

$$s_n(k) \ge \mathrm{P}\left[X_k > u_n \mid X_0 > u_n\right]$$

such that $\lim_{n\to\infty} \sum_{k=1}^{r_n} s_n(k) = \sum_{k=1}^{\infty} \lim_{n\to\infty} s_n(k) < \infty$.

Typically $s_n(k)$ will be of the form $b_n + c_k$ with $b_n = o(1/r_n)$ and $\sum_{k=1}^{\infty} c_k < \infty$. The interchangeability of the limit and the sum is then automatically fulfilled. For stochastic recurrence equations (Section 3.2), conditions (B) and (C) are verified in Example 8.3 below.

Theorem 5.1. *Suppose that $(X_t)_{t\in\mathbb{Z}}$ is a stationary, regularly varying time series and that the conditions (A(x_0)), (B), and (C) are fulfilled for some $x_0 \ge 0$. Then, for all $y_0 \in [x_0,\infty) \cap (0,\infty)$, the process $\big(\tilde{Z}_n(\phi_1), (\tilde{Z}_n(\phi_{2,x}))_{x\in[x_0,\infty)}, (\tilde{Z}_n(\phi_{3,y}))_{y\in[y_0,\infty)}\big)$ converges weakly to a centered Gaussian process $\tilde{Z}$ with covariance function given by*

$$\begin{aligned}
&\operatorname{cov}\left(\tilde{Z}(\psi_1), \tilde{Z}(\psi_2)\right) \\
&= \mathrm{E}\left[\psi_1(Y_{-1},Y_0,Y_1)\,\psi_2(Y_{-1},Y_0,Y_1)\right] + \sum_{k=1}^{\infty} \Big(\mathrm{E}\left[\psi_1(Y_{-1},Y_0,Y_1)\,\psi_2(Y_{k-1},Y_k,Y_{k+1})\right] \\
&\qquad + \mathrm{E}\left[\psi_2(Y_{-1},Y_0,Y_1)\,\psi_1(Y_{k-1},Y_k,Y_{k+1})\right]\Big)
\end{aligned} \tag{5.1}$$

for all $\psi_1, \psi_2 \in \{\phi_1, \phi_{2,x}, \phi_{3,y} : x \ge x_0, y \ge y_0\}$.

From Theorem 5.1 one may conclude the joint asymptotic normality of the forward and the backward estimator of $F^{(A_1)}$. However, additional conditions are needed to ensure that their bias is asymptotically negligible:

$$\sup_{x\in[x_0,\infty)} \left| \mathrm{P}\left[\frac{X_1}{X_0} \le x \,\middle|\, X_0 > u_n\right] - F^{(A_1)}(x) \right| = o\big((nv_n)^{-1/2}\big), \tag{5.2}$$

$$\sup_{y\in[y_0,\infty)} \left| \mathrm{E}\left[\left(\frac{X_{-1}}{X_0}\right)^{\alpha} \mathbf{1}(X_0/X_1 > y) \,\middle|\, X_0 > u_n\right] - \bar{F}^{(A_1)}(y) \right| = o\big((nv_n)^{-1/2}\big). \tag{5.3}$$

Here $\bar{F}^{(A_1)} := 1 - F^{(A_1)}$ denotes the survival function of A_1 (and hence of Θ_1). These conditions are fulfilled if nv_n tends to ∞ sufficiently slowly, because by definition of the spectral tail process and by (4.4), the left-hand sides tend to 0 if $F^{(A_1)}$ is continuous on $[x_0, \infty)$.

Corollary 5.2. *Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary, regularly varying process with a Markov spectral tail chain. If (A(x_0)), (B), (C),* (5.2) *and* (5.3) *are fulfilled for some $x_0 \geq 0$ and $y_0 \in [x_0,\infty)\cap(0,\infty)$, then*

$$(nv_n)^{1/2}\begin{pmatrix}(\hat{F}_n^{(f,A_1)}(x) - F^{(A_1)}(x))_{x\in[x_0,\infty)}\\ (\hat{F}_n^{(b,A_1)}(y) - F^{(A_1)}(y))_{y\in[y_0,\infty)}\end{pmatrix} \rightsquigarrow \begin{pmatrix}(Z^{(f,A_1)}(x))_{x\in[x_0,\infty)}\\ (Z^{(b,A_1)}(y))_{y\in[y_0,\infty)}\end{pmatrix}, \qquad n\to\infty, \tag{5.4}$$

where the limit is a centered Gaussian process whose covariance function is given by

$$\begin{aligned}\operatorname{cov}\Big(Z^{(f,A_1)}(x),\, Z^{(f,A_1)}(y)\Big) &= \bar{F}^{(A_1)}(\max(x,y)) - \bar{F}^{(A_1)}(x)\bar{F}^{(A_1)}(y),\\ \operatorname{cov}\Big(Z^{(b,A_1)}(x),\, Z^{(b,A_1)}(y)\Big) &= \mathrm{E}\left[\Theta_1^{-1}\mathbf{1}(\Theta_1 > \max(x,y))\right] - \bar{F}^{(A_1)}(x)\bar{F}^{(A_1)}(y),\end{aligned}$$

and

$$\begin{aligned}\operatorname{cov}\Big(Z^{(f,A_1)}(x),\, Z^{(b,A_1)}(y)\Big) &= \sum_{k=1}^{\infty}(\Theta_{k-1}/\Theta_k)^{\alpha}\mathbf{1}(\Theta_1 > x, \Theta_k/\Theta_{k-1} > y, Y_k > 1)\\ &\quad - \bar{F}^{(A_1)}(x)\sum_{k=1}^{\infty}\mathrm{E}\left[(\Theta_{k-1}/\Theta_k)^{\alpha}\mathbf{1}(\Theta_k/\Theta_{k-1} > y, Y_k > 1)\right]\\ &\quad - \bar{F}^{(A_1)}(y)\sum_{k=1}^{\infty}\mathrm{P}\left[\Theta_1 > x, Y_k > 1\right]\\ &\quad + \bar{F}^{(A_1)}(x)\,\bar{F}^{(A_1)}(y)\sum_{k=1}^{\infty}\mathrm{P}\left[Y_k > 1\right].\end{aligned}$$

Remark 5.3. For $x \geq 1$, we have

$$\begin{aligned}\operatorname{var}(Z^{(b,A_1)}(x)) &= \mathrm{E}\left[\Theta_1^{-1}\mathbf{1}(\Theta_1 > x)\right] - \big(\bar{F}^{(A_1)}(x)\big)^2\\ &< \mathrm{P}\left[\Theta_1 > x\right] - \big(\bar{F}^{(A_1)}(x)\big)^2\\ &= \operatorname{var}(Z^{(f,A_1)}(x)),\end{aligned}$$

provided $\bar{F}^{(A_1)}(x) = \mathrm{P}[A_1 > x] = \mathrm{P}[\Theta_1 > x] > 0$. Hence, for such x, when the tail index α is known, the backward estimator is asymptotically more efficient than the forward estimator.

Remark 5.4. While it is not too restrictive to assume that the cdf of A_1 is continuous on $(0,\infty)$, often $\mathcal{L}(A_1)$ has positive mass at 0; see Example 8.2. In this case, one may prove a version of Theorem 5.1 where $\tilde{Z}_n(\phi_{2,x})$ and $\tilde{Z}(\phi_{2,x})$ are replaced by $w(x)\tilde{Z}_n(\phi_{2,x})$ and $w(x)\tilde{Z}(\phi_{2,x})$, respectively, for $w(x) = h\big(F^{(A_1)}(x) - F^{(A_1)}(0)\big)$ and any nondecreasing, continuous function h with $h(0) = 0$. Hence, Corollary 5.2 holds true in this setting if the first coordinate in (5.4) is replaced by

$$(nv_n)^{1/2}\Big(w(x)\big(\hat{F}_n^{(\mathrm{f},A_1)}(x) - F^{(A_1)}(x)\big)\Big)_{x\in[0,\infty)} \rightsquigarrow \Big(w(x)\,Z^{(f,A_1)}(x)\Big)_{x\in[0,\infty)}, \qquad n\to\infty.$$

5.2 Unknown index of regular variation

In most applications, the index of regular variation, α, is unknown. In the definition of the backward estimator $\hat{F}_n^{(\mathrm{b},A_1)}(y)$, it must then be replaced with a suitable estimator. A popular estimator of α is the Hill-type estimator (4.11). More generally, one may consider estimators that can be written in the form

$$\hat{\alpha}_n = \frac{\sum_{i=1}^n \mathbf{1}(X_i > u_n)}{\sum_{i=1}^n \tilde{\psi}(X_i/u_n)\mathbf{1}(X_i > u_n) + R_n} \tag{5.5}$$

with a remainder term $R_n = o_P((nv_n)^{1/2})$ and a suitable function $\tilde{\psi} : [0,\infty) \to [0,\infty)$ which is a.s. continuous w.r.t. $\mathcal{L}(Y_k)$ for all $k \in \mathbb{N}\cup\{0\}$ such that $\mathrm{E}[\tilde{\psi}(Y_0)] = 1/\alpha$ and $\tilde{\psi}(x) = 0$ for all

$x \in [0,1]$. Obviously, the Hill-type estimator is of this form with $\tilde{\psi}(x) = \log(x)\,\mathbf{1}(x > 1)$. Under weak dependence conditions, other well-known estimators like the maximum likelihood estimator in a generalized Pareto model examined by Smith (1987) and the moment estimator suggested by Dekkers et al. (1989) can be written in this way too; see Drees (1998a, Example 4.1) and Drees (1998b, Example 4.1) for similar results in the case of i.i.d. sequences.

Estimators of type (5.5) can be approximated by the ratio of the generalized tail array sums corresponding to the functions ϕ_1 and $\psi(y_{-1}, y_0, y_1) := \tilde{\psi}(y_0)\mathbf{1}(y_0 > 1)$, respectively, and their asymptotic behavior can hence be derived from Theorem 2.3 of Drees and Rootzén (2010). To this end, we replace (C) with the following condition:

(C') For all $0 \le k \le r_n$ there exists

$$s_n(k) \ge \mathrm{E}\left[\max\left(\left|\tilde{\psi}\Big(\frac{X_0}{u_n}\Big)\right|, \mathbf{1}(X_0 > u_n)\right) \max\left(\left|\tilde{\psi}\Big(\frac{X_k}{u_n}\Big)\right|, \mathbf{1}(X_k > u_n)\right) \,\Big|\, X_0 > u_n\right] \quad (5.6)$$

such that $\lim_{n\to\infty} \sum_{k=1}^{r_n} s_n(k) = \sum_{k=1}^{\infty} \lim_{n\to\infty} s_n(k) < \infty$.

Moreover, there exists $\delta > 0$ such that

$$\sum_{k=1}^{r_n} \left(\mathrm{E}\left[\left|\tilde{\psi}\Big(\frac{X_0}{u_n}\Big)\tilde{\psi}\Big(\frac{X_k}{u_n}\Big)\right|^{1+\delta} \,\Big|\, X_0 > u_n\right]\right)^{1/(1+\delta)} = O(1), \qquad n \to \infty. \quad (5.7)$$

If $\tilde{\psi}$ is bounded, then (C') follows from condition (C), but in general it is more restrictive, though it can often be established by similar arguments. For example, condition (C') holds for the solutions to the stochastic recurrence equation studied in Example 8.3 and the Hill estimator, i.e., for $\tilde{\psi}(x) = \log(x)\,\mathbf{1}(x > 1)$.

The following result gives the asymptotic normality of $\hat{\alpha}_n$ centered at

$$\alpha_n := 1/\,\mathrm{E}[\tilde{\psi}(X_0/u_n) \mid X_0 > u_n]. \quad (5.8)$$

This quantity tends to α as $n \to \infty$ by the assumptions on the function $\tilde{\psi}$ and condition (C').

Lemma 5.5. *If $\hat{\alpha}_n$ is of the form* (5.5) *and if the conditions (B) and (C') hold, then*

$$(nv_n)^{1/2}(\hat{\alpha}_n - \alpha_n) = \alpha\tilde{Z}_n(\phi_1) - \alpha^2\tilde{Z}_n(\psi) + o_P(1) \rightsquigarrow \alpha\tilde{Z}(\phi_1) - \alpha^2\tilde{Z}(\psi), \qquad n \to \infty,$$

for a centered Gaussian process $\tilde{Z}$ with covariance function given by (5.1).

Similarly as in (5.2) and (5.3), we need an extra condition to ensure that the bias is asymptotically negligible:

$$\big|\mathrm{E}[\tilde{\psi}(X_0/u_n) \mid X_0 > u_n] - 1/\alpha\big| = o\big((nv_n)^{-1/2}\big), \qquad n \to \infty. \quad (5.9)$$

Now we are ready to state the asymptotic normality of the backward estimator with estimated index α, i.e.,

$$\hat{F}_n^{(\hat{b},A_1)}(x) := 1 - \frac{\sum_{i=1}^n \left(\frac{X_{i-1}}{X_i}\right)^{\hat{\alpha}_n} \mathbf{1}\left(X_i/X_{i-1} > x,\ X_i > u_n\right)}{\sum_{i=1}^n \mathbf{1}\left(X_i > u_n\right)}.$$

Corollary 5.6. *Suppose that the conditions of Corollary 5.2 and of Lemma 5.5 are fulfilled and that* (5.9) *holds. Then*

$$(nv_n)^{1/2}\begin{pmatrix}(\hat{F}_n^{(f,A_1)}(x) - F^{(A_1)}(x))_{x\in[x_0,\infty)} \\ (\hat{F}_n^{(\hat{b},A_1)}(y) - F^{(A_1)}(y))_{y\in[y_0,\infty)}\end{pmatrix} \rightsquigarrow \begin{pmatrix}(Z^{(f,A_1)}(x))_{x\in[x_0,\infty)} \\ (Z^{(\hat{b},A_1)}(y))_{y\in[y_0,\infty)}\end{pmatrix}, \qquad n \to \infty, \quad (5.10)$$

with

$$Z^{(f,A_1)}(x) = \tilde{Z}(\phi_{2,x}) - \bar{F}^{(A_1)}(x)\tilde{Z}(\phi_1),$$
$$Z^{(\hat{b},A_1)}(y) = \tilde{Z}(\phi_{3,y}) - \bar{F}^{(A_1)}(y)\tilde{Z}(\phi_1) + (\alpha^2\tilde{Z}(\psi) - \alpha\tilde{Z}(\phi_1))\,\mathrm{E}\left[\log(\Theta_1)\,\mathbf{1}(\Theta_1 > y)\right].$$

The covariance function of the limiting process can be calculated in the same way as in the proof of Corollary 5.2. In general, the resulting expressions will involve sums over all $k \in \mathbb{N}$. Moreover, it is no longer guaranteed that the backward estimator of $F^{(A_1)}(y)$ at $y > 1$ has a smaller variance than the forward estimator.

Remark 5.7. For Markovian time series which are not necessarily positive, the forward and backward estimators of $F^{(A_1)}$ and $F^{(B_1)}$ can be represented in terms of generalized tail array sums constructed from

$$X_{n,i} := \frac{(X_{i-1}, X_i, X_{i+1})}{u_n} \mathbf{1}(|X_i| > u_n).$$

When $x < 0$, for example, the backward estimator $\hat{F}_n^{(\mathrm{b},B_1)}(x)$ equals the ratio of the generalized tail array sums pertaining to

$$\begin{aligned} \phi_{4,x}(y_{-1}, y_0, y_1) &:= (-y_{-1}/y_0)^\alpha \, \mathbf{1}(y_0/y_{-1} \le x, y_0 > 1), \\ \phi_5(y_{-1}, y_0, y_1) &:= \mathbf{1}(y_0 < -1). \end{aligned}$$

Hence their limit processes can be obtained by the same methods as in the case $X_t > 0$ under obvious analogues to the conditions (A(x_0)), (B) and (C) with $v_n := \mathrm{P}\left[|X_0| > u_n\right]$.

6 Monte Carlo simulation

To compare the finite-sample performance of the forward, backward, and mixture estimators, we show results from a simulation study. We use both approaches discussed at the end of Section 4 for dealing with the problem that α is unknown.

Pseudo-random samples are generated from two different time series models.

- For the copula Markov model (Section 3.1), we choose the symmetric t-distribution with $\nu_1 = 2$ degrees of freedom as the margin G and the t-copula $C^t_{\nu_2,\rho}(u, v)$ with $\nu_2 = 2.5$ and $\rho = 0.2$ to model the temporal dependence structure. Hence, $\mathrm{P}\left[\Theta_0 = 1\right] = \mathrm{P}\left[\Theta_0 = -1\right] = 1/2$ and the index of regular variation is equal to 2. The distribution function of A_1 is calculated in Example 8.1. The simulation algorithm is described in Chen et al. (2009, Section 6.2).
- For the stochastic recurrence equation (Section 3.2), we let C_t and D_t be independent $N(1/3, 8/9)$ and $N(-10, 1)$ random variables, respectively. This choice ensures that $\mathrm{E}\left[C_t^2\right] = 1$ and that the sufficient conditions of Theorem 5 in Kesten (1973) hold. As a consequence, a stationary solution $(X_t)_{t\in\mathbb{Z}}$ exists which is regularly varying with index $\alpha = 2$.

For each model, we generate time series $(X_i)_{1\le i\le n}$ of length $n = 2\,000$ and set the threshold u_n to be the 97.5% quantile of the absolute values of the sample, i.e., we use 50 extremes for estimation. Based on 1 000 Monte Carlo repetitions, we estimate the bias, the standard deviation (SD) and the root mean squared error (RMSE) of all estimators under consideration. Note that both models are symmetric in the sense that $\mathcal{L}(A_1) = \mathcal{L}(B_1)$. Moreover, the estimators of the cdf's of A_1 and B_1 are identically distributed for the copula model, and they behave similarly for the solutions to the stochastic recurrence equation. Therefore, we report the results only for A_1.

Figure 2 shows the results for the estimators of the cdf's A_1 and A_1^* in the copula Markov model. For this model, $\mathrm{Bias}(\hat{p}) = 0.001$, $\mathrm{SD}(\hat{p}) = 0.08$, and $\mathrm{RMSE}(\hat{p}) = 0.08$. In addition, for the simulations with estimated α, we obtain $\mathrm{Bias}(\hat{\alpha}) = 0.077$, $\mathrm{SD}(\hat{\alpha}) = 0.407$, and $\mathrm{RMSE}(\hat{\alpha}) = 0.414$. In the top row, the Hill-type estimator (4.11) is used, whereas in the bottom row, we apply the signed rank transformation (4.12). The plots on the left and in the middle show the bias and the standard deviation of the forward estimator as a dotted line, of the backward estimator as a dashed line and of the mixture estimator as a solid line. The plots on the right-hand side display the relative RMSE of the backward and the mixture estimator w.r.t. the forward estimator.

As expected, the backward estimator outperforms the forward estimator for $|x| \ge 1$, whereas for arguments x close to 0 it is the forward estimator that has the lower RMSE. The mixture estimator performs much better than the backward estimator for small x, while for large values of

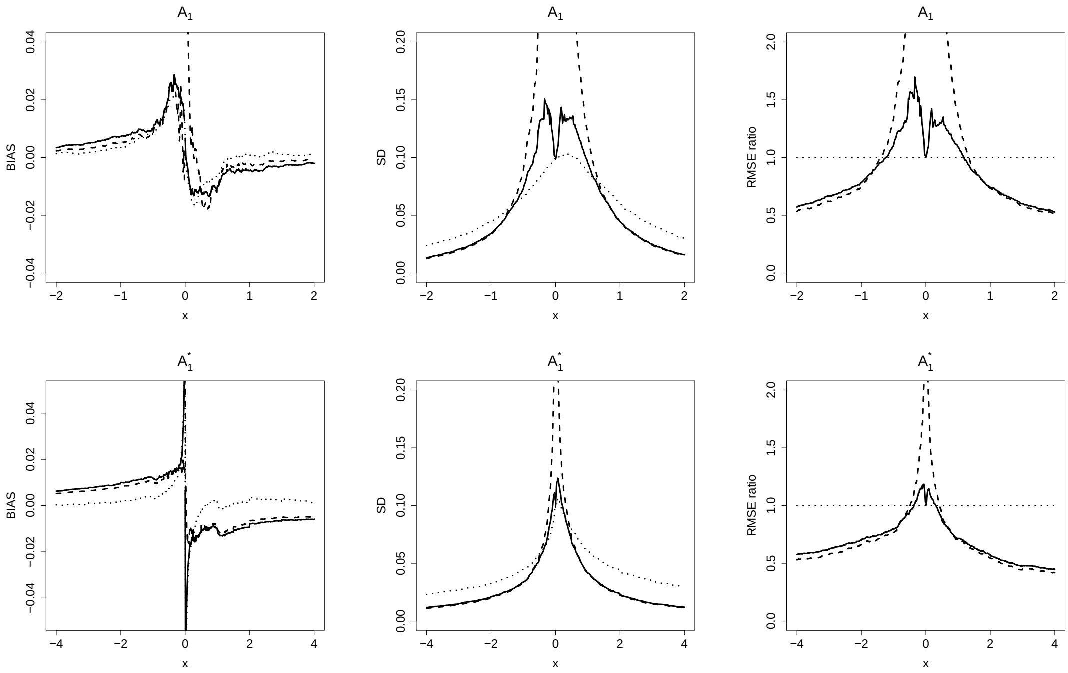

Figure 2: Simulation results for the copula Markov model. Top: Estimation of the cdf of A_1 using the plug-in estimator for α. Bottom: Estimation of the cdf of $A_1^* = \text{sign}(A_1)|A_1|^\alpha$ with the signed rank transformation. Left: bias; middle: standard deviation; right: root mean squared error ratio with respect to the forward estimator. Solid line: mixture estimator; dashed line: backward estimator; dotted line: forward estimator.

$|x|$ its RMSE is similar to the one of the backward estimator. Indeed, in the approach using rank transforms, the RMSE of the mixture estimator is never much larger than that of the forward estimator, while the mixture estimator is almost twice as efficient as the forward estimator for $|x| > 2$. Hence, it clearly outperforms both the forward and the backward estimator in this approach. In contrast, the mixture estimator performs worse than the forward estimator for small values of x if α is replaced with the Hill-type estimator.

Figure 3 shows the analogous results for the solutions of the stochastic recurrence equation. We obtain $\text{Bias}(\hat{p}) = -0.209$, $\text{SD}(\hat{p}) = 0.083$, $\text{RMSE}(\hat{p}) = 0.225$, and $\text{Bias}(\hat{\alpha}) = 0.173$, $\text{SD}(\hat{\alpha}) = 0.526$, $\text{RMSE}(\hat{\alpha}) = 0.553$. By and large, the relative performance of all estimators is similar to the one in the copula Markov model. The relative efficiency of the mixture estimator with respect to the forward estimator is a bit higher for small values of $|x|$, while it is slightly worse for larger values. Overall, the absolute estimation errors of all estimators are about 20 to 30% larger.

7 Case study

The spectral tail process of a heavy tailed time series conveys important information on its serial extremal dependence. Such extremal dependence typically arises e.g. in financial time series which exhibit clustering of extremes. By estimating the joint distribution of (Θ_0, Θ_1) and (Θ_0, Θ_{-1}), we gain insight into the dependence between extremes of consecutive observations, covering both lower and upper tails.

An interesting question is whether the distribution of a time series remains unaffected by reversing the direction of time. It is well understood how to test for such time reversibility regarding the bulk of the distribution (Beare and Seo, 2014; Chen et al., 2000). Financial time series do not have this feature in general (Chen and Kuan, 2002). This, however, does not imply that there is no time reversibility at extreme levels. Formally, such time reversibility would mean

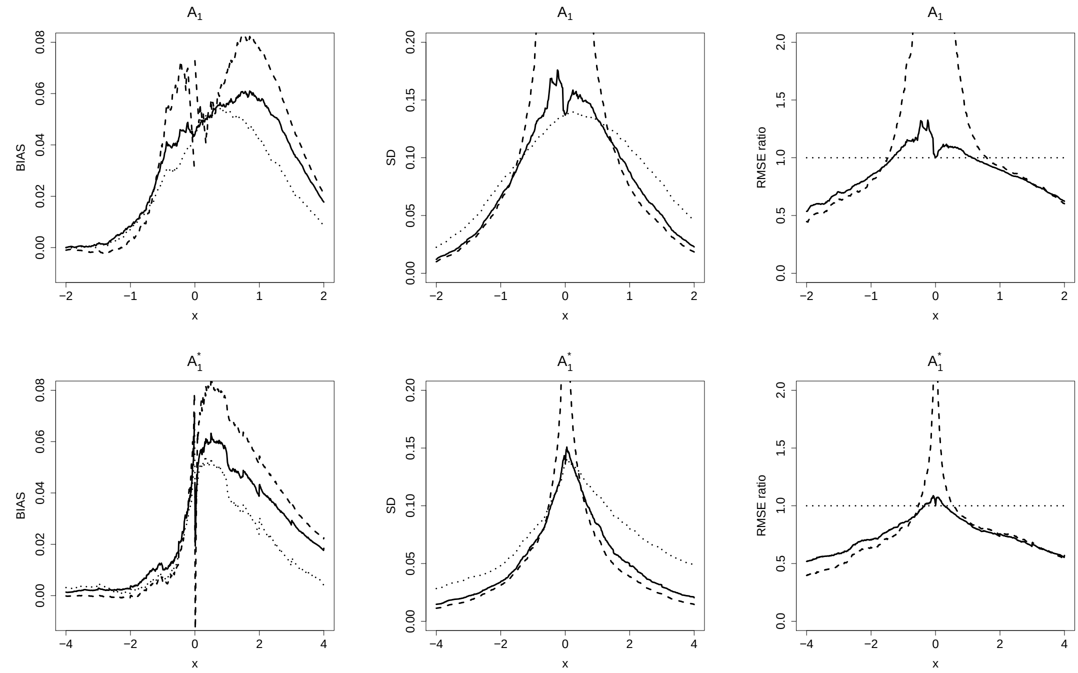

Figure 3: Simulation results for the stochastic recurrence equations. Top: Estimation of the cdf of A_1 using the plug-in estimator for α. Bottom: Estimation of the cdf of $A_1^* = \operatorname{sign}(A_1)|A_1|^\alpha$ with the signed rank transformation. Left: bias; middle: standard deviation; right: root mean squared error ratio with respect to the forward estimator. Solid line: mixture estimator; dashed line: backward estimator; dotted line: forward estimator.

that $\mathcal{L}(\Theta_0, \Theta_1) = \mathcal{L}(\Theta_0, \Theta_{-1})$ and, equivalently, $\mathcal{L}(A_1) = \mathcal{L}(A_{-1})$ and $\mathcal{L}(B_1) = \mathcal{L}(B_{-1})$, or $\mathcal{L}(A_1^*) = \mathcal{L}\left(A_{-1}^*\right)$ and $\mathcal{L}(B_1^*) = \mathcal{L}\left(B_{-1}^*\right)$.

We analyze daily log-returns of Google and UBS stock prices between 2005-01-03 and 2013-12-31 (taken from `www.google.com/finance`), leading to 2279 observations for Google and 2280 for UBS. The thresholds are set at the 95% quantiles, giving 114 extremes. For Google log-returns we obtain 53 positive extremes and 61 negative extremes, whereas for UBS it is 50 and 64 positive and negative extremes respectively. The estimated index of regular variation is equal to 2.88 for Google and 2.51 for UBS.

We discuss the results jointly for both stocks, as their extremal dependence structures exhibit similar patterns. The two data series are presented in Figure 4. The three rows show the daily closing prices (top), the daily log-returns X_i (middle) and the pertaining rank transformed log-returns $X_{n,i}^*$ as in (4.12) (bottom).

First, we analyze the absolute values of the transformed log-returns. The mixture estimates of the distribution functions of A_1^* and A_{-1}^*, shown in Figure 5, look similar, suggesting that at extreme levels, the absolute values of log-returns are time reversible.

Figure 6 displays the estimated distribution functions of A_1^*, A_{-1}^*, B_1^*, and B_{-1}^* for the rank transformed log-returns based on the mixture estimator. Since the cdf's of A_1^* and A_{-1}^* as well as the cdf's of B_1^* and B_{-1}^* apparently differ on the negative real line, the extreme values of the log-returns exhibit no time reversibility.

Most of the probability mass of A_1^* and B_{-1}^* is concentrated near the origin. This hints at asymptotic independence of X_t^+ and X_{t+1}, and of X_t and X_{t+1}^-. In contrast, the laws of A_{-1}^* and B_1^* put considerable mass on the negative half-line. This indicates that with asymptotically non-negligible probability a large loss is succeeded by a large gain. Such an event can be interpreted as a correction to an overreaction by the stock market. Indeed, the existence of long–term and short–term overreaction behaviour in various stock markets is well documented (De Bondt and

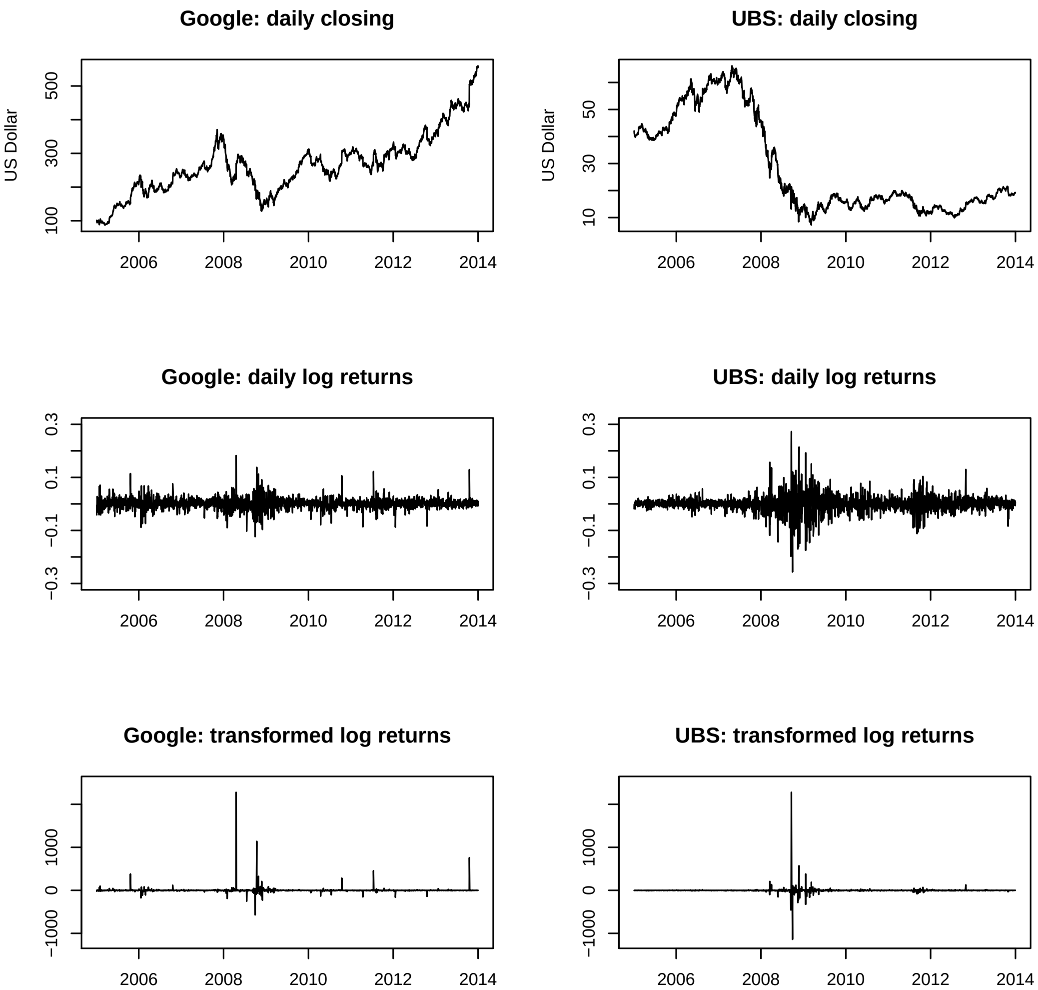

Figure 4: Google and UBS daily stock prices, daily log-returns and transformed log-returns.

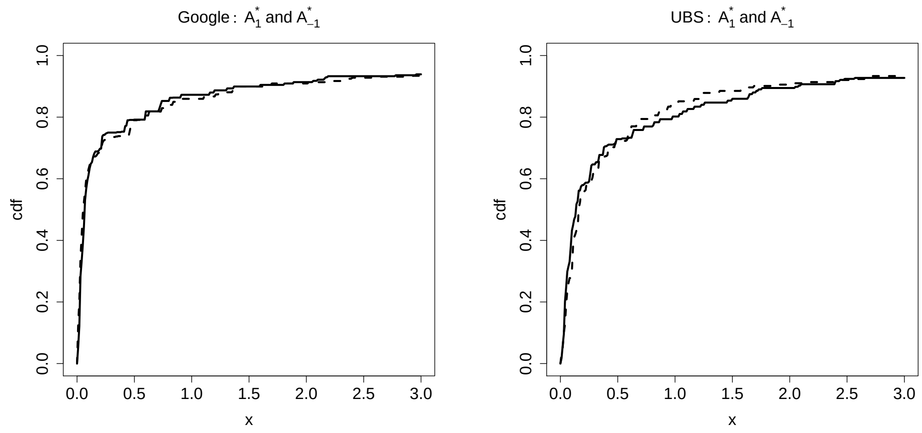

Figure 5: Estimated distribution functions of A_1^* and A_{-1}^* based on the mixture estimator for absolute values of transformed log-returns on Google (left) and UBS (right) stock prices. Solid lines: A_1^*; dashed lines: A_{-1}^*.

Thaler, 1985; Bowman and Iverson, 1998; Norli, 2010).

8 Appendix

8.1 Examples

Example 8.1 (The t-copula). In the context of the copula Markov model (Section 3.1), consider the bivariate t-copula defined by

$$C^t_{\nu,\rho}(u,v) = \int_{-\infty}^{t_\nu^{-1}(u)} \int_{-\infty}^{t_\nu^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \left\{ 1 + \frac{x^2 - 2\rho xy + y^2}{\nu\,(1-\rho^2)} \right\}^{-(\nu+2)/2} \mathrm{d}x\,\mathrm{d}y,$$

where $\rho \in (-1,1)$, t_ν is the cdf of the univariate t-distribution with $\nu > 0$ degrees of freedom and t_ν^{-1} is the corresponding quantile function. Contrary to the Gaussian copula, the t-copula allows for an asymptotically non-negligible probability of joint extremes. The t-copula is exchangeable and radially symmetric. The partial derivative of $C^t_{\nu,\rho}(u,v)$ with respect to the first coordinate equals (Demarta and McNeil, 2005, proof of Proposition 4)

$$\frac{\partial}{\partial u} C^t_{\nu,\rho}(u,v) = t_{\nu+1}\left(\left\{ \frac{t_\nu^{-1}(v)}{t_\nu^{-1}(u)} - \rho \right\} \left(\frac{\nu+1}{1-\rho^2} \right)^{1/2} \left\{ 1 + \frac{\nu}{\left(t_\nu^{-1}(u)\right)^2} \right\}^{-1/2} \operatorname{sign}(t_\nu^{-1}(u)) \right). \tag{8.1}$$

The limits (3.8) and (3.9) can be calculated using symmetry of the t-distribution and the regular variation of t_ν with index ν. Under the assumptions of Proposition 3.1, we obtain

$$\mathrm{P}\left[A_1 \le x\right] = \begin{cases} t_{\nu+1}\left(\left(x^{\alpha/\nu} - \rho\right) \left(\frac{\nu+1}{1-\rho^2}\right)^{1/2} \right) & \text{if } x \ge 0, \\ t_{\nu+1}\left(\left(-(\frac{1-p}{p}\,|x|^{-\alpha})^{-1/\nu} - \rho\right) \left(\frac{\nu+1}{1-\rho^2}\right)^{1/2} \right) & \text{if } x < 0. \end{cases} \tag{8.2}$$

As the t-copula is exchangeable, we find $\mathcal{L}(A_1) = \mathcal{L}(A_{-1})$ and $\mathcal{L}(B_1) = \mathcal{L}(B_{-1})$. Moreover, by radial symmetry we have $\eta_{1,1} = 1 - \eta_{0,0}$ and $\eta_{1,0} = 1 - \eta_{0,1}$, linking up the laws of A_1 and B_1 through (3.12) and (3.13). In particular, if $p = 1/2$, then $\mathcal{L}(A_1) = \mathcal{L}(B_1)$.

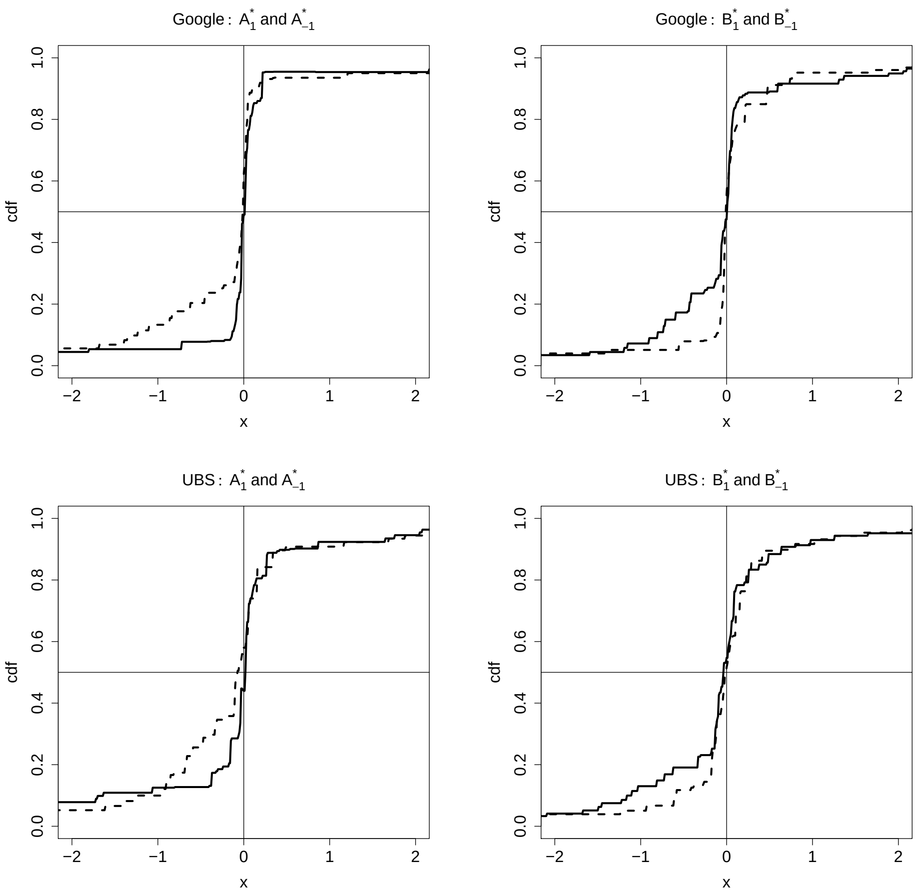

Figure 6: Estimated distribution functions of A_1^* and A_{-1}^* (left) and of B_1^* and B_{-1}^* (right) based on the mixture estimator for transformed log-returns on Google stock prices (top) and UBS stock prices (bottom). Solid lines: A_1^* and B_1^*; dashed lines: A_{-1}^* and B_{-1}^*.

Example 8.2 (Extreme–value copulas). A bivariate copula $C(u,v)$ is an extreme value copula if and only if it admits the representation

$$C(u,v) = \exp\left\{ D\left(\frac{\log v}{\log(uv)}\right) \log(uv) \right\}, \qquad (u,v) \in (0,1]^2 \setminus \{(1,1)\}.$$

The *Pickands dependence function* $D : [0,1] \to [1/2,1]$ is convex and satisfies $\max(w, 1-w) \le f(w) \le 1$ for all $w \in [0,1]$. See Gudendorf and Segers (2010) for a survey of extreme-value copulas.

If D is continuously differentiable with derivative D', then, under the conditions of Proposition 3.1,

$$\mathrm{P}[A_1 \le x] = \left\{ D\left(\frac{1}{x^\alpha + 1}\right) - \frac{1}{x^\alpha + 1} D'\left(\frac{1}{x^\alpha+1}\right) \right\} \mathbf{1}(x \ge 0),$$

whereas B_1 and B_{-1} are degenerate at 0. In particular, $\mathrm{P}[A_1 = 0] = 1 - D'(0)$. The law of A_{-1} has the same form as the one of A_1 upon replacing D by $w \mapsto D(1-w)$. If $D(w) = D(1-w)$ then C is exchangeable and $\mathcal{L}(A_{-1}) = \mathcal{L}(A_1)$.

The following two parametric families are well known:

- For the *asymmetric logistic model* (Tawn, 1988) with parameters $\theta \ge 1$ and $\psi_1, \psi_2 \in (0,1]$, we have

$$D(w) = (1-\psi_1)w + (1-\psi_2)(1-w) + \{(\psi_1 w)^\theta + (\psi_2(1-w))^\theta\}^{1/\theta},$$
$$\mathrm{P}[A_1 \le x] = 1 - \psi_2 + \psi_2 \left\{ 1 + (\psi_1/(\psi_2 x^\alpha))^\theta \right\}^{(1-\theta)/\theta}, \qquad x \ge 0.$$

  The special case $\psi_1 = \psi_2 = 1$ yields the Gumbel–Hougaard copula.

- For the *asymmetric negative logistic model* (Joe, 1990) with parameters $\theta > 0$ and $\psi_1, \psi_2 \in (0,1]$, we have

$$D(w) = 1 - \{(\psi_1 w)^{-\theta} + (\psi_2(1-w))^{-\theta}\}^{-1/\theta},$$
$$\mathrm{P}[A_1 \le x] = 1 - \psi_2 \left\{ 1 + (\psi_2 x^\alpha/\psi_1)^\theta \right\}^{-(1+\theta)/\theta}, \qquad x \ge 0.$$

  The special case $\psi_1 = \psi_2 = 1$ yields the Galambos copula.

Example 8.3 (Stochastic recurrence equations). We will show that stationary solutions to the stochastic recurrence equation (3.14) with $(C_t, D_t) \in [0,\infty)^2$ satisfy the conditions (B) and (C) in Section 5.1. Asymptotic normality of the forward and backward estimators (Corollary 5.2) follows if the cdf of $\mathcal{L}(A_1) = \mathcal{L}(C_1)$ is continuous on $[x_0, \infty)$.

We assume that the conditions in Kesten (1973) are fulfilled (see Section 2). Then $(X_t)_{t\in\mathbb{Z}}$ is geometrically β-mixing, i.e., there exist constants $\eta \in (0,1)$ and $\tau > 0$ such that $\beta_{n,k} \le \tau \eta^k$ (Doukhan, 1995, Corollary 2.4.1). Therefore, condition (B) is satisfied with $l_n = 2\log n / |\log \eta|$ and suitably chosen $r_n = o(\min\{(nv_n)^{1/2}, v_n^{-1}\})$, provided $v_n = o(1/\log n)$ and $(\log n)^2/n = o(v_n)$.

To establish condition (C), let $\Pi_{i+1,j} := \prod_{k=i+1}^{j} C_k$ and $V_k := \sum_{j=1}^{k} \Pi_{j+1,k} D_j$. Iterating (3.14) yields $X_k = V_k + \Pi_{1,k} X_0$. By independence of $(V_k, \Pi_{1,k})$ and X_0, one has

$$\begin{aligned}
\mathrm{P}[X_k > u_n \mid X_0 > u_n] &\le v_n^{-1}\,\mathrm{P}[X_0 > u_n, V_k > u_n/2] + v_n^{-1}\,\mathrm{P}[X_0 > u_n, \Pi_{1,k} X_0 > u_n/2] \\
&= \mathrm{P}[V_k > u_n/2] + v_n^{-1} \int_{u_n}^{\infty} \mathrm{P}[\Pi_{1,k} > u_n/(2t)]\, \mathcal{L}(X_0)(\mathrm{d}t).
\end{aligned}$$

There exists $\xi \in (0,\alpha)$ such that $\rho := \mathrm{E}[C_1^\xi] < 1$. Thus $\mathrm{P}[\Pi_{1,k} > u_n/(2t)] \le \mathrm{E}[\Pi_{1,k}^\xi](2t/u_n)^\xi = \rho^k (2t/u_n)^\xi$, which in turn implies

$$v_n^{-1} \int_{u_n}^{\infty} \mathrm{P}[\Pi_{1,k} > u_n/(2t)]\, \mathcal{L}(X_0)(\mathrm{d}t) \le \rho^k\, \mathrm{E}\left[(2X_0/u_n)^\xi \mid X_0 > u_n\right] \le 2^{\xi+1}\,\mathrm{E}[Y_0^\xi]\rho^k$$

for all $k \in \mathbb{N}$ and sufficiently large n. Because $\mathrm{P}[V_k > u_n/2] \le \mathrm{P}[X_k > u_n/2] \le 2^{1-\alpha} v_n$ for all $k \in \mathbb{N}$ and sufficiently large n, one may conclude that, for some constant $c > 0$,

$$\mathrm{P}[X_k > u_n \mid X_0 > u_n] \le c(v_n + \rho^k) =: s_n(k).$$

Condition (C) then follows from the fact that, as $n \to \infty$,

$$\sum_{k=1}^{r_n} s_n(k) = c r_n v_n + c \sum_{k=1}^{r_n} \rho^k \;\to\; c \sum_{k=1}^{\infty} \rho^k = \sum_{k=1}^{\infty} \lim_{n\to\infty} s_n(k) < \infty.$$

8.2 Proofs

Proof of Lemma 2.1. Note that $|X_t^*| \ge u$ if and only if $F_{|X_0|}(|X_t|) \ge 1 - 1/u$ if and only if $|X_t| \ge U(u) := F^{\leftarrow}_{|X_0|}(1 - 1/u)$ with $F_{|X_0|}$ denoting the distribution function of $|X_t|$. Moreover, regular variation of $|X_0|$ implies regular variation of U as well as $\mathrm{P}[|X_0| > u] \,/\, \mathrm{P}[|X_0| \ge u] \to 1$ and $u\,\bar{F}_{|X_0|}(U(u)) \to 1$ as $u \to \infty$. Thus, from weak convergence (2.2) we may conclude

$$\begin{aligned}
&\mathcal{L}\left(\left(\frac{X_i^*}{u} \right)_{s \le i \le t} \,\middle|\, |X_0^*| \ge u \right) \\
&= \mathcal{L}\left(\left(\frac{X_i/|X_i|}{u\,\bar{F}_{|X_0|}(|X_i|)} \cdot \mathbf{1}(X_i \ne 0) \right)_{s \le i \le t} \,\middle|\, |X_0| \ge U(u) \right) \\
&= \mathcal{L}\left(\left(\frac{X_i/U(u)}{|X_i/U(u)|} \cdot \mathbf{1}\left(\frac{X_i}{U(u)} \ne 0 \right) \cdot \frac{1}{u\,\bar{F}_{|X_0|}(U(u))} \cdot \frac{\bar{F}_{|X_0|}(U(u))}{\bar{F}_{|X_0|}\left(U(u) \frac{|X_i|}{U(u)} \right)} \right)_{s \le i \le t} \,\middle|\, |X_0| \ge U(u) \right) \\
&\rightsquigarrow \mathcal{L}\left(\left(\frac{Y_i}{|Y_i|} \cdot \mathbf{1}(Y_i \ne 0) \cdot 1 \cdot \frac{1}{|Y_i|^{-\alpha}} \right)_{s \le i \le t} \right), \qquad u \to \infty,
\end{aligned}$$

where in the last step we applied the extended continuous mapping theorem 1.11.1 of van der Vaart and Wellner (1996). Therefore, $(X_t^*)_{t\in\mathbb{Z}}$ has a tail process $(Y_t^*)_{t\in\mathbb{Z}}$ with $Y_t^* = \operatorname{sign}(Y_t)|Y_t|^{\alpha}$ and is thus regular varying with index $\alpha^* = 1$. Moreover, $\Theta_t^* = Y_t^*/|Y_0^*| = \operatorname{sign}(\Theta_t)|\Theta_t|^{\alpha}$. □

Proof of Proposition 3.1. Let g be the density function of G. For real x and u,

$$\mathrm{P}[X_1/X_0 \le x \mid X_0 > u] = \int_u^{\infty} \mathrm{P}[X_1/X_0 \le x \mid X_0 = y]\, \frac{g(y)}{1 - G(u)}\, \mathrm{d}y.$$

As a consequence, $\lim_{u\to\infty} \mathrm{P}[X_1/X_0 \le x \mid X_0 > u] = \lim_{y\to\infty} \mathrm{P}[X_1/X_0 \le x \mid X_0 = y]$, provided the latter limit exists.

We compute the conditional distribution of X_1 given $X_0 = y$. For $(x_0, x_1) \in \mathbb{R}^2$, as $G(-\infty) = 0$ and $C(0,\,\cdot\,) = 0$, we have

$$\begin{aligned}
\mathrm{P}[X_0 \le x_0, X_1 \le x_1] &= C\big(G(x_0), G(x_1)\big) - C\big(G(-\infty), G(x_1)\big) \\
&= \int_{-\infty}^{x_0} \dot{C}_1\big(G(y), G(x_1)\big)\, g(y)\, \mathrm{d}y \\
&= \mathrm{E}\Big[\mathbf{1}\{X_0 \le x_0\}\, \dot{C}_1\big(G(X_0), G(x_1)\big) \Big].
\end{aligned}$$

We find that a version of the conditional distribution of X_1 given $X_0 = y$ is given by

$$\mathrm{P}[X_1 \le x_1 \mid X_0 = y] = \dot{C}_1\big(G(y),\, G(x_1)\big), \qquad x_1 \in \mathbb{R}. \tag{8.3}$$

Assume $p > 0$. As $s \searrow 0$, the functions $[0,\infty) \to \mathbb{R} : z \mapsto \dot{C}_1(1 - s, 1 - sz)$ are assumed to converge pointwise to a continuous limit. Since these functions are monotone, the convergence holds locally uniformly, i.e., if $0 \le z(s) \to z$ as $s \searrow 0$, then $\lim_{s\searrow 0} \dot{C}_1(1-s, 1-s\,z(s)) = \lim_{s\searrow 0} \dot{C}_1(1-s, 1-s\,z) = \eta_{1,1}(z)$.

Moreover, the function $\bar{G} = 1 - G$ is regularly varying at infinity with index $-\alpha$, too. Indeed, as $u \to \infty$ we have $\mathrm{P}[X_0 > u] / \mathrm{P}[|X_0| > u] = p > 0$, whereas the function $u \mapsto \mathrm{P}[|X_0| > u]$ is regularly varying at infinity with index $-\alpha$.

For $x \in (0, \infty)$, we find

$$\begin{aligned} \mathrm{P}[X_1/X_0 \le x \mid X_0 = y] &= \dot{C}_1\big(G(y),\, G(xy)\big) \\ &= \dot{C}_1\left(1 - \bar{G}(y),\, 1 - \bar{G}(y)\,\frac{\bar{G}(xy)}{\bar{G}(y)}\right) \\ &\to \lim_{s\searrow 0} \dot{C}_1(1-s,\, 1-s\,x^{-\alpha}) = \eta_{1,1}(x^{-\alpha}), \qquad y \to \infty. \end{aligned}$$

Similarly, $\lim_{y\to\infty} G(-y)/\bar{G}(y) = (1-p)/p$ implies, for $x \in (-\infty, 0)$,

$$\begin{aligned} \mathrm{P}[X_1/X_0 \le x \mid X_0 = y] &= \dot{C}_1\big(G(y),\, G(-|x|y)\big) \\ &= \dot{C}_1\left(1 - \bar{G}(y),\, \bar{G}(y)\,\frac{G(-|x|y)}{\bar{G}(|x|y)}\,\frac{\bar{G}(|x|y)}{\bar{G}(y)}\right) \\ &\to \lim_{s\searrow 0} \dot{C}_1\big(1-s,\, s\,[(1-p)/p]\,|x|^{-\alpha}\big) = \eta_{1,0}([(1-p)/p]\,|x|^{-\alpha}), \end{aligned}$$

as $y \to \infty$. We conclude that $\mathcal{L}(X_1/X_0 \mid X_0 = y)$ converges weakly, as $y \to \infty$, to the distribution $\mathcal{L}(A_1)$ given by (3.12). By the argument at the beginning of the proof, the same then holds true for $\mathcal{L}(X_1/X_0 \mid X_0 > u)$ as $u \to \infty$.

The proof of (3.13) is entirely similar. □

Proof of Lemma 3.2. For real x, write $x^+ = \max(x, 0)$ and $x^- = \max(-x, 0)$. By the time-change formula (2.8),

$$\mathrm{E}\left[(\Theta_1^+)^\alpha\right] = \mathrm{P}[\Theta_0 = +1,\, \Theta_{-1} \ne 0], \qquad \mathrm{E}\left[(\Theta_1^-)^\alpha\right] = \mathrm{P}[\Theta_0 = -1,\, \Theta_{-1} \ne 0].$$

(The above equations even hold without conditions (i) and (ii).) Adding both identities yields, in view of (i),

$$\mathrm{P}[\Theta_{-1} \ne 0] = \mathrm{E}[|\Theta_1|^\alpha] = 1$$

and thus $\mathrm{E}\left[(\Theta_1^+)^\alpha\right] = \mathrm{P}[\Theta_0 = +1] = p$.

For $M := \Theta_1/\Theta_0$, let $\mu_+ = \mathrm{E}[(M^+)^\alpha]$ and $\mu_- = \mathrm{E}[(M^-)^\alpha]$. Then

$$\mu_+ + \mu_- = \mathrm{E}[|M|^\alpha] = \mathrm{E}[|\Theta_1|^\alpha] = 1.$$

By (ii), we have

$$\begin{aligned} p = \mathrm{E}\left[(\Theta_1^+)^\alpha\right] = \mathrm{E}\left[((M\Theta_0)^+)^\alpha\right] &= p\,\mathrm{E}\left[(M^+)^\alpha\right] + (1-p)\,\mathrm{E}\left[(M^-)^\alpha\right] \\ &= p\,(1-\mu_-) + (1-p)\,\mu_-. \end{aligned}$$

After simplification, we find

$$(1-2p)\,\mu_- = 0,$$

so that either $p = 1/2$ or $\mu_- = 0$, i.e., $\mathrm{P}[M < 0] = 0$. This yields the first statement.

Second, we show that Θ_{-1}/Θ_0 is independent of Θ_0, too. If $p \in \{0, 1\}$, then Θ_0 is degenerate; without loss of generality, assume that $0 < p < 1$. We need to show that, for bounded measurable functions $g : \mathbb{R} \to \mathbb{R}$,

$$\mathrm{E}[g(\Theta_{-1}/\Theta_0) \mid \Theta_0 = +1] = \mathrm{E}[g(\Theta_{-1}/\Theta_0) \mid \Theta_0 = -1].$$

Equivalently, we need to show that, for such g,

$$(1-p)\,\mathrm{E}[g(\Theta_{-1}/\Theta_0)\,\mathbf{1}(\Theta_0 = +1)] = p\,\mathrm{E}[g(\Theta_{-1}/\Theta_0)\,\mathbf{1}(\Theta_0 = -1)]. \tag{8.4}$$

The above formula clearly holds for constant functions g. Hence, it holds for a function g as soon as it holds for the function $y \mapsto g(y) - g(0)$. Without loss of generality, we may therefore assume that $g(0) = 0$. But then, by the time-change formula (2.8) applied to the two functions $f_\pm(\theta_0, \theta_1) = g(\theta_0/\theta_1)\,\mathbf{1}(\theta_1 = \pm 1)$ and using independence of $M = \Theta_1/\Theta_0$ and Θ_0,

$$\begin{aligned}
\mathrm{E}\left[g(\Theta_{-1}/\Theta_0)\,\mathbf{1}(\Theta_0 = +1)\right] &= \mathrm{E}\left[g(\Theta_0/\Theta_1)\,(\Theta_1^+)^\alpha\right] \\
&= \mathrm{E}\left[g(\Theta_0/\Theta_1)\,((\Theta_1/\Theta_0)^+)^\alpha\,\mathbf{1}(\Theta_0 = +1)\right] \\
&\quad + \mathrm{E}\left[g(\Theta_0/\Theta_1)\,(-(\Theta_1/\Theta_0)^+)^\alpha\,\mathbf{1}(\Theta_0 = -1)\right] \\
&= p\,\mathrm{E}\left[g(1/M)\,(M^+)^\alpha)\right] + (1-p)\,\mathrm{E}\left[g(1/M)\,(M^-)^\alpha\right], \\
\mathrm{E}\left[g(\Theta_{-1}/\Theta_0)\,\mathbf{1}(\Theta_0 = -1)\right] &= p\,\mathrm{E}\left[g(1/M)\,(M^-)^\alpha)\right] + (1-p)\,\mathrm{E}\left[g(1/M)\,(M^+)^\alpha\right].
\end{aligned}$$

Since either $p = 1/2$ or $\mathrm{P}[M < 0] = 0$, equation (8.4) follows. □

Proof of Lemma 4.1. To show (4.4), apply the time-change formula (2.7) with $s = -1$, $t = 0$, $i = -1$, and $f(y_{-1}, y_0) = \mathbf{1}(y_0/y_{-1} > x,\ y_{-1} = 1)$, where $x \geq 0$. It follows that

$$\begin{aligned}
\mathrm{P}[A_1 > x] &= \mathrm{E}\left[\mathbf{1}\left(\frac{\Theta_1}{\Theta_0} > x\right)\mathbf{1}(\Theta_0 = 1)\right] / \mathrm{P}[\Theta_0 = 1] \\
&= \mathrm{E}\left[\mathbf{1}\left(\frac{\Theta_0}{\Theta_{-1}} > x\right)\mathbf{1}(\Theta_{-1} > 0)\,|\Theta_{-1}|^\alpha\right] / \mathrm{P}[\Theta_0 = 1] \\
&= \mathrm{E}\left[A_{-1}^\alpha\,\mathbf{1}(1/A_{-1} > x)\right].
\end{aligned}$$

To show (4.5), apply the time-change formula to the function $f(y_{-1}, y_0) = \mathbf{1}(y_0/y_{-1} \leq x,\ y_{-1} = 1)$. The proofs of (4.6) and (4.7) are similar. □

Proof of Theorem 5.1. We argue similarly as in the proof of Corollary 3.6 and Remark 3.7 of Drees and Rootzén (2010): we first establish weak convergence of all finite dimensional distributions using Theorem 2.3 of that paper, and then the asymptotic equicontinuity of the processes $(\tilde{Z}_n(\phi_{2,x}))_{x \geq x_0}$ and $(\tilde{Z}_n(\phi_{3,y}))_{y \geq y_0}$ by applying Theorem 2.10, from which the assertion follows.

First we verify that conditions (C1)–(C3) of Drees and Rootzén (2010) are fulfilled so that Theorem 2.3 on the convergence of finite-dimensional distributions applies. As in the proof of Corollary 3.6, (C1) can be derived from condition (3.5) of that paper, which in turn is an easy consequence of condition (C), because by the stationarity of $(X_t)_{t \in \mathbb{Z}}$, we have

$$\begin{aligned}
\mathrm{E}\left[\left(\sum_{i=1}^{r_n}\mathbf{1}(X_{n,i} \neq 0)\right)^2\right] &= \sum_{i=1}^{r_n}\sum_{j=1}^{r_n}\mathrm{P}[X_i > u_n, X_j > u_n] \\
&\leq 2 r_n v_n \sum_{k=0}^{r_n - 1}\left(1 - \frac{k}{r_n}\right)\mathrm{P}[X_k > u_n \mid X_0 > u_n] \\
&\leq 2 r_n v_n \sum_{k=0}^{r_n - 1} s_n(k) = O(r_n v_n), \qquad n \to \infty.
\end{aligned}$$

Since all functions ϕ_1, $\phi_{2,x}$ and $\phi_{3,y}$ are bounded and since $r_n = o((n v_n)^{1/2})$, condition (C2) is obviously fulfilled.

For the convergence of all finite-dimensional marginal distributions it remains to establish condition (C3) of Drees and Rootzén (2010), i.e.,

$$\frac{1}{r_n v_n}\operatorname{cov}\left(\sum_{i=1}^{r_n}\psi_1(X_{n,i}), \sum_{j=1}^{r_n}\psi_2(X_{n,j})\right) \to \operatorname{cov}\left(\tilde{Z}(\psi_1), \tilde{Z}(\psi_2)\right), \qquad n \to \infty, \tag{8.5}$$

for all $\psi_1, \psi_2 \in \{\phi_1, \phi_{2,x}, \phi_{3,y} : x \geq x_0, y \geq y_0\}$. Similarly as above, by the stationarity of $(X_t)_{t\in\mathbb{Z}}$, the left-hand side equals

$$
\begin{aligned}
&\frac{1}{r_n v_n} \mathrm{E}\left[\sum_{i=1}^{r_n}\sum_{j=1}^{r_n} \psi_1(X_{n,i})\, \psi_2(X_{n,j})\right] + O(r_n v_n) \\
&= \mathrm{E}\left[\psi_1(X_{n,0})\, \psi_2(X_{n,0}) \mid X_0 > u_n\right] \\
&\quad + \sum_{k=1}^{r_n-1} \Big(1 - \frac{k}{r_n}\Big) \Big(\mathrm{E}\left[\psi_1(X_{n,0})\, \psi_2(X_{n,k}) \mid X_0 > u_n\right] + \mathrm{E}\left[\psi_1(X_{n,k})\, \psi_2(X_{n,0}) \mid X_0 > u_n\right]\Big).
\end{aligned}
$$

By assumption (A(x_0)), all functions under consideration are a.s. continuous and bounded. The definition of the tail process then yields

$$
\mathrm{E}\left[\psi_1(X_{n,0})\, \psi_2(X_{n,k}) \mid X_0 > u_n\right] \to \mathrm{E}\left[\psi_1(Y_{-1}, Y_0, Y_1)\, \psi_2(Y_{k-1}, Y_k, Y_{k+1})\right], \qquad n \to \infty,
$$

for all $k \geq 0$ and all $\psi_1, \psi_2 \in \{\phi_1, \phi_{2,x}, \phi_{3,y} : x \geq x_0, y \geq y_0\}$. Thus (8.5) follows by Pratt's lemma (Pratt, 1960) and condition (C), since

$$
\begin{aligned}
\Big(1 - \frac{k}{r_n}\Big) \mathrm{E}\left[\psi_1(X_{n,0})\, \psi_2(X_{n,k}) \mid X_0 > u_n\right] \quad &\leq \quad \max(1, \tilde{x}_0^{-2\alpha})\, \mathrm{P}\left[X_k > u_n \mid X_0 > u_n\right] \\
&\leq \quad \max(1, \tilde{x}_0^{-2\alpha})\, s_n(k).
\end{aligned}
$$

In the second step, the asymptotic equicontinuity of $(\tilde{Z}_n(\phi_{2,x}))_{x\in[x_0,\infty)}$ and $(\tilde{Z}_n(\phi_{3,y}))_{y\in[y_0,\infty)}$ follows from Theorem 2.10 of Drees and Rootzén (2010) if their conditions (D1), (D2'), (D3), (D5) and (D6) are verified. Note that (D1) is obvious, that (D5) is an immediate consequence of the separability of the processes, and that (D2') follows from $r_n = o((nv_n)^{1/2})$ and the boundedness of all functionals $\phi_{2,x}$ and $\phi_{3,y}$. Moreover, because the maps $x \mapsto \phi_{2,x}(y_{-1}, y_0, y_1)$ and $x \mapsto \phi_{3,x}(y_{-1}, y_0, y_1)$ are decreasing, condition (D6) can be concluded in the same way as in the case $d = 1$ of Example 3.8 in Drees and Rootzén (2010).

It remains to establish the continuity condition (D3) for the semi-norm generated by the cdf $F^{(A_1)}$, i.e.,

$$
\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \sup_{\substack{y>x\geq x_0 \\ F^{(A_1)}(y)-F^{(A_1)}(x)\leq\delta}} \frac{1}{r_n v_n} \mathrm{E}\left[\left(\sum_{i=1}^{r_n} \big(\phi_{2,y}(X_{n,i}) - \phi_{2,x}(X_{n,i})\big)\right)^2\right] = 0 \qquad (8.6)
$$

and an analogous condition for $\phi_{3,y}$. By the usual stationarity argument, the expectation on the left-hand side can be bounded by a multiple of

$$
r_n v_n \sum_{k=0}^{r_n-1} \mathrm{P}\left[X_1/X_0 \in (x,y],\, X_{k+1}/X_k \in (x,y],\, X_k > u_n \mid X_0 > u_n\right].
$$

Now, for all fixed $M > 0$, as $n \to \infty$,

$$
\begin{aligned}
\sum_{k=0}^{M} \mathrm{P}\left[\frac{X_1}{X_0} \in (x,y], \frac{X_{k+1}}{X_k} \in (x,y], X_k > u_n \,\Big|\, X_0 > u_n\right] &\leq (M+1)\, \mathrm{P}\left[\frac{X_1}{X_0} \in (x,y] \,\Big|\, X_0 > u_n\right] \\
&\to (M+1)\big(F^{(A_1)}(y) - F^{(A_1)}(x)\big) \leq (M+1)\delta, \quad (8.7)
\end{aligned}
$$

uniformly for all $y > x \geq x_0$ by condition (A(x_0)). Moreover, condition (C) implies

$$
\begin{aligned}
&\sum_{k=M+1}^{r_n} \mathrm{P}\left[\frac{X_1}{X_0} \in (x,y], \frac{X_{k+1}}{X_k} \in (x,y], X_k > u_n \,\Big|\, X_0 > u_n\right] \\
&\qquad \leq \sum_{k=M+1}^{r_n} s_n(k) \to \sum_{k=M+1}^{\infty} \lim_{n\to\infty} s_n(k) < \infty. \quad (8.8)
\end{aligned}
$$

By choosing M sufficiently large, the right-hand side of (8.8) can be made arbitrarily small. Given such M, by choosing δ small, the right-hand side of (8.7) can be made arbitrarily small too. Equation (8.6) follows. The convergence statement (8.6) involving the functions $(\phi_{3,y})_{y\geq y_0}$ can be concluded in a similar way. □

Proof of Corollary 5.2. For notational simplicity, we write $\bar{F}$ for $\bar{F}^{(A_1)} = 1 - F^{(A_1)}$ and use the notations $\hat{\bar{F}}_n^{(\mathrm{f},A_1)}(x) := 1 - \hat{F}_n^{(\mathrm{f},A_1)}(x)$ and $\hat{\bar{F}}_n^{(\mathrm{b},A_1)}(y) := 1 - \hat{F}_n^{(\mathrm{b},A_1)}(y)$ for the estimators of the survival function. Check that

$$\begin{aligned}
&(nv_n)^{1/2}\left(\hat{\bar{F}}_n^{(\mathrm{f},A_1)}(x) - \mathrm{P}\left[X_1/X_0 > x \mid X_0 > u_n\right]\right) \\
&= (nv_n)^{1/2}\left(\frac{\sum_{i=1}^n \phi_{2,x}(X_{n,i})}{\sum_{i=1}^n \phi_1(X_{n,i})} - \frac{\mathrm{P}\left[X_1/X_0 > x, X_0 > u_n\right]}{v_n}\right) \\
&= (nv_n)^{1/2}\left(\frac{n\,\mathrm{E}\left[\phi_{2,x}(X_{n,0})\right] + (nv_n)^{1/2}\tilde{Z}_n(\phi_{2,x})}{nv_n + (nv_n)^{1/2}\tilde{Z}_n(\phi_1)} - \frac{n\,\mathrm{E}\left[\phi_{2,x}(X_{n,0})\right]}{nv_n}\right) \\
&= \frac{\tilde{Z}_n(\phi_{2,x}) - \mathrm{P}\left[X_1/X_0 > x \mid X_0 > u_n\right]\tilde{Z}_n(\phi_1)}{1 + (nv_n)^{-1/2}\tilde{Z}_n(\phi_1)},
\end{aligned}$$

and likewise

$$\begin{aligned}
&(nv_n)^{1/2}\left(\hat{\bar{F}}_n^{(\mathrm{b},A_1)}(y) - \mathrm{E}\left[(X_{-1}/X_0)^\alpha\,\mathbf{1}(X_0/X_{-1} > y) \mid X_0 > u_n\right]\right) \\
&\qquad = \frac{\tilde{Z}_n(\phi_{3,y}) - \mathrm{E}\left[(X_{-1}/X_0)^\alpha\,\mathbf{1}(X_0/X_{-1} > y) \mid X_0 > u_n\right]\tilde{Z}_n(\phi_1)}{1 + (nv_n)^{-1/2}\tilde{Z}_n(\phi_1)}.
\end{aligned}$$

Hence, in view of (5.2) and (5.3), Theorem 5.1 implies (5.4) with $Z^{(f,A_1)}(x) = \bar{F}(x)\tilde{Z}(\phi_1) - \tilde{Z}(\phi_{2,x})$ and $Z^{(b,A_1)}(y) = \bar{F}(y)\tilde{Z}(\phi_1) - \tilde{Z}(\phi_{3,y})$.

The covariance structure of the limiting process follows by direct calculations. Nonnegativity of X_0 implies $Y_0 > 1$ a.s. Moreover, as $(\Theta_t)_{t\in\mathbb{Z}}$ is a Markov spectral tail chain, the random variables $Y_k/Y_{k-1} = \Theta_k/\Theta_{k-1}$, $k \in \mathbb{Z}$, are independent; for $k \in \mathbb{N}$, their common survival function is $\bar{F}$. From Theorem 5.1, we obtain

$$\begin{aligned}
\operatorname{var}(\tilde{Z}(\phi_1)) &= 1 + 2\sum_{k=1}^{\infty} \mathrm{P}\left[Y_k > 1\right] \\
\operatorname{cov}\left(\tilde{Z}(\phi_1), \tilde{Z}(\phi_{2,x})\right) &= \mathrm{P}\left[\Theta_1 > x\right] + \sum_{k=1}^{\infty}\big(\mathrm{P}\left[\Theta_{k+1}/\Theta_k > x, Y_k > 1\right] + \mathrm{P}\left[\Theta_1 > x, Y_k > 1\right]\big) \\
&= \bar{F}(x)\sum_{k=0}^{\infty}\mathrm{P}\left[Y_k > 1\right] + \sum_{k=1}^{\infty}\mathrm{P}\left[\Theta_1 > x, Y_k > 1\right], \\
\operatorname{cov}\left(\tilde{Z}(\phi_{2,x}), \tilde{Z}(\phi_{2,y})\right) &= \mathrm{P}\left[\Theta_1 > \max(x,y)\right] + \sum_{k=1}^{\infty}\big(\mathrm{P}\left[\Theta_1 > x, \Theta_{k+1}/\Theta_k > y, Y_k > 1\right] \\
&\qquad + \mathrm{P}\left[\Theta_1 > y, \Theta_{k+1}/\Theta_k > x, Y_k > 1\right]\big) \\
&= \bar{F}(\max(x,y)) + \sum_{k=1}^{\infty}\big(\mathrm{P}\left[\Theta_1 > x, Y_k > 1\right]\bar{F}(y) \\
&\qquad + \mathrm{P}\left[\Theta_1 > y, Y_k > 1\right]\bar{F}(x)\big).
\end{aligned}$$

Similarly,

$$\operatorname{cov}\left(\tilde{Z}(\phi_1), \tilde{Z}(\phi_{3,x})\right) = \bar{F}(x) \sum_{k=0}^{\infty} \mathrm{P}\left[Y_k > 1\right] + \sum_{k=1}^{\infty} \mathrm{E}\left[(\Theta_{k-1}/\Theta_k)^{\alpha} \mathbf{1}(\Theta_k/\Theta_{k-1} > x, Y_k > 1)\right]$$

$$\operatorname{cov}\left(\tilde{Z}(\phi_{3,x}), \tilde{Z}(\phi_{3,y})\right) = \bar{F}(\max(x,y)) + \sum_{k=1}^{\infty} \Big(\bar{F}(x)\, \mathrm{E}\left[(\Theta_{k-1}/\Theta_k)^{\alpha} \mathbf{1}(\Theta_k/\Theta_{k-1} > y, Y_k > 1)\right]$$
$$+ \bar{F}(y)\, \mathrm{E}\left[(\Theta_{k-1}/\Theta_k)^{\alpha} \mathbf{1}(\Theta_k/\Theta_{k-1} > x, Y_k > 1)\right] \Big)$$

$$\operatorname{cov}\left(\tilde{Z}(\phi_{2,x}), \tilde{Z}(\phi_{3,y})\right) = \bar{F}(x)\bar{F}(y) \sum_{k=0}^{\infty} \mathrm{P}\left[Y_k > 1\right]$$
$$+ \sum_{k=1}^{\infty} \mathrm{E}\left[(\Theta_{k-1}/\Theta_k)^{\alpha} \mathbf{1}(\Theta_1 > x, \Theta_k/\Theta_{k-1} > y, Y_k > 1)\right].$$

The asymptotic covariance functions of the forward and the backward estimators can then be derived as follows:

$$\begin{aligned}
&\operatorname{cov}\left(Z^{(f,A_1)}(x), Z^{(b,A_1)}(y)\right) \\
&= \operatorname{cov}\left(\tilde{Z}(\phi_{2,x}), \tilde{Z}(\phi_{3,y})\right) - \bar{F}(x) \operatorname{cov}\left(\tilde{Z}(\phi_1), \tilde{Z}(\phi_{3,y})\right) - \bar{F}(y) \operatorname{cov}\left(\tilde{Z}(\phi_1), \tilde{Z}(\phi_{2,x})\right) \\
&\quad + \bar{F}(x)\, \bar{F}(y) \operatorname{var}(\tilde{Z}(\phi_1)) \\
&= \bar{F}(x)\, \bar{F}(y) \sum_{k=1}^{\infty} \mathrm{P}\left[Y_k > 1\right] - \bar{F}(x) \sum_{k=1}^{\infty} \mathrm{E}\left[(\Theta_{k-1}/\Theta_k)^{\alpha}\, \mathbf{1}(\Theta_k/\Theta_{k-1} > y, Y_k > 1)\right] \\
&\quad - \bar{F}(y) \sum_{k=1}^{\infty} \mathrm{P}\left[\Theta_1 > x, Y_k > 1\right] + \sum_{k=1}^{\infty} \mathrm{E}\left[(\Theta_{k-1}/\Theta_k)^{\alpha}\, \mathbf{1}(\Theta_1 > x, \Theta_k/\Theta_{k-1} > y, Y_k > 1)\right].
\end{aligned}$$

The other covariances can be calculated in a similar way. □

Proof of Lemma 5.5. By similar arguments as used in the proof of Theorem 5.1, one can show that under the present conditions the conclusion of Theorem 5.1 remain valid if the family of functions is extended to $\{\phi_1, \phi_{2,x}, \phi_{3,y}, \psi : x \geq x_0, y \geq y_0\}$. Hence the assertion follows from

$$\begin{aligned}
\hat{\alpha}_n - \alpha_n &= \frac{nv_n + (nv_n)^{1/2} \tilde{Z}_n(\phi_1)}{n\, \mathrm{E}[\tilde{\psi}(X_0/u_n)\, \mathbf{1}(X_0 > u_n)] + (nv_n)^{1/2} \tilde{Z}_n(\psi) + R_n} - \alpha_n \\
&= \alpha_n \frac{1 + (nv_n)^{-1/2} \tilde{Z}_n(\phi_1)}{1 + \alpha_n((nv_n)^{-1/2} \tilde{Z}_n(\psi) + (nv_n)^{-1} R_n)} - \alpha_n \\
&= \alpha_n (nv_n)^{-1/2} \frac{\tilde{Z}_n(\phi_1) - \alpha_n \left\{Z_n(\psi) + (nv_n)^{-1/2} R_n\right\}}{1 + \alpha_n \left\{(nv_n)^{-1/2} \tilde{Z}_n(\psi) + (nv_n)^{-1} R_n\right\}} \\
&= (nv_n)^{-1/2} \big(\alpha \tilde{Z}_n(\phi_1) - \alpha^2 Z_n(\psi) + o_P(1)\big)\big(1 + o_P(1)\big), \qquad n \to \infty.
\end{aligned}$$

In the last step we have used stochastic boundedness of $\tilde{Z}_n(\psi)$ and $\tilde{Z}_n(\phi_1)$, which follows from their weak convergence, and the assumption that $R_n = o_P((nv_n)^{1/2})$ as $n \to \infty$. □

Proof of Corollary 5.6. A Taylor expansion of the function $t \mapsto z^t$ yields

$$z^{\hat{\alpha}_n} - z^{\alpha} = z^{\alpha} \log(z)\, (\hat{\alpha}_n - \alpha) + \frac{1}{2} z^{\alpha + \lambda(\hat{\alpha}_n - \alpha)} (\log z)^2 (\hat{\alpha}_n - \alpha)^2$$

for some (random) $\lambda = \lambda_{z,\alpha} \in (0,1)$. Because $z^{\tilde{\alpha}}(\log z)^2$ is bounded for all $\tilde{\alpha}$ in a neighborhood of α and all $z \leq 1/y_0$, it follows that on the event $\{X_i/X_{i-1} \geq y_0\}$, we have

$$\left| \Big(\frac{X_{i-1}}{X_i}\Big)^{\hat{\alpha}_n} - \Big(\frac{X_{i-1}}{X_i}\Big)^{\alpha} - (nv_n)^{-1/2} \Big(\frac{X_{i-1}}{X_i}\Big)^{\alpha} \log\Big(\frac{X_{i-1}}{X_i}\Big) \big(\alpha \tilde{Z}_n(\phi_1) - \alpha^2 \tilde{Z}_n(\psi)\big) \right| \leq C(\hat{\alpha}_n - \alpha)^2$$

for some constant C depending only on α and y_0 (but not on i or n). Hence, by Lemma 5.5, as $n \to \infty$,

$$
\begin{aligned}
&(nv_n)^{-1/2} \sum_{i=1}^n \left(\left(\frac{X_{i-1}}{X_i}\right)^{\hat\alpha_n} - \left(\frac{X_{i-1}}{X_i}\right)^{\alpha} \right) \mathbf{1}\Big(\frac{X_{i-1}}{X_i} > y, X_i > u_n\Big) \\
&= \frac{\alpha \tilde Z_n(\phi_1) - \alpha^2 \tilde Z_n(\psi)}{nv_n} \sum_{i=1}^n \left(\frac{X_{i-1}}{X_i}\right)^{\alpha} \log\left(\frac{X_{i-1}}{X_i}\right) \mathbf{1}\Big(\frac{X_i}{X_{i-1}} > y, X_i > u_n\Big) \\
&\qquad + o_P\Big((nv_n)^{-1} \sum_{i=1}^n \mathbf{1}(X_i > u_n)\Big) \\
&= \frac{\alpha \tilde Z_n(\phi_1) - \alpha^2 \tilde Z_n(\psi)}{nv_n} \sum_{i=1}^n \left(\frac{X_{i-1}}{X_i}\right)^{\alpha} \log\left(\frac{X_{i-1}}{X_i}\right) \mathbf{1}\Big(\frac{X_i}{X_{i-1}} > y, X_i > u_n\Big) + o_P(1).
\end{aligned}
$$

The last equality follows from the weak convergence of $\tilde Z_n(\phi_1)$.

As in the proof of Theorem 5.1, one may establish weak convergence of

$$
\Bigg((nv_n)^{-1/2} \sum_{i=1}^n \left(\frac{X_{i-1}}{X_i}\right)^{\alpha} \log\left(\frac{X_{i-1}}{X_i}\right) \mathbf{1}\Big(\frac{X_{i-1}}{X_i} > y, X_i > u_n\Big) - \mathrm{E}\left[\left(\frac{X_{i-1}}{X_i}\right)^{\alpha} \log\left(\frac{X_{i-1}}{X_i}\right) \mathbf{1}\Big(\frac{X_{i-1}}{X_i} > y, X_i > u_n\Big) \right] \Bigg)_{y \ge y_0}
$$

to a centered Gaussian process. In particular, as $n \to \infty$,

$$
\begin{aligned}
&(nv_n)^{-1} \sum_{i=1}^n \left(\frac{X_{i-1}}{X_i}\right)^{\alpha} \log\left(\frac{X_{i-1}}{X_i}\right) \mathbf{1}\Big(\frac{X_{i-1}}{X_i} > y, X_i > u_n\Big) \\
&= \mathrm{E}\left[\left(\frac{X_{-1}}{X_0}\right)^{\alpha} \log\left(\frac{X_{-1}}{X_0}\right) \mathbf{1}\Big(\frac{X_0}{X_{-1}} > y\Big) \,\Big|\, X_0 > u_n \right] + o_P(1) \\
&\rightsquigarrow \mathrm{E}\left[\Theta_{-1}^{\alpha} \log(\Theta_{-1})\, \mathbf{1}(1/\Theta_{-1} > y)\right] =: \ell_y,
\end{aligned}
$$

uniformly for $y \ge y_0$. By the time-change formula (2.7) with $i = -1$, $s = t = 0$ and $f(x) = -\log(x)\, \mathbf{1}_{(y,\infty)}(x)$, one has $\ell_y = -\mathrm{E}\left[\log(\Theta_1)\, \mathbf{1}(\Theta_1 > y)\right]$.

It follows that

$$
\begin{aligned}
&(nv_n)^{-1/2} \sum_{i=1}^n \left(\left(\frac{X_{i-1}}{X_i}\right)^{\hat\alpha_n} \mathbf{1}\Big(\frac{X_i}{X_{i-1}} > y, X_i > u_n\Big) - \mathrm{E}\left[\left(\frac{X_{i-1}}{X_i}\right)^{\alpha} \mathbf{1}\Big(\frac{X_i}{X_{i-1}} > y, X_i > u_n\Big) \right] \right) \\
&\qquad = \tilde Z_n(\phi_{3,y}) + \ell_y \left(\alpha Z_n(\phi_1) - \alpha^2 \tilde Z_n(\tilde\psi)\right) + o_P(1), \qquad n \to \infty.
\end{aligned}
$$

Proceed as in the proof of Theorem 5.1 to arrive at the assertion. □

Acknowledgments

H. Drees was supported by the "Deutsche Forschungsgemeinschaft", projects FOR 1735 and JA2160/1. J. Segers was supported by contract "Projet d'Actions de Recherche Concertées" No. 12/17-045 of the "Communauté française de Belgique" and by IAP research network Grant P7/06 of the Belgian government (Belgian Science Policy). M. Warchoł was funded by a "mandat d'aspirant" of the "Fonds de la Recherche Scientifique" (FNRS).

References

Basrak, B., R. A. Davis, and T. Mikosch (2002). Regular variation of GARCH processes. *Stochastic Processes and their Applications 99*(1), 95 – 115.

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

Beare, B. K. and J. Seo (2014, 4). Time irreversible copula-based markov models. *Econometric Theory FirstView*, 1–38.

Bortot, P. and S. Coles (2000). A sufficiency property arising from the characterization of extremes of markov chains. *Bernoulli 6*(1), 183–190.

Bowman, R. G. and D. Iverson (1998). Short-run overreaction in the New Zealand stock market. *Pacific-Basin Finance Journal 6*(5), 475–491.

Chen, X. and Y. Fan (2006). Estimation of copula-based semiparametric time series models. *Journal of Econometrics 130*(2), 307–335.

Chen, X., W. B. Wu, and Y. Yi (2009). Efficient estimation of copula-based semiparametric markov models. *The Annals of Statistics 37*(6B), 4214–4253.

Chen, Y.-T., R. Y. Chou, and C.-M. Kuan (2000). Testing time reversibility without moment restrictions. *Journal of Econometrics 95*(1), 199–218.

Chen, Y.-T. and C.-M. Kuan (2002). Time irreversibility and EGARCH effects in US stock index returns. *Journal of Applied Econometrics 17*(5), 565–578.

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

Davis, R. A., T. Mikosch, and Y. Zhao (2013). Measures of serial extremal dependence and their estimation. *Stochastic Processes and their Applications 123*(7), 2575–2602.

De Bondt, W. F. M. and R. Thaler (1985). Does the stock market overreact? *The Journal of Finance 40*(3), 793–805.

Dekkers, A., J. Einmahl, and L. de Haan (1989). A moment estimator for the index of an extreme-value distribution. *Annals of Statistics 17*(4), 1833–1855.

Demarta, S. and A. J. McNeil (2005). The t copula and related copulas. *International Statistical Review 73*(1), 111–129.

Doukhan, P. (1995). *Mixing. Properties and Examples.* New York: Springer.

Drees, H. (1998a). A general class of estimators of the extreme value index. *Journal of Statistical Planning and Inference 66*(1), 95–112.

Drees, H. (1998b). On smooth statistical tail functionals. *Scandinavian Journal of Statistics 25*(1), 187–210.

Drees, H. and H. Rootzén (2010). Limit theorems for empirical processes of cluster functionals. *The Annals of Statistics 38*(4), 2145–2186.

Goldie, C. M. (1991). Implicit renewal theory and tails of solutions of random equations. *The Annals of Applied Probability 1*(1), 126–166.

Gudendorf, G. and J. Segers (2010). Extreme-value copulas. In *Copula Theory and Its Applications*, Lecture Notes in Statistics. Springer Berlin Heidelberg.

Janßen, A. and H. Drees (2013). A stochastic volatility model with flexible extremal dependence structure. Available at http://arxiv.org/abs/1310.4621.

Janßen, A. and J. Segers (2014). Markov tail chains. *Advances in Applied Probability, forthcoming*.

Joe, H. (1990). Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statistics & Probability Letters 9*(1), 75–81.

Kesten, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Mathematica 131* (1), 207–248.

Kulik, R. and P. Soulier (2013). Heavy tailed time series with extremal independence. Available at http://arxiv.org/abs/1307.1501.

Larsson, M. and S. I. Resnick (2012). Extremal dependence measure and extremogram: the regularly varying case. *Extremes 15* (2), 231–256.

Leadbetter, M. (1983). Extremes and local dependence in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 65* (2), 291–306.

Ledford, A. and J. Tawn (1996). Statistics for near independence in multivariate extreme values. *Biometrika 83* (1), 169–187.

Ledford, A. and J. Tawn (2003). Diagnostics for dependence within time series extremes. *Journal of the Royal Statistical society, Series B 65* (2), 521–543.

Norli, A. (2010). Short run stock overreaction: Evidence from Bursa Malaysia. *Journal of Economics and management 4* (2), 319–333.

Perfekt, R. (1997). Extreme value theory for a class of Markov chains with values in $\mathbb{R}^d$. *Advances in Applied Probability 29* (1), 138–164.

Pratt, J. W. (1960). On interchanging limits and integrals. *Annals of Mathematical Statistics 31* (1), 74–77.

Segers, J. (2007). Multivariate regular variation of heavy-tailed markov chains. Technical report, Université catholique de Louvain. Available at http://arxiv.org/abs/math/0701411.

Smith, R. (1987). Estimating tails of probability distributions. *Annals of Statistics 15* (3), 1174–1207.

Smith, R. L. (1992). The extremal index for a Markov chain. *Journal of Applied Probability 29* (1), 37–45.

Tawn, J. A. (1988). Bivariate extreme value theory: Models and estimation. *Biometrika 75* (3), 397–415.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence of Empirical Processes.* New York: Springer.

Yun, S. (2000). The distributions of cluster functionals of extreme events in a dth-order markov chain. *Journal of Applied Probability 37* (1), 29–44.