HIDDEN TAIL CHAINS AND RECURRENCE EQUATIONS FOR DEPENDENCE PARAMETERS ASSOCIATED WITH EXTREMES OF STATIONARY HIGHER-ORDER MARKOV CHAINS

IOANNIS PAPASTATHOPOULOS 🆔,* ** AND
ADRIAN CASEY,* *University of Edinburgh*
JONATHAN A. TAWN,*** *Lancaster University*

Abstract

We derive some key extremal features for stationary kth-order Markov chains that can be used to understand how the process moves between an extreme state and the body of the process. The chains are studied given that there is an exceedance of a threshold, as the threshold tends to the upper endpoint of the distribution. Unlike previous studies with $k > 1$, we consider processes where standard limit theory describes each extreme event as a single observation without any information about the transition to and from the body of the distribution. Our work uses different asymptotic theory which results in non-degenerate limit laws for such processes. We study the extremal properties of the initial distribution and the transition probability kernel of the Markov chain under weak assumptions for broad classes of extremal dependence structures that cover both asymptotically dependent and asymptotically independent Markov chains. For chains with $k > 1$, the transition of the chain away from the exceedance involves novel functions of the k previous states, in comparison to just the single value, when $k = 1$. This leads to an increase in the complexity of determining the form of this class of functions, their properties, and the method of their derivation in applications. We find that it is possible to derive an affine normalization, dependent on the threshold excess, such that non-degenerate limiting behaviour of the process, in the neighbourhood of the threshold excess, is assured for all lags. We find that these normalization functions have an attractive structure that has parallels to the Yule–Walker equations. Furthermore, the limiting process is always linear in the innovations. We illustrate the results with the study of kth-order stationary Markov chains with exponential margins based on widely studied families of copula dependence structures.

Keywords: conditional extremes; extremal index; homogeneous and stationary Markov chain; recurrence equation; tail chain

2010 Mathematics Subject Classification: Primary 60GXX
Secondary 60G70

1. Introduction

The extreme value theory of sequences of independent and identically distributed (i.i.d.) random variables has often been generalized to include the situation where the variables are no longer independent, as in the monograph of [15], where for stationary processes the focus is on long-range dependence conditions and local clustering of extremes as measured by the extremal index. Among the most useful stochastic processes are positive recurrent Markov chains, with a continuous state space, which provide the backbone of a broad range of statistical models for stationary time series. Such models have attracted considerable interest in the analysis of extremes of stochastic processes. These processes meet the required long-range dependence conditions [19, 27], so it is the local clustering properties of their extreme values that are of interest. By considering the behaviour of the process when it is extreme, that is, when it exceeds a high threshold, [27] showed that, under certain circumstances, the times of extreme events of stationary Markov chains that exceed a high threshold converge to a homogeneous Poisson process and that the limiting characteristics of the values within an extreme event, including the extremal index θ with $\theta \in (0, 1]$ being the reciprocal of the mean extreme event duration, can be derived as the threshold converges to the upper endpoint of the marginal distribution. This limit result only reveals the behaviour of the process whilst it remains at the same level of marginal extremity as the threshold, and therefore it is only informative about the temporal structure of extreme events for a subset of processes, i.e. those with $\theta < 1$. This excludes all processes with $\theta = 1$, where extreme values occur in temporal isolation, with no apparent clustering of extreme values revealed through this limit theory. Processes with $\theta = 1$ include all Gaussian processes, so they can exhibit strong temporal dependence. The behaviour of extreme events of these processes is of interest, but cannot be studied through the existing methods. So we are interested in deriving the detailed characteristics of a positive recurrent Markov chain within an extremal event, irrespective of whether $\theta < 1$ or $\theta = 1$, to provide more insights than are currently available.

We focus on real-valued stationary kth-order Markov processes $\{X_t : t \in \mathbb{Z}\}$, for $k \in \mathbb{N}$, with the marginal distribution function F of X_t and with a copula [12] for the joint distributions of $(X_{t-k}, \ldots, X_t)$, for all $t \in \mathbb{Z}$, which is invariant to time, so that the one-step forward transition probability kernel π is invariant to t, as it is a function of F and this copula. Motivated by the limitations of the established limit theory for stationary Markov chains, we seek to understand better the behaviour of such processes within an extreme event under less restrictive conditions, by using a more refined limit theory. Specifically, our analysis aims to characterize the temporal behaviour of extreme events transitioning between an extreme state and the body of the distribution irrespective of whether the process has $\theta < 1$ or $\theta = 1$.

The case where $k = 1$ has been well studied. Under weak conditions, [27, 31] show that the powerful general Poisson limit gives that the extremal index is either $\theta < 1$ or $\theta = 1$, respectively, depending on whether $\chi_1 > 0$ or $\chi_1 = 0$, where

$$\chi_1 = \lim_{u \to 1} \Pr\{F(X_t) > u \mid F(X_{t-1}) > u\},$$

for all $t \in \mathbb{Z}$. These two limiting properties are known as asymptotic dependence and asymptotic independence of (X_{t-1}, X_t), respectively, in the literature on bivariate extremes. To derive greater detail about the behaviour within extreme events for asymptotically dependent Markov chains, the appropriate strategy is to study *tail chains* [11]. A tail chain arises as a limiting process after witnessing an extreme state, under rescaling of the future Markov chain by the

value of the process in the extreme state, resulting in the tail chain being driven by a random walk. Tail chains fail to reveal the detailed structure of extreme events for any asymptotically independent process. For $k = 1$, [20] take a different approach, using a new limiting theory involving *hidden tail chains*, defined below by the limit (2), which treat both asymptotically dependent and asymptotically independent chains in a unified theory. They establish the extremal event properties that we require more generally for when $k > 1$. So the focus of our paper is similar to that of [20], but with the added difficulty that the complexity of extreme events substantially increases with k.

Markov processes with order $k > 1$ extend the traditional concept of first-order Markov processes by incorporating information from the $k \in \mathbb{N}$ most recent states into the transition behaviour of the next state. The fundamental importance of these higher-order Markov processes lies in their ability to capture and model dependencies in time-series data in real-world scenarios, which first-order Markov processes would fail to do. Specifically, the first-order process can account only for the current level of the process and not whether at that time the process has just moved up or down; $k = 2$ is required in such a scenario, with larger k needed as more subtle memory of the past is important to determine the future behaviour. To the best of our knowledge, key characteristics of the extremal behaviour of higher-order Markov chains have not been dealt with in depth, yet these are crucial for understanding the evolution of extreme events of random processes and for providing well-founded parametric models that can be used for inference, prediction, and assessment of risk, e.g. [36, 37].

Modelling stationary kth-order Markov processes requires the modelling of $(k + 1)$-dimensional multivariate distributions of $(X_{t-k}, \ldots, X_t)$ to describe the transition distributions π of $X_t \mid (X_{t-1}, \ldots, X_{t-k})$, so the extremal properties of these processes cannot be characterized as simply as by the measure χ_1. In fact, it is known from multivariate extreme-value theory that the challenge grows exponentially with k [3, 30]. [11, 26] work on tail chains when $k > 1$, but restrict themselves to processes which, in addition to other conditions, require that $\chi_j > 0$ for all $j = 1, \ldots, k$, where

$$\chi_j = \lim_{u \to 1} \mathbb{P}(F(X_t) > u \mid F(X_{t-j}) > u),$$

for all $t \in \mathbb{Z}$ (which is termed *full pairwise asymptotic dependence* across all variables in the transition), and restrict the Markov chain from moving from an extreme state to the body of the process in a single step. Even in these restrictive cases few results exist, e.g. [11, 23, 39]; however, we provide some extensions of these. We also derive the extensive extremal properties for full pairwise asymptotically independent processes, i.e. with $\chi_j = 0$ for $j = 1, \ldots, k - 1$. Finally, we determine the behaviour for extreme events for a class of processes which allow a subset of the $k > 1$ consecutive states to be in the body of the process while the rest are in an extreme state. This class of process falls between previous investigations and our core developments here.

To work with hidden tail chains, we study the effect of different dependence structures for stationary Markov chains with marginal distributions with exponential tails, for which [20] show that more general results can be achieved when $k = 1$ when using affine normalizations than when the marginals have regularly varying tails, which is the assumption typically used in studying tail chains. There is no loss of generality in making such a transform, as through the probability integral transform we can transform from any marginal distribution to any other, e.g. from regularly varying tails to being in the Gumbel max-domain of attraction.

Without loss of generality, we assume that $\sup\{x : F(x) < 1\} = \infty$. For $k > 1$, we assume that there exist $k-1$ norming functions $a_t : \mathbb{R} \to \mathbb{R}$ and $b_t : \mathbb{R} \to \mathbb{R}_+$, for $t = 1, \ldots, k-1$, such that

$$\left\{\frac{X_t - a_t(X_0)}{b_t(X_0)} : t = 1, \ldots, k-1\right\} \,\Big|\, \{X_0 > u\} \xrightarrow{d} \{Z_t : t = 1, \ldots, k-1\} \tag{1}$$

as $u \to \infty$, where $\xrightarrow{d}$ denotes convergence in distribution and $(Z_1, \ldots, Z_{k-1})$ is a random vector that is non-degenerate in each component. Then our aim is to find conditions that guarantee the existence of an infinite sequence of additional functions $a_t : \mathbb{R} \to \mathbb{R}$ and $b_t : \mathbb{R} \to \mathbb{R}_+$ for $t = k, k+1, \ldots$, such that

$$\left\{\frac{X_t - a_t(X_0)}{b_t(X_0)} : t = 1, 2, \ldots\right\} \,\Big|\, \{X_0 > u\} \xrightarrow{d} \{Z_t : t = 1, 2, \ldots\}, \tag{2}$$

where each Z_t is non-degenerate, with the limit process $\{Z_t : t = 1, 2, \ldots\}$ termed the hidden tail chain. The hidden tail chain generalizes the tail chain studied by [11], as with our marginal choice, the tail chains require norming functions to be $a_t(x) = x$ and $b_t(x) = 1$ for all t, and any Z_t can be degenerate at $\{-\infty\}$. In cases where we find that $a_t(x)/x \to \alpha_t < 1$ as $x \to \infty$ for all $t = 1, 2, \ldots, k-1$ (that is, as [9] show, the process has asymptotic pairwise independence for all lags up to $k-1$), the tail chain degenerates as $\{-\infty, -\infty, \ldots\}$ but the hidden tail chain is non-degenerate and stochastic for all components. Furthermore, we find that if the process is asymptotically dependent for all lags $t = 1, \ldots, k-1$, i.e. $a_t(x) \sim x$ and $b_t(x) \sim 1$ as $x \to \infty$ for all $t = 1, 2, \ldots, k-1$, then the hidden tail chain is identical to the tail chain. So the hidden tail chain reveals important structure of the extreme events lost by the tail chain when the tail chain becomes degenerate, but it equals the tail chain otherwise. Thus hidden tail chains have wider use than tail chains.

Our primary targets are, under weak conditions, to find how the first $k-1$ norming functions $a_t(\cdot)$ and $b_t(\cdot)$ in the limit (1) control those in the limit (2), where $t \geq k$, and to identify the transition dynamics of the hidden tail chain along its index and across its state space. For the former, to find the behaviour of the $t \geq k$ norming functions requires a step-change in approach relative to the case when $k = 1$, studied by [20]. In particular, the transitions involve novel functions, a and b, of the k previous values, in comparison to just the single value when $k = 1$. This instantly makes the problem more challenging, as not only do the transition functions have more arguments, but there can be interaction effects from these arguments. Here we develop results for determining the form of this class of functions and present a method of deriving them in applications. We find some parallels between the extremal properties of the norming functions and the Yule–Walker equations, used in standard time-series analysis [35, 38]. We also make the surprising finding that we can always express the hidden tail chain in the form of a non-stationary scaled autoregressive process. Specifically,

$$Z_t = \psi_t^a(\mathbf{Z}_{t-k:t-1}) + \psi_t^b(\mathbf{Z}_{t-k:t-1})\,\varepsilon_t \quad \text{for all } t > k, \tag{3}$$

where $\mathbf{Z}_{t-k:t-1} = (Z_{t-k}, \ldots, Z_{t-1})$, $\psi_t^a : \mathbb{R}^k \to \mathbb{R}$, $\psi_t^b : \mathbb{R}^k \to \mathbb{R}_+$ are continuous update functions which fall in a particular class of functions, and $\{\varepsilon_t : t = 1, 2, \ldots\}$ is a sequence of non-degenerate i.i.d. innovations.

Using the values of a_t, b_t and the properties of Z_t, as $t \to \infty$, we are able to investigate how the Markov chain returns to a non-extreme state following the occurrence of an extreme state.

We focus almost exclusively on forward-in-time hidden tail chains, as in the limit (1), but we also briefly discuss back-and-forth hidden tail chains, expanding on the equivalent feature for tail chains that [11] study. The limit theory developed in this paper is the first that considers asymptotic independence when studying extreme values of any structured process other than a first-order Markov processes. The extension to stationary kth-order Markov processes opens up the possibility of developing similar theory for much broader classes of graphical models. The study of multivariate extreme values on graphical structures has been a rich vein of research recently, with several influential papers, such as [1, 5, 6, 29]. However, all of these papers focus on the case in which all underlying distributions of cliques on the graph are asymptotically dependent. We believe that the results in this paper will help to unlock these approaches to enable the case when some, or all, cliques have asymptotic independence.

Organization of the paper. In Section 2, we state our main theoretical results for higher-order Markov chains with affine update functions under rather broad assumptions. In particular, in that section we relax the requirement for the Markov chain to be stationary, instead assuming only that it is homogeneous. This allows us to show that the new developments here apply more generally. However, from Section 3 onwards we restrict attention to stationary processes, as they allow for the derivation of a much a more refined characterization of extreme events and enable comparisons with past literature. Specifically, in Section 3 we study hidden tail chains of full pairwise asymptotically dependent and full pairwise asymptotically independent stationary Markov chains with standardized marginal distributions. Based on the theory established in Sections 2 and 3 we study a number of special cases and obtain either simpler or closed-form solutions in Sections 4 and 5. Specifically, in Section 4 we characterize closed-form solutions for the norming functions for a class of asymptotically independent Markov chains, with the structure of these functions paralleling that of the autocovariance in Yule–Walker equations. In Section 5, we provide examples of Markov chains constructed from widely studied classes of joint distributions, such as Gaussian, max-stable, and inverted max-stable copulas, and provide a detailed analysis for particular parametric sub-families of these classes. All proofs are postponed to Appendix A.

Some notation. We use the following notation linked to vector and multivariate function operations. Vectors are typeset in bold and vector algebra is interpreted as componentwise throughout the paper. For example, when $\boldsymbol{x} = (x_1, \ldots, x_k)$ and $\boldsymbol{x}' = (x'_1, \ldots, x'_k)$ are two vectors of the same size, $\boldsymbol{x}/\boldsymbol{x}' = (x_1/x'_1, \ldots, x_k/x'_k)$. We also use $d\boldsymbol{x}_{0:t}$ as shorthand for $dx_0 \times \cdots \times dx_t$. The notation $\boldsymbol{x}^\top \boldsymbol{y}$ is reserved for the scalar product of two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^k$, that is, $\boldsymbol{x}^\top \boldsymbol{y} = \sum_{i=1}^k x_i y_i$. For a sequence of measurable functions $\{g_t\}_{t \in \mathbb{N}}$ and real numbers $\{x_t\}_{t \in \mathbb{N}}$, the notation $\boldsymbol{g}_{t-k:t-1}(x)$ and $\boldsymbol{x}_{t-k:t-1}$, for $t, k, t - k \in \mathbb{N}$, is used to denote $(g_{t-k}(x), \ldots, g_{t-1}(x))$ and $(x_{t-k}, \ldots, x_{t-1})$, respectively. By convention, univariable functions on vectors are applied componentwise; e.g. if $f : \mathbb{R} \to \mathbb{R}$, $\boldsymbol{x} \in \mathbb{R}^k$, then $f(\boldsymbol{x}) = (f(x_1), \ldots, f(x_k))$. The symbols $\boldsymbol{0}_p$ and $\boldsymbol{1}_p$, where $p \in \mathbb{N}$, are used to denote the vectors $(0, \ldots, 0) \in \mathbb{R}^p$ and $(1, \ldots, 1) \in \mathbb{R}^p$. We use the notation $\|\boldsymbol{x}\|$ for the L_1 norm of a k-dimensional vector $\boldsymbol{x}$. For a Cartesian coordinate system $\mathbb{R}^k$ with coordinates $x_1, \ldots, x_k$, ∇ is defined by the partial derivative operators as $\nabla = \sum_{i=1}^k (\partial/\partial x_i) \boldsymbol{e}_i$ for an orthonormal basis $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k\}$. The gradient vector of a differentiable function $f : \mathbb{R}^k \to \mathbb{R}$ at $\boldsymbol{x}$ is denoted by $\nabla f(\boldsymbol{x}) = ((\nabla f)_1(\boldsymbol{x}), \ldots, (\nabla f)_k(\boldsymbol{x}))$.

We use the following notation for sets and special functions. The closure of a set A is denoted by $\overline{A}$. The $(k-1)$-dimensional unit simplex $\{\boldsymbol{\omega} \in [0, \infty)^k : \|\boldsymbol{\omega}\| = 1\}$, $k \in \mathbb{N}$, is denoted by Δ^{k-1}. Given cones $\mathcal{K} \subseteq \mathbb{R}_+^k$ and $\mathcal{K}' \subseteq \mathbb{R}_+^k$, a function $f : \mathcal{K} \to \mathcal{K}'$ is called homogeneous of

degree $\rho \in \mathbb{R}$ if $f(\lambda \boldsymbol{x}) = \lambda^\rho f(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathcal{K}$ and $\lambda > 0$. Given cones $\mathcal{K} \subseteq \mathbb{R}_+^k$ and $\mathcal{K}' \subseteq \mathbb{R}_+^k$, a map $f : \mathcal{X} \to \mathcal{X}'$ with $\mathcal{X} \subseteq \mathbb{R}_+^k$ and $\mathcal{X}' \subseteq \mathbb{R}_+^k$ is said to be order-preserving if $\boldsymbol{y} - \boldsymbol{x} \in \mathcal{K}$ implies $f(\boldsymbol{y}) - f(\boldsymbol{x}) \in \mathcal{K}'$. An identity map, denoted by $\mathrm{id} : \mathcal{X} \to \mathcal{X}$, takes every element in a set $\mathcal{X}$ and maps it back to itself.

Notation linked to convergence is defined as follows. For a topological space E we denote its Borel-σ-algebra by $\mathscr{B}(E)$ and the set of bounded continuous functions on E by $C_b(E)$. If f_n, f are real-valued functions on E, we say that f_n converges uniformly on compact sets to f if for any compact $C \subset E$ the convergence $\lim_{n \to \infty} \sup_{x \in C} |f_n(x) - f(x)| = 0$ holds true. Moreover, f_n is said to converge uniformly on compact sets to infinity if $\inf_{x \in C} f_n(x) \to \infty$ for compact sets $C \subset E$ as $n \to \infty$. Weak convergence of measures on E is abbreviated by $\xrightarrow{\mathrm{w}}$. For random elements $X, X_1, X_2, \ldots$ defined on the same probability space, we say $\{X_n\}$ converges in distribution to X, and we write $X_n \xrightarrow{d} X$, if the distributions P_n of the X_n converge weakly to the distribution P of X, that is, if $P_n \xrightarrow{\mathrm{w}} P$.

2. Theory for hidden tail chains of homogeneous Markov chains

2.1. Overview

In this section we present results for $\{X_t : t \in \mathbb{Z}\}$ being a homogeneous Markov chain, with an extreme value at X_0. That is, we assume that we have the distribution function F_0 for X_0, for $k > 1$ with an initial conditional joint distribution function $F_{1:k-1|0}(\boldsymbol{x}_{1:k-1} \mid x_0) = \Pr(X_{1:k-1} \leq \boldsymbol{x}_{1:k-1} \mid X_0 = x_0)$, for $\boldsymbol{x}_{1:k-1} \in \mathbb{R}^{k-1}$, and we have time-invariant forward and backward one-step transition probability kernels π and π^- given by

$$\pi(\boldsymbol{x}_{t-k:t-1}, x_t) := \mathbb{P}(X_t \leq x_t \mid \boldsymbol{X}_{t-k:t-1} = \boldsymbol{x}_{t-k:t-1}) \quad \text{and}$$

$$\pi^-(\boldsymbol{x}_{-(t-k):-(t-1)}, x_{-t}) := \mathbb{P}(X_{-t} \leq x_{-t} \mid \boldsymbol{X}_{-(t-k):-(t-1)} = \boldsymbol{x}_{-(t-k):-(t-1)}) \tag{4}$$

for $t \geq k$ and $t \geq 1$, respectively. So in Section 2 we do not impose structure on the marginal distribution of X_t for $t \neq 0$, or derive how the results relate to the joint distributions of $\boldsymbol{X}_{t-k:t}$ for any t; that is considered in Section 3, where stationarity assumptions are made.

Under weak assumptions, in Section 2 we show how the functions $a_t(x), b_t(x)$ for $t = 1, \ldots, k-1$ and the update functions of (1) and (3), respectively, are connected to transition functionals a and b and the hidden tail chain. From the chosen initial conditional distribution $F_{1:k-1|0}$, we can derive the norming functions $a_t(x)$ and $b_t(x)$ for $t = 1, \ldots, k-1$. We consider three different scenarios for their asymptotic behaviour. Specifically, in Section 2.3, we have $a_t(x)/b_t(x) \to \infty$ as $x \to \infty$. In Section 2.4 we cover the case where $a_t(x) = 0$ for all $x > 0$ and $b_t(x) \to \infty$ as $x \to \infty$, but implicitly that section also covers the case $a_t(x) = O(b_t(x))$ as $x \to \infty$, where in that case the associated $a_t(x)/b_t(x)$ term in (2) may tend to a constant. In Section 2.5 we consider the cases where both $a_t(x)$ and $b_t(x)$ are invariant to x. Sections 2.3–2.5 focus on forward hidden tail chains, whereas in Section 2.6 we look at the joint behaviour of backward and forward hidden tail chains; we therefore drop the term *forward* in describing π until then.

2.2. Marginal standardization

To facilitate the generality of our theoretical developments, our assumptions about the margins of the process throughout Section 2 only concern the tail behaviour of the random variable at which we condition the Markov process to exceed a level. This assumption is in the

style of theoretical approaches in conditional extreme value theory [8] and is made precise by Assumption A_0.

Assumption 1. *The distribution F_0 has upper endpoint ∞, and there exist a non-degenerate probability distribution H_0 on $[0, \infty)$ and a measurable norming function $\sigma(v) > 0$ such that*

$$\frac{F_0(v + \sigma(v)dx)}{\overline{F}_0(v)} \xrightarrow{\text{w}} H_0(dx) \quad \text{as } v \to \infty.$$

From [24], the limit distribution H_0 can be identified by a generalized Pareto distribution with a non-negative shape parameter, i.e. $H_0(x) = 1 - (1 + \xi x)^{-1/\xi}$, for $x > 0$ and $\xi \geq 0$. This covers the Pareto distribution, with power-law decay, when $\xi > 0$, and the exponential distribution when $\xi = 0$, taken as the limit as $\xi \to 0$.

2.3. Transitional behaviour for chains with location and scale norming

In this section we consider homogeneous Markov chains where the initial conditional distribution $F_{1:k-1\,|\,0}$ is such that in the limit (1) we have functions $a_t(x)/b_t(x) \to \infty$ as $x \to \infty$ for all $t = 1, \ldots, k-1$ when $k > 1$. Specifically, Assumption A_1 ensures that, given that an extreme event $\{X_0 = v\}$ occurs at time $t = 0$, there exist normalizing functions $a_t(x)$ and $b_t(x)$ for $t = 1, \ldots, k-1$ such that if the arguments of $F_{1:k-1\,|\,0}$ are affine-transformed using these functions, then a non-degenerate limiting initial conditional distribution G is obtained for $F_{1:k-1\,|\,0}$, in the limit as $v \to \infty$. This provides the first $k-1$ renormalized states of the Markov process after X_0.

Assumption 2. *(Behaviour of initial states in the presence of an extreme event). If $k > 1$, the initial conditional joint distribution function $F_{1:k-1\,|\,0}$ is such that there exist the following:*

(i) *for $t = 1, \ldots, k-1$, measurable functions $a_t : \mathbb{R} \to \mathbb{R}$ and $b_t : \mathbb{R} \to \mathbb{R}_+$, satisfying $a_t(v) + b_t(v)\,x \to \infty$ as $v \to \infty$, for all fixed $x \in \mathbb{R}$;*

(ii) *a distribution G supported on $\mathbb{R}^{k-1}$ that has non-degenerate margins such that*

$$\mathbb{P}\left(\frac{\boldsymbol{X}_{1:k-1} - \boldsymbol{a}_{1:k-1}(v)}{\boldsymbol{b}_{1:k-1}(v)} \in dz_{1:k-1} \,\Big|\, X_0 = v \right) \xrightarrow{\text{w}} G(dz_{1:k-1}) \quad \text{as } v \to \infty.$$

Note that such conditions are not required by [20] when $k = 1$. Assumption 2 implies that $a_t(v) \to \infty$ and $b_t(v) = o(a_t(v))$ as $v \to \infty$, since if $b_t(v)$ grew as fast as, or faster than, $a_t(t)$, then for a suitably selected $x \in \mathbb{R}$, with x negative, we would have that $a_t(v) + b_t(v)x \to -\infty$ as $v \to \infty$.

Remark 1. When saying that a distribution is supported on a subset A of $\mathbb{R}^k$, we do not allow the distribution to place mass at the boundary ∂A of A. This is an important distinction which restricts the class of possible initial conditional distribution functions $F_{1:k-1\,|\,0}$ that the homogeneous Markov chain can possess. See Example 4 in Section 5.3 for a stationary Markov chain which has limit G with mass on the boundary, and hence breaks the conditions of Assumption 2.

After we have initialized the states $X_0, \ldots, X_{k-1}$, a complete characterization of the advancing sequence of states for $t \geq k$ is given by the one-step transition probability kernel $\pi(\boldsymbol{x}_{t-k\,:\,t-1}, x_t)$. To motivate our next assumption about the behaviour of the transition probability kernel of the process, consider how a complete characterization may be given for

higher-order Markov processes with $k > 1$ using induction on $\mathbb{N}$. Fix a $t \geq k > 1$ and assume that there exist sequences of norming functions a_i and b_i, $i = 1, \ldots, t - 1$, such that

$$\frac{\boldsymbol{X}_{1:t-1} - \boldsymbol{a}_{1:t-1}(X_0)}{\boldsymbol{b}_{1:t-1}(X_0)} \mid \{X_0 > u\} \xrightarrow{d} \boldsymbol{Z}_{1:t-1} \quad \text{as } u \to \infty,$$

where each Z_i is a random variable with a non-degenerate distribution on $\mathbb{R}$. Therefore, what is required is to assert that, under the induction hypothesis, we can find a_t and b_t such that for $\boldsymbol{a}_{1:t} = (\boldsymbol{a}_{1:t-1}, a_t)$ and $\boldsymbol{b}_{1:t} = (\boldsymbol{b}_{1:t-1}, b_t)$, we have that $\{\boldsymbol{X}_{1:t} - \boldsymbol{a}_{1:t}(X_0)\}/\boldsymbol{b}_{1:t}(X_0) \mid \{X_0 > u\} \xrightarrow{d} \boldsymbol{Z}_{1:t}$, as $u \to \infty$, where Z_t is a random variable with a non-degenerate distribution supported on $\mathbb{R}$. To motivate our assumptions that guarantee this latter convergence, it suffices to consider marginal convergence, that is, the case where the distribution of $\{X_t - a_t(X_0)\}/b_t(X_0) \mid \{X_0 > u\}$ converges weakly under the induction hypothesis. Standard calculations give that

$$\mathbb{P}\left(\frac{X_t - a_t(X_0)}{b_t(X_0)} \leq x_t \,\Big|\, X_0 > u\right) \tag{5}$$

$$= \int_u^\infty \int_{\mathbb{R}^{t-1}} \frac{\mathbb{P}\left(\boldsymbol{X}_{0:t-1} \in d\boldsymbol{x}_{0:t-1}\right)}{\overline{F}_0(u)} \left[\int_{-\infty}^{x_t} \mathbb{P}\left(\frac{X_t - a_t(X_0)}{b_t(X_0)} \in dz_t \,\Big|\, \boldsymbol{X}_{0:t-1} = \boldsymbol{x}_{0:t-1}\right)\right]$$

for $t \geq k$. This expression can be simplified by exploiting the Markov chain properties of $\{X_t\}$ and by rearranging the expression to clarify the connections with transitions from the initial $k - 1$ states after an extreme event by a change of the variables being integrated. First, replace $a_t(X_0)$ by $a_t(x_0)$ in the innermost integral, by virtue of the conditioning on the exact value of X_0 being equal to x_0. Second, use the Markov property so that the conditioning on all previous states is reduced to conditioning on the previous k states. Third, change variables to $z_0 = \{x_0 - u\}/\sigma(u)$ and $z_i = \{x_i - a_i(x_0)\}/b_i(x_0)$, for $i = 1, \ldots, t - 1$. This sequence of operations shows that (5) equals

$$\int_0^\infty \frac{F_0\{v_u(dz_0)\}}{\overline{F}_0(u)} \times$$

$$\left[\int_{\mathbb{R}^{t-1}} \mathbb{P}\left(\frac{\boldsymbol{X}_{1:t-1} - \boldsymbol{a}_{1:t-1}(v_u(z_0))}{\boldsymbol{b}_{1:t-1}(v_u(z_0))} \in d\boldsymbol{z}_{1:t-1} \,\Big|\, X_0 = v_u(z_0)\right) \tag{6}\right.$$

$$\left.\times \left\{\int_{-\infty}^{x_t} \mathbb{P}\left(\frac{X_t - a_t(v_u(z_0))}{b_t(v_u(z_0))} \in dz_t \,\Big|\, \frac{\boldsymbol{X}_{t-k:t-1} - \boldsymbol{a}_{t-k:t-1}(v_u(z_0))}{\boldsymbol{b}_{t-k:t-1}(v_u(z_0))} = \boldsymbol{z}_{t-k:t-1}\right)\right\}\right]$$

for $t \geq k$, where $a_0(x) = x$, $b_0(x) = 1$ for all $x \in \mathbb{R}$, and $v_u(z_0) = u + \sigma(u) z_0$. Hence, convergence of the innermost integral in the curly parentheses in (6) is necessary for marginal convergence of the probability (5) as $v \to \infty$. To further simplify this integral, for $\boldsymbol{z} \in \mathbb{R}^{k-1}$, let

$$\boldsymbol{A}_t(v, \boldsymbol{z}) = \boldsymbol{a}_{t-k:t-1}(v) + \boldsymbol{b}_{t-k:t-1}(v)\,\boldsymbol{z}. \tag{7}$$

Then, observe that for functions $a \colon \mathbb{R}^k \to \mathbb{R}$ and $b \colon \mathbb{R}^k \to \mathbb{R}_+$, which for the present are arbitrary functions but are made precise in Proposition 1, the innermost integral can be written as

$$
\int_{-\infty}^{x_t} \mathbb{P}\left(\frac{X_t - a(\boldsymbol{X}_{t-k:t-1})}{b(\boldsymbol{X}_{t-k:t-1})} \in \right.
$$

$$
\left. \frac{dz_t}{\psi_{t,u}^b(\boldsymbol{X}_{t-k:t-1})} - \psi_{t,u}^a(\boldsymbol{X}_{t-k:t-1}) \,\middle|\, \boldsymbol{X}_{t-k:t-1} = \boldsymbol{A}_t(v_u(z_0), \boldsymbol{z}_{t-k:t-1}) \right)
$$

$$
= \int_{-\infty}^{x_t} \pi\left(\boldsymbol{A}_t(v_u(z_0), \boldsymbol{z}_{t-k:t-1}), \right.
$$

$$
\left. a(\boldsymbol{X}_{t-k:t-1}) + b(\boldsymbol{X}_{t-k:t-1}) \left(\frac{dz_t}{\psi_{t,u}^b(\boldsymbol{X}_{t-k:t-1})} - \psi_{t,u}^a(\boldsymbol{X}_{t-k:t-1}) \right) \right),
$$

where $\psi_{t,u}^a$ and $\psi_{t,u}^b$ are given in the expression (8) and depend on the functions a, b and the normings a_t, b_t, $t \geq k$. Writing the integral in this way provides the connection between the convergence of the transition probability kernel π, defined in the expression (4), and the required marginal convergence.

These observations motivate our next assumption, which serves as an extension to higher order ($k > 1$) of the conditions established by [20] in the context of first-order Markov chains. Specifically, to establish the convergence of this rearranged integral, the oscillation of the functions a and b in a neighbourhood of infinity needs to be controlled, i.e. a and b need to be chosen so that the functions $\psi_{t,u}^a$ and $\psi_{t,u}^b$ converge locally uniformly to real-valued limits (ψ_t^a and ψ_t^b, respectively), as $u \to \infty$. These conditions are made precise by Assumption 3.

Assumption 3. *(Behaviour of the next state of the process as the previous states become extreme.) Let $k \geq 1$, $a_0(x) = x$, and $b_0(x) = 1$. If $k > 1$, suppose that for $X_{1:k-1}$, Assumption 2 holds with norming functions $\boldsymbol{a}_{1:k-1}$ and $\boldsymbol{b}_{1:k-1}$. Then assume that π is such that there exist the following:*

(i) *for $t = k, k+1, \ldots$, measurable functions $a_t \colon \mathbb{R} \to \mathbb{R}$ and $b_t \colon \mathbb{R} \to \mathbb{R}_+$, continuous update functions $\psi_t^a \colon \mathbb{R}^k \to \mathbb{R}$, $\psi_t^b \colon \mathbb{R}^k \to \mathbb{R}_+$, and measurable functions $a \colon \mathbb{R}^k \to \mathbb{R}$, $b \colon \mathbb{R}^k \to \mathbb{R}_+$, such that for all $\boldsymbol{z} \in \mathbb{R}^k$*

$$
\psi_{t,v}^a(\boldsymbol{z}_v) := \frac{a(\boldsymbol{A}_t(v, \boldsymbol{z}_v)) - a_t(v)}{b_t(v)} \to \psi_t^a(\boldsymbol{z}) \quad \text{and} \quad \psi_{t,v}^b(\boldsymbol{z}_v) := \frac{b(\boldsymbol{A}_t(v, \boldsymbol{z}_v))}{b_t(v)} \to \psi_t^b(\boldsymbol{z}) \tag{8}
$$

whenever $\boldsymbol{z}_v \to \boldsymbol{z}$ as $v \to \infty$, where $\boldsymbol{A}_t(v, \boldsymbol{z})$ is defined by (7);

(ii) *a non-degenerate distribution K_A supported on $\mathbb{R}$, such that for all $\boldsymbol{z} \in \mathbb{R}^k$ and for any $f \in C_b(\mathbb{R})$*

$$
\int_{\mathbb{R}} f(x)\pi[\boldsymbol{A}_t(v, \boldsymbol{z}_v), a(\boldsymbol{A}_t(v, \boldsymbol{z}_v)) + b(\boldsymbol{A}_t(v, \boldsymbol{z}_v))\, dx] \to \int_{\mathbb{R}} f(x) K_A(dx), \quad t = k, k+1, \ldots, \tag{9}
$$

whenever $\boldsymbol{z}_v \to \boldsymbol{z}$ as $v \to \infty$.

We are now able to characterize the hidden tail chain for this type of homogeneous Markov chain. Specifically, we combine assumptions about the marginal tail behaviour of X_0 (through H_0), the conditional limiting behaviour of the initial state vector given that X_0 is extreme (described by G), and the constraints on the chain's transitional behaviour, represented by the functions a and b, through π. This is asserted by Theorem 1.

Theorem 1. *Let $\{X_t : t \in \mathbb{Z}\}$ be a homogeneous kth-order Markov chain satisfying Assumptions 1, 2, and 3. Then as $v \to \infty$*

$$\left(\frac{X_0 - v}{\sigma(v)}, \frac{X_1 - a_1(X_0)}{b_1(X_0)}, \dots, \frac{X_t - a_t(X_0)}{b_t(X_0)} \right) \Big| \{X_0 > v\} \overset{d}{\longrightarrow} (E_0, Z_1, \dots, Z_t), \quad (10)$$

where

  (i) *$E_0 \sim H_0$ and the vector $(Z_1, Z_2, \dots, Z_t)$ are independent, for all $t \geq 1$,*

  (ii) *$Z_0 = 0$ almost surely (a.s.), $(Z_1, \dots, Z_{k-1}) \sim G$, and*

$$Z_t = \psi_t^a(\mathbf{Z}_{t-k\,:\,t-1}) + \psi_t^b(\mathbf{Z}_{t-k\,:\,t-1})\, \varepsilon_t, \qquad t = k, k+1, \dots, \quad (11)$$

   for a sequence of i.i.d. random variables $\varepsilon_k, \varepsilon_{k+1}, \dots$ with non-degenerate marginal distribution function K_A, defined by the limit (9).

Theorem 1 provides a highly structured limiting hidden tail chain, with the first value distributed as a generalized Pareto variable, with non-negative shape parameter, which is independent of the rest of the hidden tail; an initial conditional distribution given by G; and subsequently a kth-order autoregressive behaviour. In Theorem 1, we do not attempt to clarify what properties ψ_t^a and ψ_t^b possess, or to give how the transition functionals a and b are derived. Additional assumptions in later sections provide these details. However, under the current assumptions, Proposition 1 makes a connection between the required a_t and b_t functions for $t \geq k$ with these functions for $1 \leq t \leq k - 1$.

Proposition 1. *Let $a : \mathbb{R}^k \to \mathbb{R}$ and $b : \mathbb{R}^k \to \mathbb{R}_+$ be measurable maps. Let $t \geq k$ and $\mathbf{z} \in \mathbb{R}^k$. The following statements are equivalent:*

  (i) *There exist measurable functions $a_t : \mathbb{R} \to \mathbb{R}$, $b_t : \mathbb{R} \to \mathbb{R}_+$ and continuous functions $\psi_t^a : \mathbb{R}^k \to \mathbb{R}$ and $\psi_t^b : \mathbb{R}^k \to \mathbb{R}_+$ such that the convergence (8) holds.*

  (ii) *There exist continuous functions $\lambda_t^a : \mathbb{R}^k \to \mathbb{R}$ and $\lambda_t^b : \mathbb{R}^k \to \mathbb{R}_+$ such that for all $\mathbf{z} \in \mathbb{R}^k$*

$$\frac{a(\mathbf{A}_t(v, \mathbf{z}_v)) - a(\mathbf{A}_t(v, \mathbf{0}_k))}{b(\mathbf{A}_t(v, \mathbf{0}_k))} \to \lambda_t^a(\mathbf{z}) \quad \text{and} \quad \frac{b(\mathbf{A}_t(v, \mathbf{z}_v))}{b(\mathbf{A}_t(v, \mathbf{0}_k))} \to \lambda_t^b(\mathbf{z})$$

   whenever $\mathbf{z}_v \to \mathbf{z}$ as $v \to \infty$, where a and b are as defined in Assumption 3 and $\mathbf{A}_t(v, \mathbf{z})$ is as defined by (7).

2.4. Transitional behaviour for nonnegative chains with only scale norming

Consider nonnegative homogeneous Markov chains, i.e. with $F_t(0) = 0$ for all $t \in \mathbb{Z}$, where the initial conditional distribution $F_{1\,:\,k-1\,|\,0}$ is such that in the limit (1) we have functions such that there is no need for norming of the location, i.e. we can take $a_t(X_0) = 0$, and yet we need a scaling as $b_t(x) \to \infty$ for $t = 1, \dots, k - 1$, when $k > 1$. As in [20], we require extra care

relative to Section 2.3 since the convergences in Assumption 3(i) will be satisfied for all $x \in (0, \infty)$, but not all $x \in [0, \infty)$. Hence, we have to control the mass of the limiting renormalized initial conditional distribution and the limiting renormalized transition probability kernel of the Markov process. The strategy in this case is otherwise similar to that of Section 2.3, so we give only the statements of the key equivalent results.

Assumption 4. *(Behaviour of initial states in the presence of an extreme event.)* If $k > 1$, the initial conditional joint distribution function $F_{1:k-1|0}$ is such that there exist measurable functions $b_t : \mathbb{R}_+ \to \mathbb{R}_+$ for $t = 1, \ldots, k-1$ such that $b_t(v) \to \infty$ as $v \to \infty$, and a non-degenerate distribution function G on $[0, \infty)^k$, with no mass at any of the half-planes $C_j = \{(z_1, \ldots, z_{k-1}) \in [0, \infty)^{k-1} : z_j = 0\}$, i.e. $G(\{C_j\}) = 0$ for $j = 1, \ldots, k-1$, such that

$$\mathbb{P}\left(\frac{\boldsymbol{X}_{1:k-1}}{\boldsymbol{b}_{1:k-1}(v)} \in dz_{1:k-1} \,\Big|\, X_0 = v \right) \xrightarrow{\text{w}} G(dz_{1:k-1}) \quad \text{as } v \to \infty.$$

Assumption 5. *(Behaviour of the next state of the process as the previous states become extreme.)* Let $k \geq 1$ and $b_0(x) = x$, and when $k > 1$, additionally suppose that for $\boldsymbol{X}_{1:k-1}$, Assumption 4 holds with norming functions $\boldsymbol{b}_{1:k-1}$. Then assume that π is such that there exist the following:

(i) for $t = k, k+1, \ldots$, measurable functions $b_t : \mathbb{R}_+ \to \mathbb{R}_+$, continuous update functions $\psi_t^b : \mathbb{R}_+^k \to \mathbb{R}_+$, and a measurable function $b : \mathbb{R}_+^k \to \mathbb{R}_+$ such that for all $\delta_1, \ldots, \delta_k > 0$ and $z \in [\delta_1, \infty) \times \ldots \times [\delta_k, \infty)$,

$$\lim_{v \to \infty} \frac{b(\boldsymbol{B}_t(v, z_v))}{b_t(v)} = \psi_t^b(z) > 0, \tag{12}$$

whenever $z_v \to z$ as $v \to \infty$ and $\sup\{\|z\|_\infty : z \in A_c\} \to 0$ as $c \downarrow 0$, where $\boldsymbol{B}_t(v, z) := \boldsymbol{b}_{t-k:t-1}(v) z$ and $A_c = \{z \in (0, \infty)^k : \psi_t^b(z) \leq c\}$, with the convention that $\sup(\emptyset) = 0$;

(ii) a non-degenerate distribution K_B supported on $[0, \infty)$ with no mass at $\{0\}$, that is, $K_B\{0\} = 0$, such that, for any $f \in C_b(\mathbb{R}_+)$,

$$\int_{\mathbb{R}_+} f(x) \, \pi[\boldsymbol{B}_t(v, z_v), b(\boldsymbol{B}_t(v, z_v))\, dx] \to \int_{\mathbb{R}_+} f(x) \, K_B(dx), \qquad t = k, k+1, \ldots, \tag{13}$$

whenever $z_v \to z$ as $v \to \infty$.

Theorem 2. *Let $\{X_t : t \in \mathbb{Z}\}$ be a homogeneous Markov chain satisfying Assumptions 1, 4, and 5. Then as $v \to \infty$*

$$\left(\frac{X_0 - v}{\sigma(v)}, \frac{X_1}{b_1(X_0)}, \cdots, \frac{X_t}{b_t(X_0)} \right) \,\Big|\, \{X_0 > v\} \xrightarrow{d} (E_0, Z_1, \ldots, Z_t), \tag{14}$$

where

(i) $E_0 \sim H_0$ and the vector $(Z_1, Z_2, \ldots, Z_t)$ are independent for any $t \geq 1$,

(ii) $Z_0 = 1$ a.s., $(Z_1, \ldots, Z_{k-1}) \sim G$, and

$$Z_t = \psi_t^b(\boldsymbol{Z}_{t-k:t-1}) \varepsilon_t, \quad t = k, k+1, \ldots, \tag{15}$$

for a sequence of i.i.d. random variables $\varepsilon_k, \varepsilon_{k+1}, \ldots$ with non-degenerate marginal distribution function K_B defined by the limit (13).

Remark 2. Theorem 2 appears simply to be the kth-order extension of Theorem 3.1 in [14], but it differs: H_0 includes both Pareto and exponential tails, so it is in a more general class than the Pareto family considered by [14].

2.5. Transitional behaviour for near extremally independent chains

In this section, we consider homogeneous Markov chains where the initial conditional distribution $F_{1:k-1|0}$ is such that there no norming of the location and no norming of the scale are needed for the limit (1) to hold. This case resembles the formulation of Theorem 1, but has $a_t(v) = 0$ and $b_t(v) = 1$ for all $t \geq 1$. The next assumption ensures that after an extreme event at time $t = 0$, a non-degenerate distribution, given by $\{X_0 = v\}$, is obtained in the limit as $v \to \infty$ for the first k states of the Markov process, without any renormalization.

Assumption 6. *(Behaviour of next states in the presence of an extreme event.) If $k > 1$, the initial conditional joint distribution function $F_{1:k-1|0}$ is such that there exists a distribution G supported on $\mathbb{R}^k$ that has non-degenerate margins such that*

$$\mathbb{P}(X_{1:k} \in dz_{1:k} \mid X_0 = v) \xrightarrow{\mathrm{w}} G(dz_{1:k}) \quad \text{as } v \to \infty.$$

[9] showed that Assumption 6 holds for the Morgenstern copula, with exponential marginals, for $k \geq 1$. A related assumption also appears [18] for the case $k = 1$ with (X_0, X_1) being nonnegative random variables. Here, we note that if $X_{0:k}$ has the independence copula, then $G(z_{1:k}) = \prod_{j=1}^k F_j(z_j)$, whereas cases with $G_j(z) \geq F_j(z)$ ($G_j(z) \leq F_j(z)$) for all $z \in \mathbb{R}$, with G_j being the jth marginal distribution of G, correspond to positive (negative) near extremal independence at lag j in the hidden tail chain.

Assumptions 1 and 6 are sufficient to establish the weak convergence of the conditioned Markov chain to a hidden tail chain in Theorem 3 below. The proof of this theorem follows along the lines of the proof of Theorem 1 and is omitted for brevity.

Theorem 3. *Let $\{X_t : t \in \mathbb{Z}\}$ be a homogeneous kth-order Markov chain satisfying Assumptions 1 and 6. Then as $v \to \infty$*

$$\left(\frac{X_0 - v}{\sigma(v)}, X_1, \ldots, X_t \right) \Big| \{X_0 > v\} \xrightarrow{d} (E_0, Z_1, \ldots, Z_t), \tag{16}$$

where

(i) *$E_0 \sim H_0$ and the vector $(Z_1, Z_2 \ldots, Z_t)$ are independent for any $t \geq 1$,*

(ii) *$(Z_1, \ldots, Z_k) \sim G$ and*

$$Z_t = \pi^{-1}(\mathbf{Z}_{t-k:t-1}, U_t), \qquad t = k+1, k+2, \ldots,$$

where $\{U_t\}$ is a sequence of i.i.d. uniform(0,1) random variables for $t \geq k+1$, π is the one-step transition probability kernel for the original Markov chain, and $\pi^{-1} : \mathbb{R}^k \times (0, 1) \to \mathbb{R}$ with $\pi^{-1}(z, u) := \inf\{x \in \mathbb{R} : \pi(z, x) > u\}$.

We note that if $X_{0:k}$ has the independence copula, then $\pi^{-1}(z, u)$ is independent of z.

2.6. Back-and-forth hidden tail chains

In the discussion above, formally the entities we have referred to as tail chains and hidden tail chains are in fact forward tail and hidden tail chains [cf. 14]. These describe the behaviour

of the Markov chain only forward in time from a large observation. There is also the parallel interest in backward tail and hidden tail chains, to give how the chain evolves into an extreme event, and the joint behaviour of the two, known as back-and-forth tail processes.

Here we focus on an extension of the back-and-forth tail chains developed by [11]. The properties of the backward hidden tail chain are similar in structure to those of the forward hidden tail chain identified in Sections 2.3–2.4. To save repetition, here we outline the back-and-forth hidden tail chains for the assumptions in Section 2.3 only. For this purpose, it suffices to consider a straightforward extension of Assumption 3 which allows us to characterize the backward behaviour of the chain from an extreme event by requiring a functional normalization for the backward chain $X_{-s} \,|\, X_{-(s-1)\,:\,-(s-k)}$, $s \in \mathbb{N}$. Clearly, if the chain is time-reversible, then Assumption 3 holds backwards with the same functional normalizations a and b and the same limit distribution K_A. In general, however, there is no mathematical connection between these forward and backward quantities; Assumption 7 below considers this more general case.

Assumption 7. *(Behaviour of the backward state of the process.) Let $k \geq 1$, $a_0(x) = x$, and $b_0(x) = 1$. If $k > 1$, suppose that for $\boldsymbol{X}_{1\,:\,k-1}$, Assumption 2 holds with norming functions $\boldsymbol{a}_{1\,:\,k-1}$ and $\boldsymbol{b}_{1\,:\,k-1}$. Then assume that π^- is such that there exist the following:*

(i) for $s = 1, 2, \ldots$, measurable functions $a_{-s} : \mathbb{R} \to \mathbb{R}$ and $b_{-s} : \mathbb{R} \to \mathbb{R}_+$, continuous update functions $\psi_{-s}^{a^-} : \mathbb{R}^k \to \mathbb{R}$, $\psi_{-s}^{b^-} : \mathbb{R}^k \to \mathbb{R}_+$, and measurable functions $a^- : \mathbb{R}^k \to \mathbb{R}$, $b^- : \mathbb{R}^k \to \mathbb{R}_+$ such that, for all $z \in \mathbb{R}^k$,

$$\frac{a^-(\boldsymbol{A}_{-s}(v, z_v)) - a_{-s}(v)}{b_{-s}(v)} \to \psi_{-s}^{a^-}(z) \quad \text{and} \quad \frac{b^-(\boldsymbol{A}_{-s}(v, z_v))}{b_{-s}(v)} \to \psi_{-s}^{b^-}(z), \qquad (17)$$

whenever $z_v \to z$ as $v \to \infty$, where $\boldsymbol{A}_{-s}(v, z) := \boldsymbol{a}_{-s+1\,:\,-s+k}(v) + \boldsymbol{b}_{-s+1\,:\,-s+k}(v)\,z$;

(ii) a non-degenerate distribution K_A^- supported on $\mathbb{R}$, such that for all $z \in \mathbb{R}^k$ and for any $f \in C_b(\mathbb{R})$,

$$\int_{\mathbb{R}} f(x)\pi^-[\boldsymbol{A}_{-s}(v, z_v), a^-(\boldsymbol{A}_{-s}(v, z_v)) + b^-(\boldsymbol{A}_{-s}(v, z_v))\,dx] \to \int_{\mathbb{R}} f(x)K_A^-(dx), \qquad (18)$$

$s = 1, 2, \ldots$, whenever $z_v \to z$ as $v \to \infty$.

The back-and-forth hidden tail chain is presented in Theorem 4. For the sake of brevity, we do not include its proof, as this is identical to the proof of Theorem 1.

Theorem 4. *Let $\{X_t : t \in \mathbb{Z}\}$ be a homogeneous kth-order Markov chain satisfying Assumptions 1, 2, 3, and 7. Then as $v \to \infty$*

$$\left(\frac{\boldsymbol{X}_{-s\,:\,-1} - \boldsymbol{a}_{-s\,:\,-1}(X_0)}{\boldsymbol{b}_{-s\,:\,-1}(X_0)}, \frac{X_0 - v}{\sigma(v)}, \frac{\boldsymbol{X}_{1\,:\,t} - \boldsymbol{a}_{1\,:\,t}(X_0)}{\boldsymbol{b}_{1\,:\,t}(X_0)} \right) \Bigg| \{X_0 > v\}$$

$$\xrightarrow{d} (\boldsymbol{Z}_{-s\,:\,-1}, E_0, \boldsymbol{Z}_{1\,:\,t}), \qquad t, s \in \mathbb{N},$$

where

    (i) $E_0 \sim H_0$ is independent of the vector $(\boldsymbol{Z}_{-s\,:\,-1}, \boldsymbol{Z}_{1\,:\,t})$ for each $s, t \geq 1$,

    (ii) $Z_0 = 0$ a.s., $\boldsymbol{Z}_{1\,:\,k-1} \sim G$,

$$Z_t = \psi_t^a(\boldsymbol{Z}_{t-k\,:\,t-1}) + \psi_t^b(\boldsymbol{Z}_{t-k\,:\,t-1})\,\varepsilon_t, \qquad t = k, k+1, \ldots,$$

and

$$Z_{-s} = \psi_{-s}^{a^-}(\mathbf{Z}_{-s+1\,:\,-s+k}) + \psi_{-s}^{b^-}(\mathbf{Z}_{-s+1\,:\,-s+k})\,\varepsilon_{-s}, \qquad s = 1, 2, \ldots,$$

for independent sequences of i.i.d. random variables $\{\varepsilon_{-s}\}_{s=1}^{\infty}$ *and* $\{\varepsilon_t\}_{t=k}^{\infty}$, *where* $\varepsilon_{-s} \sim K_A^-$ *and* $\varepsilon_t \sim K_A$, *and* K_A^- *and* K_A *are defined by the limits* (18) *and* (9), *respectively.*

In general there is a relationship between the forward and backward hidden tail chains. When $k = 1$ they are independent, but when $k > 1$ and $t + s > k$, Z_t is conditionally independent of Z_{-s} given $(\mathbf{Z}_{-s+1\,:\,-1}, \mathbf{Z}_{1\,:\,t-1})$. Hence, given any consecutive block of terms in the back- and forth hidden tail chain of size k, the values before and after this block are independent. We remark that the precise dependence conditions between the forward and backward hidden tail chains have been given for the case where only $a_j(x) = x$ and $b_j(x) = 1$ for all $j \neq 0$ by [11]. Subsequently, we focus on the forward hidden tail chain and do not address the inter- connections between the different $a_j(x)$ and $b_j(x)$ for positive and negative j.

Remark 3. When $k = 1$, by Proposition 1 we can choose, without loss of generality, $a(v) = a_1(v)$ and $b(v) = b_1(v)$ so that $\psi_1^a(0) = \{a(v) - a_1(v)\}/b_1(v) = 0$ and $\psi_1^b(0) = b(v)/b_1(v) = 1$. Consequently, (11) implies that $Z_1 = \varepsilon_1$ and thus, the special case of $k = 1$ in Theorem 1 corre- sponds to the results of [20]. The methods of identifying the functions a and b more generally are discussed in Sections 3 and 5.

3. Dependence and recurrence equations under stationarity

3.1. Stationarity and parametric conditional extremes models

For homogeneous Markov chains, the theory presented in Section 2 generalizes to kth-order homogeneous Markov chains (with $k > 1$) the results for $k = 1$ presented in [20]. Although working with homogeneous chains embeds the theory in a rather broad setting, it is impossible to explore more details of the results of Theorems 1–4 without imposing further structure, and we do this by assuming stationarity of the Markov chains for the remainder of the paper.

So we now have a common marginal distribution function F over t, with this distribution satisfying Assumption A_0, unlike previously, when this was assumed only for the variable X_0. Specifically, we assume that the stationary Markov chain $\{X_t\}$ has unit-exponential marginal distributions, that is, $F(x) := \mathbb{P}(X_t \leq x) = (1 - \exp(-x))_+$, for $t \in \mathbb{Z}$, which implies that the limit distribution H_0 in Assumption 1 is also unit-exponential. This marginal choice gives the clearest mathematical formulation for our needs [20], and if a stationary Markov chain $\{\widetilde{X}_t\}$ has marginal distribution $\widetilde{F} \neq F$ then transformation by the probability integral transform $X_t = -\log[1 - \widetilde{F}(\widetilde{X}_t)]$, for $t \in \mathbb{Z}$, gives the required properties. Furthermore, from stationarity we have a time-invariant copula for $\mathbf{X}_{t-k\,:\,t}$, for all $t \in \mathbb{Z}$, which together with the marginals can be used to derive the initial conditional distribution $F_{1\,:\,k-1|0}$ and the corresponding $a_t(x)$ and $b_t(x)$ for $t = 1, \ldots, k - 1$, as well as the forms of both π and π^-, and to show that they are also time-invariant functions.

With the marginal distribution now fully defined, we look to narrow down the proper- ties of the copula into a single class that covers all of the scenarios discussed in Section 2. This approach is in the style of copula methods, where the assumption of identical margins is typical when identifying the extremal dependence structure of a random vector. In partic- ular, [9] found that for a broad range of copula models for a random vector $\mathbf{X}_{0\,:\,d}$, $d \in \mathbb{N}$, with exponential-tailed random variables, the conditional distribution of the renormalized states $\{\mathbf{X}_{1\,:\,d} - \boldsymbol{a}_{1\,:\,d}(X_0)\}/\boldsymbol{b}_{1\,:\,d}(X_0)$, given $X_0 > v$, weakly converges to some distribution with

non-degenerate margins. They also identified that this convergence typically holds with the normalization functions taking the simple form $\boldsymbol{a}_{1:d}(v) = \boldsymbol{\alpha}_{1:d} v$ and $\boldsymbol{b}_{1:d}(v) = v^\beta \mathbf{1}_d$, where $(\boldsymbol{\alpha}_{1:d} = (\alpha_1, \ldots, \alpha_d), \beta) \in [0, 1]^d \times [0, 1)$. The parameters α_t and β have a simple interpretation and control the strength of extremal association between the variables X_0 and X_t, for $t = 1, \ldots, d$. Informally, in the presence of an extreme event X_0 with $X_0 > v$ and v sufficiently large, we may then think of X_t as $X_t = \alpha_t X_0 + X_0^\beta Z_t$ where Z_t has a non-degenerate distribution. Thus, α_t and β are slope and scale parameters, respectively, with larger values of α_t indicating stronger linear dependence between X_t and the big values of X_0, and, for fixed α_t, the larger values of β indicating a more diffuse distribution for $X_t \mid X_0 > v$. The links to Section 2 and full pairwise asymptotically dependent and full pairwise asymptotically independent kth-order Markov processes can now be made much clearer. Specifically, $\alpha_t \in (0, 1]$ for all $t = 1, \ldots, k$ corresponds to a family of copulas satisfying the conditions of Section 2.3 with the special case of $\alpha_t = 1$ and $\beta = 0$ for all $t = 1, \ldots, k - 1$ being a full pairwise asymptotically dependent Markov chain, with $\chi_t > 0$ for $t = 1, \ldots, k - 1$, so it only arises as a special case in Section 2.3. In contrast, full pairwise asymptotically independent processes arise as a broad class in Section 2.3 and all scenarios in Sections 2.4 and 2.5. Specifically, we have pairwise asymptotic independence in Section 2.3 when $\alpha_t \in [0, 1)$ for all $t = 1, \ldots, k - 1$, in Section 2.4 when $\alpha_t = 0$ for all $t = 1, \ldots, k - 1$ and $\beta \in (0, 1)$, and in Section 2.5 when $\alpha_t = 0$ for all $t = 1, \ldots, k - 1$ and $\beta = 0$. Here we assume that the formulation of [9] holds for the initial conditional distribution, i.e. giving α_t for $t = 1, \ldots, k - 1$ and β, and we focus on deriving the structure of α_t and β_t for all $t \geq k$ and the stochastic recurrence properties of the hidden tail chains for both full pairwise asymptotically dependent and full pairwise asymptotically independent kth-order stationary Markov chains.

3.2. Full pairwise asymptotically dependent Markov chains

Corollary 1. (Full pairwise asymptotically dependent Markov chains).

Let $\{X_t : t \in \mathbb{Z}\}$ be a kth-order stationary Markov chain with unit-exponential margins. Suppose that Assumption 2 holds with $a_t(x) = x$ and $b_t(x) = 1$, for $t = 1, \ldots, k - 1$. Suppose further that Assumption 3 holds with functions a and b, defined there, such that a is non-zero and continuous, and $\exp\{a(\log \boldsymbol{x})\}$, $\boldsymbol{x} \in \mathbb{R}_+^k$, is 1-homogeneous; that is,

$$a(v\mathbf{1}_k + \boldsymbol{y}) - v = a(\boldsymbol{y}) \quad \text{for all } v \in \mathbb{R} \text{ and all } \boldsymbol{y} \in \mathbb{R}^k, \tag{19}$$

with $a(\mathbf{0}_k) \leq 0$ and $b(\cdot) \equiv 1$. Then the convergence (10) holds with $a_t(x) = x$ and $b_t(x) = 1$ for $t \geq k$, and

$$Z_t = a(\mathbf{Z}_{t-k:t-1}) + \varepsilon_t, \quad t = k, k+1, \ldots,$$

for a sequence $\{\varepsilon_t\}_{t=k}^\infty$ of i.i.d. random variables with a distribution K_A, as defined in the limit (9), supported on $\mathbb{R}$. Furthermore, $\mathbb{E}(Z_t) < 0$ for all $t \geq 1$.

So in terms of the parameters of [9], here $\alpha_t = 1$ and $\beta_t = 0$ for all $t \in \mathbb{Z}$, and consequently the process is pairwise asymptotically dependent for (X_0, X_t), i.e., $\chi_t > 0$, for any $t \geq k$. Although this gives that $a_t(x) = x$ and $b_t(x) = 1$ for all $t \geq 1$, suggesting that asymptotically dependent Markov chains stay extreme forever after witnessing an extreme value, they do in fact return to the body of the distribution thanks to the negative drift of the tail chain; that is, $\mathbb{E}(Z_t) < 0$ for all $t \geq 1$, which ensures that the Markov chain will return to the body regardless of the behaviour of the norming functions. Corollary 1 refines our understanding of

full pairwise asymptotically dependent Markov chains by identifying a class of update functions *a* whose defining property is (19), also known as topical maps, which can be regarded as nonlinear generalizations of row-stochastic matrices [17].

3.3. Full pairwise asymptotically independent Markov chains

As there are different types of dependence that lead to full pairwise asymptotic independent behaviour, we separate our findings into Corollaries 2 and 3, which correspond to the conditions of Sections 2.3 and 2.4, respectively.

Corollary 2. (Full pairwise asymptotically independent Markov chains with location and scale norming.)

Let $\{X_t : t \in \mathbb{Z}\}$ be a kth-order stationary Markov chain with unit-exponential margins. Suppose Assumption 2 holds with $a_t(v) = \alpha_t v$, $\alpha_t \in (0, 1)$ and $b_t(v) = v^{\beta_t}$, $\beta_t = \beta \in [0, 1)$, for $t = 1, \ldots, k - 1$. Suppose that Assumption 3 holds with the function a being a twice continuously differentiable, order-preserving 1-homogeneous function, with $a(\mathbf{1}_k) < 1$, and the function b being β-homogeneous when $\beta \in (0, 1)$ and unity when $\beta = 0$. Then the convergence (10) holds with, for $t \geq k$, $a_t(v) = \alpha_t v$ and $b_t(v) = v^{\beta}$, where

$$\alpha_t = a(\boldsymbol{\alpha}_{t-k:t-1}), \qquad t = k, k+1, \ldots, \tag{20}$$

with $\alpha_t \in (0, 1)$ for all $t \geq k$, and

$$Z_t = \nabla a(\boldsymbol{\alpha}_{t-k:t-1})^\top \mathbf{Z}_{t-k:t-1} + b(\boldsymbol{\alpha}_{t-k:t-1}) \varepsilon_t, \qquad t = k, k+1, \ldots, \tag{21}$$

for a sequence $\{\varepsilon_t\}_{t=k}^{\infty}$ of i.i.d. random variables from a non-degenerate distribution K_A, defined in the limit (9), on $\mathbb{R}$. Consequently, the process is pairwise asymptotically independent for (X_0, X_t), i.e. $\chi_t = 0$ for any $t \geq k$, and $\alpha_t \to 0$ as $t \to \infty$.

In Section 4 we are able to explicitly solve the recurrence equation (20) for a flexible parametric class of the function *a*, and find a geometric decay to zero in α_t as *t* increases. Even for an arbitrary functional *a* satisfying the weak assumptions of Corollary 2, considerable insight into the behaviour of the hidden tail chain is achieved from Corollary 2. It shows that the norming functions a_t, $t = k, k+1, \ldots$, have a particularly neat structure, not least $a_t(X_0) = \alpha_t X_0$, where α_t is determined by the recurrence equation (20) of the *k* previous values $\boldsymbol{\alpha}_{t-k:t-1}$ through the 1-homogeneous function *a*.

Since Corollary 2 gives that $\alpha_t \to 0$ as $t \to \infty$, this leads to there eventually being no location-norming in the limit, which is consistent with the independence case. However, for practical usage we need to consider the limit at $t \to \infty$ whilst also allowing the level of extremity of X_0 to increase, i.e. the value $v \to \infty$. We address these issues for the cases $\beta = 0$ and $\beta \in (0, 1)$. When $\beta = 0$, the behaviour of the forward Markov chain as $t \to \infty$ and $v \to \infty$ is almost entirely given by Corollary 2, as $X_t | \{X_0 = v\} = \alpha_t v + Z_t + o_p(1)$. If we can suitably link $t \to \infty$ with $v \to \infty$, the location term $\alpha_t v$ will tend to zero and X_t will converge to the process $\{Z_t\}$, which is a non-degenerate autoregressive process. So, with such combined limiting operations, we have that $X_t | \{X_0 = v\}$ returns to the body of the distribution as $t \to \infty$, becoming independent of X_0. Here, if there were a constant $A \in (0, 1)$ with $\alpha_t \sim A^t$ as $t \to \infty$, i.e. the α_t exhibit geometric decay, as in the parametric class for *a* in Section 4, then we would need $t / \log(v) \to \infty$ as $v \to \infty$ for this result to hold. When $0 < \beta < 1$ the limiting behaviour of the forward hidden chain $X_t | \{X_0 = v\}$ as $t \to \infty$ is only partially implied by Corollary 2. This is because $Z_t \xrightarrow{p} 0$, since both the location and the scale terms of ε_t in (21) tend to zero, but its

scaling v^β tends to infinity. Consequently, the limiting behaviour is determined by the relative speed of convergence of $\alpha_t\, v \to 0$ and $v^\beta\, Z_t \xrightarrow{p} 0$ if we link the growth rate of t to that of v.

In general we can view the recurrence relation in (20) as the parallel of the Yule–Walker equations; hence we term them the *extremal Yule–Walker equations*. The Yule–Walker equations provide a recurrence relation for the autocorrelation function in standard time series that is used to determine the dependence properties of a linear Markov process. For a kth-order linear Markov process $Y_t = \sum_{i=1}^{k} \phi_i\, Y_{t-k+i-1} + \eta_t$ with $\{\eta_t\}_{t=-\infty}^{\infty}$ a sequence of zero-mean, common-finite-variance, and uncorrelated random variables, where the set of regression parameters $\phi_1, \ldots, \phi_k$ are real-valued constants such that the characteristic polynomial $1 - \phi_k\, z - \phi_{k-1}\, z^2 - \cdots - \phi_1\, z^k \neq 0$ on $\{z \in \mathbb{C} : |z| \le 1\}$, the Yule–Walker equations relate the autocorrelation function of the process $\rho_t = \mathrm{cor}(Y_{s-t}, Y_s)$ at lag t with the regression parameters $\phi_1, \ldots, \phi_k$ and the k lagged autocorrelations according to $\rho_t = \sum_{i=1}^{k} \phi_i\, \rho_{t-i}$, $t \in \mathbb{Z}$. The sequence $\{\alpha_t\}$ has a similar structure for extremes via the recurrence (20).

In Corollary 2 we exclude the case $\beta < 0$ considered by [9], which corresponds to the case where location-only normalization gives limits that are degenerate, with all limiting mass at $\{0\}$. For simplicity, the theory developed in this paper deals only with positive extremal association in Markov chains, and hence, in Corollary 2, the case where X_0 and X_t exhibit negative extremal association is also ruled out. We note, however, that this latter case potentially can be accommodated by suitable transformations of the marginal distributions, e.g. by standardizing margins to standard Laplace distributions and then allowing $\alpha_t < 0$; see for example [13] and [20 Theorem 3]. However, some modifications in the conditions will be required to address the complications that some arguments will tend to $-\infty$ whilst the conditioning variable will still go to ∞.

Corollary 3. (Full pairwise asymptotically independent Markov chains with only scale norming) *Let $\{X_t : t \in \mathbb{Z}\}$ be a kth-order stationary Markov chain with unit-exponential margins. Suppose that Assumption 4 holds with $b_t(x) = x^\beta$, $\beta \in (0, 1)$, for $t = 1, \ldots, k-1$. Suppose further that Assumption 5 holds with the function b being continuous and β-homogeneous. Then the convergence (14) holds with $b_t(x) = x^{\beta_t}$, $\beta_t \in (0, 1)$, $t \ge k$, where β_t satisfies the recurrence relation $\log \beta_t = \log \beta + \log(\max_{i=1,\ldots,k} \beta_{t-i})$, with $\beta_1 = \ldots = \beta_{k-1} = \beta$. This gives the solution*

$$\log \beta_t = (\lfloor 1 + (t-1)/k \rfloor) \log \beta, \qquad t \ge k,$$

where $\lfloor x \rfloor$ denotes the integer part of x. It follows that $\beta_t \in (0, 1)$ for all $t \ge k$ and that $\beta_t \to 0$ as $t \to \infty$. Also, for $t \ge k$ we have

$$Z_t = \begin{cases} b(Z_{t-k}, \mathbf{0}_{k-1})\, \varepsilon_t & \text{when } \mathrm{mod}_k(t) = 0, \\ b(Z_{t-k}, \ldots, Z_{t-1})\, \varepsilon_t & \text{when } \mathrm{mod}_k(t) = 1, \\ b(Z_{t-k}, \ldots, Z_{t-j}, \mathbf{0}_{j-1})\, \varepsilon_t & \text{when } \mathrm{mod}_k(t) = j \in \{2, \ldots, k-1\}, \end{cases}$$

for a sequence $\{\varepsilon_t\}_{t=k}^{\infty}$ of i.i.d. random variables with distribution K_B, defined by the limit (13), supported on $\mathbb{R}_+$ and $Z_0 = 1$ a.s. As a consequence, the process is pairwise asymptotically independent for (X_0, X_t), i.e. $\chi_t = 0$, for any $t \ge k$.

In contrast to Corollary 2, where the location parameter α_t changed with t, here it is the power parameter β_t of the scale function that is decaying. As with Corollary 2 we find a form of geometric decay in the dependence parameters β_t as t increases, leading eventually to extremal independence ($\alpha_t = 0$ and $\beta_t \to 0$) in the limit as $t \to \infty$, so that X_t returns to the body of

the distribution as it becomes independent of X_0. In particular, β_t decays geometrically to 0 stepwise, with steps at every k lags. At time t the resulting hidden tail chain depends only on the last j values, with $j = \mathrm{mod}_k(t)$.

4. A class of recurrence relations for dependence parameters in asymptotically independent Markov chains with closed-form solutions

The results of Section 3 provide insight into the form of the norming and updating functions of Theorems 1 and 2, not least for asymptotically independent Markov chains where $(\alpha_t, \beta_t) \neq (1, 0)$ for all $t > 0$. A precise formulation of the location and scale parameters α_t and β_t for $t \geq k$, however, depends on the forms of functionals $a(\cdot)$ and $b(\cdot)$ which are opaque even when these are assumed to be homogeneous functionals. Motivated by examples considered in Section 5.3, here we give an explicit characterization of the solution to the extremal Yule–Walker equations (20) in Corollary 2 for a parsimonious parametric subclass of functionals $a_{\mathcal{M}}$, of a, which embeds many of the examples of Section 5.

For $\quad \boldsymbol{x} = (x_1, \ldots, x_k) \in \mathbb{R}_+^k, \quad 0 < c < 1, \quad \delta \in \overline{\mathbb{R}}, \quad$ and $\quad (\gamma_1, \ldots, \gamma_k) \in \Delta^{k-1} \quad$ with $\min_{j=1,\ldots,k} (\gamma_j) > 0$, consider the function $a_{\mathcal{M}} : \mathbb{R}_+^k \to \mathbb{R}_+$ defined by

$$a_{\mathcal{M}}(\boldsymbol{x}; \delta) = c \left\{ \frac{\gamma_1 (\gamma_1 x_1)^\delta + \cdots + \gamma_k (\gamma_k x_k)^\delta}{\gamma_1^{1+\delta} + \cdots + \gamma_k^{1+\delta}} \right\}^{1/\delta}. \tag{22}$$

Here $a_{\mathcal{M}}$ satisfies the conditions of Corollary 2, as it is order-preserving and 1-homogeneous, and the bound on c ensures that $a(\mathbf{1}_k) < 1$. The functional $a_{\mathcal{M}}$ is continuous in $\delta \in \overline{\mathbb{R}}$, though its values at $\delta = -\infty, 0, \infty$ need careful treatment as they are not immediately apparent from (22). Specifically,

$$a_{\mathcal{M}}(\boldsymbol{x}, 0) = \lim_{\delta \to 0} a_{\mathcal{M}}(\boldsymbol{x}, \delta) = c \, x_1^{\gamma_1} x_2^{\gamma_2} \cdots x_k^{\gamma_k}, \tag{23}$$

and

$$a_{\mathcal{M}}(\boldsymbol{x}, \infty) = c \, \frac{\max_{j=1,\ldots,k} (\gamma_j x_j)}{\max_{j=1,\ldots,k} \gamma_j}, \quad a_{\mathcal{M}}(\boldsymbol{x}, -\infty) = c \, \frac{\min_{j=1,\ldots,k} (\gamma_j x_j)}{\min_{j=1,\ldots,k} (\gamma_j)}. \tag{24}$$

Proposition 2. *Consider the function $a_{\mathcal{M}}$ defined by (22). Suppose that the $s \in \mathbb{N}$ distinct (possibly complex) roots of the characteristic polynomial*

$$x^k - c^\delta \left(\frac{\gamma_k^{1+\delta}}{\gamma_1^{1+\delta} + \cdots + \gamma_k^{1+\delta}} \right) x^{k-1} - \cdots - c^\delta \left(\frac{\gamma_1^{1+\delta}}{\gamma_1^{1+\delta} + \cdots + \gamma_k^{1+\delta}} \right) = 0$$

are $r_1, \ldots, r_s$ with multiplicities $m_1, \ldots, m_s$, $\sum_i m_i = k$. Then the solution of the recurrence relation (20) with $a(\boldsymbol{x}) = a_{\mathcal{M}}(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}_+^k$, subject to the initial condition $(\alpha_1, \ldots, \alpha_{k-1}) \in (0, 1)^{k-1}$, is

$$\alpha_t = \left(\sum_{i=1}^s \left(C_{i0} + C_{i1} t + \cdots + C_{i,m_i-1} t^{m_i-1} \right) r_i^t \right)^{1/\delta} \quad \text{for } t = k, k+1, \ldots,$$

where the constants $C_{i0}, \ldots, C_{i,m_i-1}$, $i = 1, \ldots, s$, are uniquely determined by the initial condition via the system of equations $\alpha_0 = 1$ and

$$\alpha_t = \left(\sum_{i=1}^s \left(C_{i0} + C_{i1} t + \cdots + C_{i,m_i-1} t^{m_i-1} \right) r_i^t \right)^{1/\delta} \quad \text{for } t = 0, \ldots, k-1.$$

From Corollary 2, it follows that the sequence $\{\alpha_t\}$ in Proposition 2 satisfies $\alpha_t \to 0$ as $t \to \infty$. Let $I_r = \{I \in \{1, \dots, s\} : |r_I| = \max_{i=1,\dots,s} |r_i|\}$. Under the assumption that $|I_r| = 1$, we have that α_t satisfies $\alpha_t \sim C_{I,m_I-1} \, t^{(m_I-1)/\delta} r_I^{t/\delta} \to 0$, $I \in I_r$, $\delta \in \mathbb{R} \backslash \{0\}$, as $t \to \infty$.

Remark 4. Although $a_{\mathcal{M}}$ in (22) is defined for any $\delta \in \overline{\mathbb{R}}$, it is not evident from Proposition 2 what form the solution takes when $\delta = 0$ or when $\delta = \pm\infty$. These cases are considered separately below.

Case $\delta \to 0$: A logarithmic transformation in the limit (23) results in the linear non-homogeneous recurrence relation

$$\log \alpha_t - \gamma_1 \log \alpha_{t-1} - \cdots - \gamma_k \log \alpha_{t-k} = \log c.$$

Suppose that the $s \in \mathbb{N}$ distinct (possibly complex) roots of the characteristic polynomial

$$x^k - \gamma_k \, x^{k-1} - \cdots - \gamma_1 = 0 \tag{25}$$

are $r_1, \dots, r_s$ with multiplicities $m_1, \dots, m_s$, $\sum m_i = k$. Then the solution of the recurrence (20) is

$$\alpha_t = \exp\left\{ \sum_{i=1}^{s} \left(C_{i0} + C_{i1} \, t + \cdots + C_{i,m_i-1} t^{m_i-1} \right) r_i^t + \frac{c}{\gamma_1 + 2\gamma_2 + \cdots + k\gamma_k} t \right\},$$

for $t = k, k+1, \dots$, where the constants $C_{i0}, \dots, C_{i,m_i-1}$, $i = 1, \dots, s$, are uniquely determined by the system of equations

$$\alpha_t = \exp\left\{ \sum_{i=1}^{s} \left(C_{i0} + C_{i1} \, t + \cdots + C_{i,m_i-1} t^{m_i-1} \right) r_i^t + \frac{c}{\gamma_1 + 2\gamma_2 + \cdots + k\gamma_k} t \right\},$$

for $t = 0, \dots, k-1$, with $\alpha_0 = 1$ and $\alpha_t \in (0, 1)$ for $t = 1, \dots, k-1$.

Case $|\delta| \to \pm\infty$: Using forward substitution, we have that for $\delta \to \infty$ the solution of (20) is

$$\alpha_t = c^t \max_{i=1,\dots,k} (d_{t-i} \, \alpha_{i-1}), \qquad d_{t-i} = \max \prod_{n=1}^{k} \gamma_{k+1-n}^{j_n}, \qquad t \geq k, \tag{26}$$

where the maximum for d_{t-i} in (26) is taken over $0 \leq j_1 \leq \dots \leq j_k \leq t-i$ such that $\sum_{m=1}^{t-i} m j_m = t-i$. The case $\delta \to -\infty$ is obtained by replacing the maximum operator in (26) by the minimum operator.

5. Results for kernels based on important copula classes

5.1. Strategy for finding norming functionals

Our results in Section 3 provide powerful results for stationary kth-order Markov processes which derive the behaviour of hidden tail chains over all lags, given the appropriate norming functions a_t, b_t, for lags $t = 1, \dots, k-1$, and the norming functionals a and b after an extreme event at time $t = 0$. However, these results do not explain how to derive these quantities. Here we discuss general strategies for how to find these norming functions, with Section 5.3 providing a step-by-step illustration of how these strategies are implemented.

The methods for finding a_t, b_t for $t = 1, \ldots, k-1$ are well established; specifically, these can be obtained from Theorem 1 of [9]. So the novelty here is in the derivation of a and b, which was not required in [20]. Here we explain the general strategy for the case $a \neq 0$ and note that the case $a = 0$ and $b \neq 1$ is handled in a similar manner. Assuming that the conditional distribution of $X_k \mid X_{0:k-1}$ admits a Lebesgue density almost everywhere, a similar argument as in the proof of Theorem 1 in [9] guarantees that the functionals a and b can be identified, up to type, by finding the functional forms of a and b, which satisfy the following two asymptotic properties: for all $z \in \mathbb{R}^k$,

$$\lim_{v \to \infty} \mathbb{P}\left(X_k < a(X_{0:k-1}) \mid X_{0:k-1} = A_t(v, z)\right) = p \quad \text{for some } p \in (0, 1), \tag{27}$$

and

$$b(A_t(v, z)) \sim \frac{\mathbb{P}\left(X_k > a(X_{0:k-1}) \mid X_{0:k-1} = A_t(v, z)\right)}{[d\,\mathbb{P}\left(X_k \leq y \mid X_{0:k-1} = A_t(v, z)\right)/dy]|_{y=a(A_t(v,z))}} \quad \text{as } v \to \infty, \tag{28}$$

where $A_t(v, z)$ is defined by (7) with $a_{t-k:t-1}$ and $b_{t-k:t-1}$ as in Assumption 3. The expression (28) can be cumbersome to use in practice, so we resort to asymptotic inversion in order to identify b. In particular, to find a representative form for b, we make an informed choice based on the leading-order terms in an asymptotic expansion, as $v \to \infty$, of the conditional distribution in (27) to obtain

$$\mathbb{P}\left(X_k < a(X_{0:k-1}) + b(X_{0:k-1})\,y \mid X_{0:k-1} = A_t(v, z)\right) \xrightarrow{\text{w}} K_A(y), \tag{29}$$

for all $y \in \mathbb{R}$ and $z \in \mathbb{R}^k$, where K_A is a non-degenerate distribution on $\mathbb{R}$.

5.2. Examples of copula families and transition probability kernels

To illustrate the results in Theorems 1 and 2, we study the extremal behaviour of kth-order stationary Markov chains with unit-exponential margins, with transition probability kernels derived from the copula of $k + 1$ consecutive values given in Section 5.3. Here we define the classes of copula families, and the associated transition probability kernels, that we use in Section 5.3 to derive the hidden tail chain behaviour. We have selected these copula families so we have families which give full pairwise asymptotic dependence, full pairwise asymptotic independence, and cases in between these; see Section 1 for the definition of these extremal process types. We can achieve all of these different properties from studying Gaussian, max-stable, and inverted max-stable copula families. The theory which motivates these copulas does not matter here; we simply view them as a range of interesting and well-known copula families whose extremes we study in a Markov setting.

First we present the link between a general $(k+1)$-dimensional copula C and its associated transition probability kernel π when working with a kth-order stationary Markov process with exponential margins. Let F denote the joint distribution function of a random vector $X = (X_0, \ldots, X_k)$, assumed to be absolutely continuous with respect to Lebesgue measure with unit-exponential margins, that is, $F_i(x) = F_E(x) = (1 - \exp(-x))_+$, $i = 0, \ldots, k$. Writing $C : [0, 1]^{k+1} \to [0, 1]$ for the copula of X, that is, $C(\boldsymbol{u}) = F\left(F_E^{\leftarrow}(u_0), \ldots, F_E^{\leftarrow}(u_k)\right)$, where $\boldsymbol{u} = (u_0, \ldots, u_k) \in [0, 1]^{k+1}$, we define the Markov kernel $\pi_E : \mathscr{B}(\mathbb{R}^k) \to [0, 1]$ of the stationary process by

$$\pi_E(\boldsymbol{x}_{0:k-1}, x_k) = \left[\frac{\partial^k}{\partial u_0 \cdots \partial u_{k-1}} C(\boldsymbol{u}_{0:k-1}, u_k) \Big/ \frac{\partial^k}{\partial u_0 \cdots \partial u_{k-1}} C(\boldsymbol{u}_{0:k-1}, 1) \right],$$

with $\boldsymbol{u}_{0:k} = \{1 - \exp(-\boldsymbol{x}_{0:k})\}_+$. Assuming the copula function satisfies appropriate conditions that ensure stationarity [12], the initial distribution $F(\boldsymbol{x}_{0:k-1}, \infty)$ is the k-dimensional invariant distribution of a Markov process with unit-exponential margins and kernel π_E.

Gaussian copula: Consider a stationary Gaussian autoregressive process with positive dependence transformed componentwise to have exponential marginal distributions. Let $\boldsymbol{\Sigma} \in \mathbb{R}^{(k+1)\times(k+1)}$ be a $(k+1)$-dimensional Toeplitz correlation matrix, that is, $\boldsymbol{\Sigma} = (\rho_{|i-j|})_{1\leq i,j\leq k+1}$ with $\rho_0 = 1$, $\rho_i > 0$, for $i = 1, \ldots, k$, which is assumed to be positive definite, and let $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1} = (q_{i-1,j-1})_{1\leq i,j\leq k+1}$. The distribution function of the standard $(k+1)$-dimensional Gaussian with mean $\boldsymbol{0}_{k+1}$ and positive definite variance–covariance matrix $\boldsymbol{\Sigma}$, in exponential margins, is

$$F(\boldsymbol{x}_{0:k}) = \int_{-\infty}^{\Phi^{\leftarrow}\{1-\exp(-x_k)\}} \cdots \int_{-\infty}^{\Phi^{\leftarrow}\{1-\exp(-x_0)\}} \frac{\{\det(\boldsymbol{Q})\}^{1/2}}{(2\pi)^{(k+1)/2}} e^{-s^\top \boldsymbol{Q} s/2} \, \mathrm{d}s,$$

for $\boldsymbol{x}_{0:k} \in \mathbb{R}_+^{k+1}$ and $\boldsymbol{s} = (s_0, \ldots, s_k)^\top$, where $\Phi^{\leftarrow} : [0,1] \to \mathbb{R}$ denotes the quantile function of the standard normal distribution function $\Phi(\cdot)$. This choice of correlation function ensures stationarity of the resulting Markov chain, and the joint distribution gives the transition probability kernel

$$\pi_E(\boldsymbol{x}_{0:k-1}, x_k) = \pi_G\{\Phi^{\leftarrow}\{1 - \exp(-\boldsymbol{x}_{0:k-1})\}, \Phi^{\leftarrow}\{1 - \exp(-1/x_k)\}\},$$

for $(\boldsymbol{x}_{0:k-1}, x_k) \in \mathbb{R}^k \times \mathbb{R}$, where the kernel π_G is the full conditional distribution function of the multivariate normal, given by

$$\pi_G(\boldsymbol{x}_{0:k-1}, x_k) = \Phi\left[q_{kk}^{1/2}\left\{x_k - \sum_{t=0}^{k-1}\left(-\frac{q_{tk}}{q_{kk}}\right)x_t\right\}\right],$$

where Φ is the standard normal distribution function. The condition $\rho_i > 0$, for $i = 1, \ldots, k$, appears restrictive but is made to simplify the presentation. If we worked with standard Laplace marginals, instead of exponential marginals, as say in [13], the presentation would be equally simple for any values $|\rho_i| > 0$, $i = 1, \ldots, k$, of the correlation matrix $\boldsymbol{\Sigma}$.

Max-stable copula: A class of transition probability kernels for asymptotically dependent Markov processes is obtained from the class of multivariate extreme-value distributions [25]. The $(k+1)$-dimensional distribution function of the multivariate extreme-value distribution with exponential margins is given by

$$F(\boldsymbol{x}_{0:k}) = \exp(-V(\boldsymbol{y}_{0:k})), \quad \text{where } \boldsymbol{y}_{0:k} = T(\boldsymbol{x}_{0:k}) := -1/\log(1 - \exp(-\boldsymbol{x}_{0:k})), \quad (30)$$

for $\boldsymbol{x}_{0:k} \in \mathbb{R}_+^{k+1}$, with $V : \mathbb{R}_+^{k+1} \to \mathbb{R}_+$ a -1-homogeneous function, known as the exponent function, given by

$$V(\boldsymbol{y}_{0:k}) = \int_{\Delta^k} \max_{i=0,\ldots,k}(\omega_i/y_i)\, H(\mathrm{d}\boldsymbol{\omega}), \quad (31)$$

where H is termed the spectral measure on Δ^k that has total mass $k+1$ and satisfies the moment constraints $\int_{\Delta^k} \omega_i H(\mathrm{d}\boldsymbol{\omega}) = 1$, for $i = 0, \ldots, k$. Throughout this section, we assume that V has continuous mixed partial derivatives of all orders, which ensures

that a density for F exists [4]. For any $J \subseteq [k]$, we write V_J to denote the higher-order partial derivative $\partial^{|J|} V(\boldsymbol{x}_{0:k}) / \prod_{j \in J} \partial x_j$ and Π_m for the set of partitions of $[m] :=$ $\{0, 1, \dots, m\}$, where $m = 0, \dots, k$. Furthermore, for a vector $\boldsymbol{z} = (\boldsymbol{x}_{0:m}, \boldsymbol{x}_{m+1:k})$, we write $V(\boldsymbol{z}) = V(\boldsymbol{x}_{0:m}, \boldsymbol{x}_{m+1:k})$. For $m = 0, \dots, k-1$ we define $V(\boldsymbol{x}_{0:m}, \infty \cdot \boldsymbol{1}_{k-m}) :=$ $\lim_{\boldsymbol{x}_{m+1:k} \to \infty \cdot \boldsymbol{1}_{k-m}} V(\boldsymbol{x}_{0:m}, \boldsymbol{x}_{m+1:k})$, and for $J \subseteq [m]$, we define $V_J(\boldsymbol{x}_{0:m}, \infty \cdot \boldsymbol{1}_{k-m}) :=$ $\partial^{|J|} V(\boldsymbol{x}_{0:m}, \infty \cdot \boldsymbol{1}_{k-m}) / \prod_{j \in J} \partial x_j$. Stationarity is achieved by requiring that for any set $A \subset \mathbb{Z}$, the distributions of $\{X_i : i \in A\}$ and $\{X_i : i \in B\}$ are identical when the set B is a translation of the set A, i.e., when there exists a unique $\omega \in \mathbb{Z}$ such that $B = \{x + \omega : x \in A\}$. Given the Markov property, stationarity is ensured if V satisfies the property that

$$\lim_{\boldsymbol{x}_{[k] \backslash A} \to \infty} V(\boldsymbol{x}) \Big|_{\boldsymbol{x}_A = \boldsymbol{y}} = \lim_{\boldsymbol{x}_{[k] \backslash B} \to \infty} V(\boldsymbol{x}) \Big|_{\boldsymbol{x}_B = \boldsymbol{y}}, \qquad \boldsymbol{y} \in \mathbb{R}_+^{|A|}, \tag{32}$$

for any set $A \subseteq [k]$, where B is a translation of the set A, with $B \subseteq [k]$. The transition probability kernel induced by the multivariate extreme value copula, in exponential margins, is

$$\pi_E(\boldsymbol{x}_{0:k-1}, x_k) = \frac{\left[\sum_{p \in \Pi_{k-1}} (-1)^{|p|} \prod_{J \in p} V_J(\boldsymbol{y}_{0:k}) \right]}{\left[\sum_{p \in \Pi_{k-1}} (-1)^{|p|} \prod_{J \in p} V_J(\boldsymbol{y}_{0:k-1}, \infty) \right]}$$

$$\times \exp\{V(\boldsymbol{y}_{0:k-1}, \infty) - V(\boldsymbol{y}_{0:k})\}, \tag{33}$$

where $(\boldsymbol{x}_{0:k-1}, x_k) \in \mathbb{R}^k \times \mathbb{R}$ and with $\boldsymbol{y}_{0:k}$ as defined in (30).

Inverted max-stable: The final class of transition kernels is based on the class of inverted max-stable distributions [16, 21]. The specification of this distribution is most elegantly expressed in terms of its $(k+1)$-dimensional survivor function. In exponential margins, this is expressed as

$$\overline{F}(\boldsymbol{x}_{0:k}) = \exp\left(-V(1/\boldsymbol{x}_{0:k})\right), \tag{34}$$

where V denotes an exponent function as defined by (31). To ensure stationarity, V is assumed to satisfy the conditions (32). This distribution gives the transition probability kernel

$$\pi^{\text{inv}}(\boldsymbol{x}_{0:k-1}, x_k) = 1 - \pi_E[-\log\{1 - \exp(-1/\boldsymbol{x}_{0:k-1})\}, -\log\{1 - \exp(-1/x_k)\}], \tag{35}$$

where $(\boldsymbol{x}_{0:k-1}, x_k) \in \mathbb{R}^k \times \mathbb{R}$ and π_E is as given by (33).

5.3. Examples of norming function and hidden tail chains

For a range of examples of kth-order Markov processes, we illustrate how the theory we have developed is applicable, and we derive the forms of the required norming functions and identify the properties of the hidden tail chains. The examples include the full pairwise asymptotically dependent max-stable distribution family, and specifically two subclasses known as the logistic [2] and Hüsler–Reiss [7, 10] dependence structures. They also include two classes of full pairwise asymptotically independent distributions, the Gaussian copula and the inverted max-stable distribution with logistic dependence structure. We also illustrate the sub-asymptotic behaviour of these hidden tail chains in Figure 1, through simulation of the Markov process after a large event. Our proofs that the required the weak convergence of each transition probability kernel satisfies the assumptions of Section 2 are presented in Appendices A.5–A.9, where we implement step by step the strategy we outlined in Section 5.1 to find the required norming functions.
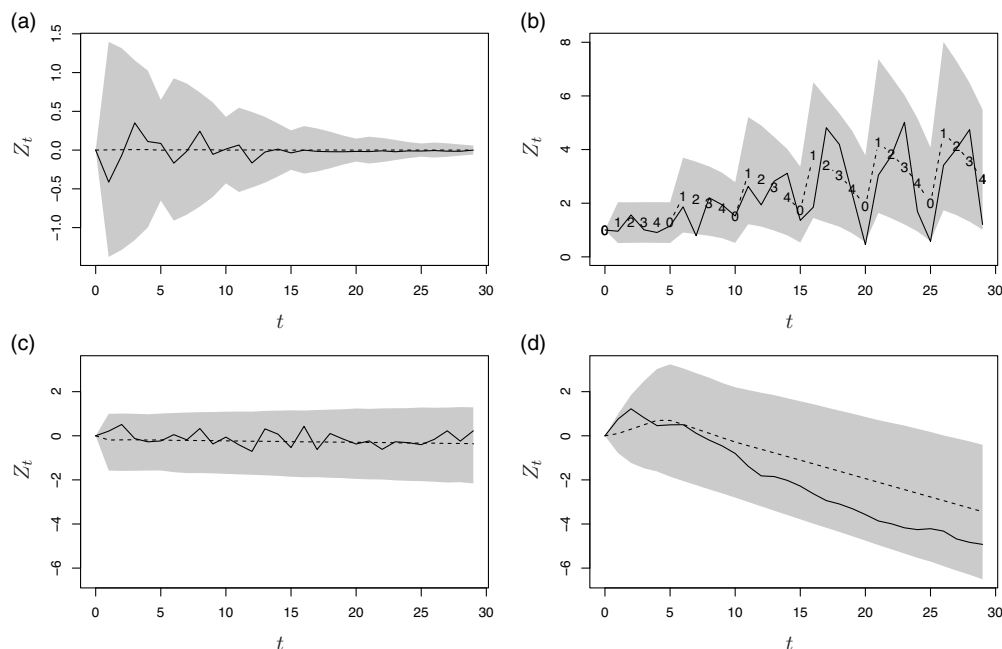
FIGURE 1. Properties for each hidden tail chain for Examples 1–3 with $k = 5$. Presented for each chain are pointwise 2.5% and 97.5% quantiles of the sampling distribution (shaded region), mean of the sampling distribution (dashed line), and one realization from the (hidden) tail chain (solid line). The copula of $X_{0:k}$ used to derive the (hidden) tail chain comes from (a) the standard multivariate Gaussian copula with Toeplitz positive definite covariance matrix Σ generated by the vector (1, 0.70, 0.57, 0.47, 0.39, 0.33), and (b) the inverted logistic with $\alpha = \log(\mathbf{1}_{k+1}^\top \Sigma^{-1} \mathbf{1}_{k+1}) / \log k = 0.27$ (here the value of the function $\mathrm{mod}_k(t)$ is also highlighted on the mean function of the time series with numbers ranging from 0 to 4 for all t), (c) the logistic copula with $\alpha = 0.32$, and (d) the Hüsler–Reiss copula with Toeplitz positive definite covariance matrix generated by the vector (1, 0.9, 0.7, 0.5, 0.3, 0.1). The parameters for all copulas are chosen so that the coefficient of residual tail dependence η [16] and the extremal coefficient θ [2] are equal for the copulas in panels (a) and (b), and for those in panels (c) and (d), respectively.

We also consider an example of a Markov process not covered by the theory we have developed, but for which we can directly derive the norming functions and the hidden tail chain behaviour. This is a second-order stationary Markov chain with a transition probability kernel using a max-stable distribution which permits the possibility of sudden switches from extreme to non-extreme states and vice versa. In this setting, a novel form of normalization of the transition probability kernel is required, which, together with the associated hidden tail chain, carries information about the mechanism that governs the sudden transitions. In Figure 2 we illustrate that the sub-asymptotic properties of this process are captured by our asymptotic results. Although the development of general theory for this type of process is beyond the scope of this paper, this example serves to motivate future extensions of our theory.

Example 1. (Stationary Gaussian autoregressive process—positive dependence.) For this copula, under the notation and conditions described in Section 5.2, [9, Section 8.6] showed that Assumption 2 holds with norming functions $a_i(v) = \rho_i^2 v$, $b_i(v) = v^{1/2}$, that is, $\alpha_i = \rho_i^2$ and $\beta_i = 1/2$, for $i = 1, \ldots, k-1$ and initial limiting distribution $G(z_{1:k-1}) = \Phi_k(z_{1:k-1}; \Sigma_0)$,
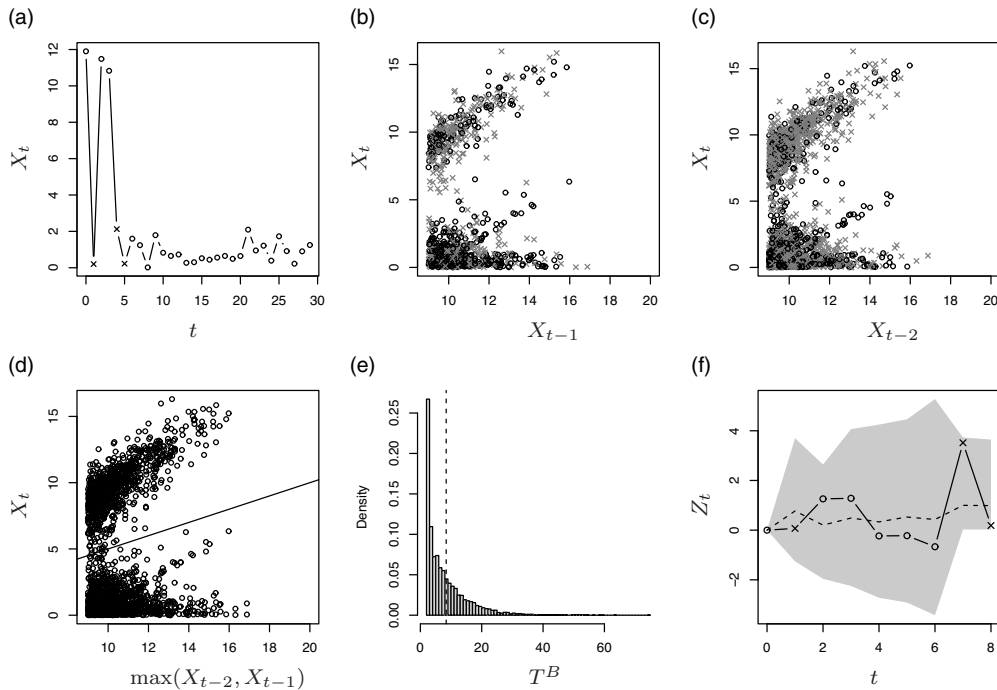
FIGURE 2. (a) Time series plot showing a single realization from the second-order Markov chain with asymmetric logistic dependence (46) initialized from the distribution of $X_0 \mid X_0 > 9$. For this realization, there are three change-points T_1^X, T_2^X, and T_3^X, which are highlighted with crosses. (b) Scatter-plot of states $\{(X_{t-1}, X_t) : X_{t-1} > 9\}$ drawn from 10^3 realizations of the Markov chain initialized from the distribution of $X_0 \mid X_0 > 9$. Points for which $X_{t-2} < 9$ and $X_{t-2} \geq 9$ are highlighted with grey crosses and black circles, respectively. (c) Scatter-plot of consecutive states (X_{t-2}, X_t). Points for which $X_{t-1} < 9$ and $X_{t-1} \geq 9$ are highlighted with grey crosses and black circles, respectively. (d) Scatter-plot of states $\{(\max(X_{t-2}, X_{t-1}), X_t) : \max(X_{t-2}, X_{t-1}) > 9\}$ and line $X_t = c \max(X_{t-2}, X_{t-1})$ with $c = \frac{1}{2}$ superposed. (e) Histogram of termination time T^B obtained from 10^4 realizations from the hidden tail chain. The Monte Carlo estimate of the mean of the distribution is 8.42 and shown with a dashed vertical line. (f) Pointwise 2.5% and 97.5% quantiles of the sampling distribution (shaded region), mean of the sampling distribution (dashed line), and one realization from the hidden tail chain (solid line), conditioned on $T^B = 8$. The value of the latent Bernoulli process B_t is highlighted with a cross when $B_t = 0$ and with a circle when $B_t = 1$. For all plots presented, $\theta_0 = \theta_1 = \theta_2 = \theta_{01} = \theta_{02} = 0.3$, $\theta_{012} = 0.1$, and $v_{01} = v_{02} = v_{012} = 0.5$.

where $z_{1:k-1} \in \mathbb{R}^{k-1}$ and $\Phi_k(\,\cdot\,; \Sigma_0)$ denotes the cumulative distribution function of the k-dimensional multivariate normal distribution with a zero mean vector and covariance matrix $\Sigma_0 = (2\rho_i \rho_j(\rho_{|j-i|} - \rho_i \rho_j))_{1 \leq i,j \leq k-1}$. Appendix A.5 shows that Assumption 3 holds with norming functionals

$$a(\boldsymbol{u}) = \left(\sum_{i=1}^{k} \phi_i u_i^{1/2} \right)^2, \qquad b(\boldsymbol{u}) = a(\boldsymbol{u})^{1/2}, \qquad \boldsymbol{u} = (u_1, \dots, u_k) \in \mathbb{R}_+^k, \qquad (36)$$

where $\phi_i = -q_{k-i,k}/q_{kk}$, $i = 1, \ldots, k$, denote the first k partial autocorrelation coefficients of the stationary Gaussian process (on Gaussian margins), and the transition probability kernel of the renormalized Markov chain converges weakly to the distribution

$$K_A(x) = \Phi\left((q_{kk}/2)^{1/2} x\right), \qquad x \in \mathbb{R}. \tag{37}$$

Corollary 2 asserts that a suitable location normalization after $t \geq k$ steps has $\alpha_t = a(\boldsymbol{\alpha}_{t-k\,:\,t-1}) = \rho_t^2$ and $\beta_t = 1/2$. This equivalence arises from the a function in (36) and the Yule–Walker equations for stationary Gaussian autoregressive processes, that is, $\rho_t = \sum_{i=1}^k \phi_i \rho_{t-i}$, for $t \geq k$. As all stationary Gaussian finite-order Markov chains have that $\rho_t \to 0$ geometrically as $t \to \infty$, it follows that α_t does likewise.

Now consider the hidden tail chain. The gradient vector of a is

$$\nabla a(\boldsymbol{u}) = \left\{ (\phi_j u_j)^{-1/2} \sum_{i=1}^k \phi_i u_i^{1/2} : j = 1, \ldots, k \right\}.$$

Thus, based on the Yule–Walker equations, we have

$$\nabla a(\boldsymbol{\alpha}_{t-k\,:\,t-1})^\top = \rho_t(\phi_1/\rho_{t-1}, \ldots, \phi_k/\rho_{t-k}).$$

Also, $b(\boldsymbol{\alpha}_{t-k\,:\,t-1}) = a(\boldsymbol{\alpha}_{t-k\,:\,t-1})^{1/2} = \rho_t$. This leads to the scaled autoregressive hidden tail chain

$$Z_t = \rho_t \sum_{i=1}^k \frac{\phi_i}{\rho_{t-i}} Z_{t-i} + \rho_t \, \varepsilon_t, \qquad t \geq k, \tag{38}$$

and $\{\varepsilon_t\}_{t=k}^\infty$ is a sequence of i.i.d. random variables with distribution K_A given by (37). The hidden tail chain is a non-stationary kth-order autoregressive Gaussian process with zero mean and autocovariance function $\mathrm{cov}(Z_{t-s}, Z_t) = (2\rho_{t-s}\,\rho_t(\rho_s - \rho_{t-s}\,\rho_t))$ when $t \neq s$. The variance of the process satisfies $\mathrm{var}(Z_t) = \mathcal{O}(\rho_t^2)$ as $t \to \infty$; hence the hidden tail chain degenerates to 0 in the limit as $t \to \infty$. This long-term degenerative behaviour is illustrated in panel (*a*) of Figure 1.

Remark 5. The location functional a in (36) can be written in the form (22) with $c = \sum_{i=1}^k \phi_i$ and $\gamma_i = \phi_i^{2/3} / \sum_j \phi_j^{2/3}$ for $i = 1, \ldots, k$.

Example 2. (Inverted max-stable copula with logistic dependence.) Consider a stationary kth-order Markov chain with a $(k+1)$-dimensional survivor function (34) and exponent function of logistic type given by

$$V(\boldsymbol{y}_{0\,:\,k}) = \|\boldsymbol{y}_{0\,:\,k}^{-1/\alpha}\|^\alpha, \qquad \boldsymbol{y}_{0\,:\,k} \in \mathbb{R}_+^d, \tag{39}$$

where $\alpha \in (0, 1)$, which gives a stationary process as V is an exchangeable function. [9, Section 8.5] showed that Assumption 4 holds with $b_i(v) = v^{1-\alpha}$, that is, $\beta_i = 1 - \alpha$ for $i = 1 \ldots, k-1$, and limiting initial conditional distribution $G(\boldsymbol{z}) = \prod_{i=1}^{k-1} \left\{ 1 - \exp\left(-\alpha z_i^{1/\alpha}\right) \right\}$, $\boldsymbol{z} \in (0, \infty)^{k-1}$. Appendix A.6 shows that Assumption 5 holds with normalizing functionals

$$a(\boldsymbol{u}) = 0, \qquad b(\boldsymbol{u}) = \|\boldsymbol{u}^{1/\alpha}\|^{\alpha\,(1-\alpha)}, \qquad \boldsymbol{u} = (u_1, \ldots, u_k) \in \mathbb{R}_+^k, \tag{40}$$

and the transition probability kernel of the renormalized Markov chain converges weakly to the distribution

$$K_B(x) = 1 - \exp\left(-\alpha \, x^{1/\alpha}\right), \qquad x \in \mathbb{R}_+,$$

as $u \to \infty$. Corollary 3 asserts that a suitable normalization after $t \geq k$ steps is $a_t(v) = 0$, $\log b_t(v) = ((1-\alpha)^{1+\lfloor (t-1)/k \rfloor}) \log v$, which leads to the scaled random walk hidden tail chain

$$Z_t = \begin{cases} \|(Z_{t-k}, \mathbf{0}_{k-1})^{1/\alpha}\|^{\alpha(1-\alpha)} \varepsilon_t & \text{when } \mathrm{mod}_k(t) = 0, \\ \|\mathbf{Z}_{t-k:t-1}^{1/\alpha}\|^{\alpha(1-\alpha)} \varepsilon_t & \text{when } \mathrm{mod}_k(t) = 1, \\ \|(\mathbf{Z}_{t-k:t-j}^{1/\alpha}, \mathbf{0}_{j-1})\|^{\alpha(1-\alpha)} \varepsilon_t & \text{when } \mathrm{mod}_k(t) = j \in \{2, \ldots, k-1\}, \end{cases}$$

where $\{\varepsilon_t\}_{t=k}^\infty$ is a sequence of i.i.d. random variables with distribution K_B.

This hidden tail chain is a non-stationary process; specifically, after a logarithmic transformation, it is a non-stationary nonlinear kth-order autoregressive process. The first element of the process is $Z_0 = 1$ a.s., and the next $k-1$ elements of the process $\mathbf{Z}_{1:k-1}$ are i.i.d. positive random variables with distribution function $K_B(x)$, $x > 0$. Subsequent elements Z_t, for $t \geq k$, have distributions that vary in both mean and variance. For b given by (40) and any $\boldsymbol{x}_{1:k} \in \mathbb{R}_+^k$, we see that $b(\boldsymbol{x}_{1:k}) > b(\boldsymbol{x}_{2:k}, 0) > \cdots > b(x_k, \mathbf{0}_{k-1})$. This leads to an oscillating behaviour which is illustrated in panel (*b*) of Figure 1. In particular, both the mean and variance of the hidden tail chain can be seen to decrease in a segment of k consecutive time points $(s, s+1, \ldots, s+k-1)$ for any $s > k$ such that $\mathrm{mod}_k(s) = 1$.

Example 3. (Multivariate extreme value copula—all mass on interior of simplex.) [9, Section 8.4] showed that if the spectral measure H in (31) places no mass on the boundary of Δ^k, then Assumption 2 holds for the distribution (30) with norming functions $a_i(v) = v$, $b_i(v) = 1$, i.e. $\alpha_i = 1$, $\beta_i = 0$, for $i = 1, \ldots, k-1$, and limiting distribution

$$G(z_{1:k-1}) = -V_0[\exp\{(0, z_{1:k-1})\}, \infty], \qquad z_{1:k-1} \in \mathbb{R}^{k-1}. \tag{41}$$

In this set-up we trivially have that the functional $b \equiv 1$. For determining the functional a, Appendix A.7 shows that for any choice of a satisfying the condition (19), the limit integral (9) is $\int_{\mathbb{R}} f(x) K(dx; z_{1:k-1})$, with $z_{1:k-1} \in \mathbb{R}^{k-1}$, which is not in the form required by Assumption A_2, as the K term is not independent of $z_{1:k-1}$ as is K_A in the statement of the assumption. Specifically, we find that K has the following form:

$$K(x; z_{1:k-1}) = \frac{V_{0:k-1}[\exp(z_{0:k-1}), \exp(a(z_{0:k-1}) + x)]}{V_{0:k-1}[\exp(z_{0:k-1}), \infty]}, \qquad x \in \mathbb{R}. \tag{42}$$

Without additional assumptions on the max-stable copula, i.e. of the exponent measure V, it seems impossible to find the form of the location functional a to make K, in (42), independent of $z_{1:k-1}$, or to know whether such a functional even exists. Without Assumption A_2 we cannot use Corollary 1 to find the norming functions for $t \geq k$. To get around this problem, here we can make an additional assumption about V, which ensures that K is independent of $z_{1:k-1}$ and thus that Assumption A_2 holds. The new assumption exploits the property that both $V_{0:k-1}(\cdot, \infty)$ and $V_{0:k-1}(\cdot)$ are $-(k+1)$-homogeneous functions, so the map

$$\mathbb{R}_+^{k+1} \ni \boldsymbol{y}_{0:k} \mapsto V_{0:k-1}(\boldsymbol{y}_{0:k-1}, y_k)/V_{0:k-1}(\boldsymbol{y}_{0:k-1}, \infty) \in \mathbb{R}_+ \tag{43}$$

is 0-homogeneous, and this latter property restricts the possible forms the function $V_{0:k-1}$ can take. One such simple form is given by Property K_1, which holds for a wide variety of parametric models for the exponent measure, including the logistic dependence and Hüsler–Reiss dependence structures discussed below.

Assumption 8. *There exist a continuous function $a_P : \mathbb{R}_+^k \to \mathbb{R}_+$ which is 1-homogeneous, and a non-degenerate distribution function K_P on $\mathbb{R}_+$, such that the following hold:*

(i) $K_P^{\leftarrow}(p^\star) = 1$ *for some* $p^\star \in (0, 1)$, *where* $K_P^{\leftarrow}(p) = \inf\{x \in \mathbb{R}_+ : K_P(x) > p\}$;

(ii) $V_{0:k-1}(\boldsymbol{y}_{0:k-1}, y_k) = V_{0:k-1}(\boldsymbol{y}_{0:k-1}, \infty) K_P\{y_k/a_P(\boldsymbol{y}_{0:k-1})\}$, *for all* $\boldsymbol{y}_{0:k} \in \mathbb{R}_+^k$.

When Assumption 8 holds, additional information about the location functional a and the limit distribution K_A of Corollary 1 can be given. This is established in Proposition 3 below.

Proposition 3. *Suppose that for a max-stable distribution with exponent function V, Assumption 8 holds. Let $a(\boldsymbol{x}_{0:k-1}) = \log[a_P\{\exp(\boldsymbol{x}_{0:k-1})\}]$, $\boldsymbol{x}_{0:k-1} \in \mathbb{R}^k$, and assume that there exists a right-inverse $\mathbb{R} \times \mathbb{R}_+^k \ni (q, \boldsymbol{y}_{0:k-1}) \mapsto V_{0:k-1}^{\leftarrow}(q;\boldsymbol{y}_{0:k-1}) \in \mathbb{R}_+$ such that $V_{0:k-1}\{\boldsymbol{y}_{0:k-1}, V_{0:k-1}^{\leftarrow}(q;\boldsymbol{y}_{0:k-1})\} = q$ for all q and $\boldsymbol{y}_{0:k-1}$ in the domain of $V_{0:k-1}^{\leftarrow}$. Then the following hold:*

(i) The location functional a satisfies the property (19), and for all $\boldsymbol{x}_{0:k-1} \in \mathbb{R}^k$,

$$a(\boldsymbol{x}_{0:k-1}) = \log V_{0:k-1}^{\leftarrow}\left\{p^\star V_{0:k-1}\left(e^{\boldsymbol{x}_{0:k-1}}, \infty\right); e^{\boldsymbol{x}_{0:k-1}}\right\}. \qquad (44)$$

(ii) Assumption 3 holds with normalizing functionals a as given by (44) and $b \equiv 1$, and $K_A(x) = K_P(e^x)$, $x \in \mathbb{R}$.
(iii) For all $x \in \mathbb{R}$,

$$K_P(e^x) = \frac{V_{0:k-1}\left\{\exp\left(z_{0:k-1}^\star\right), \exp(x)\right\}}{V_{0:k-1}\left\{\exp\left(z_{0:k-1}^\star\right), \infty\right\}},$$

where $z_{0:k-1}^\star$ satisfies $a_P\left(z_{0:k-1}^\star\right) = 1$.

So, for a Markov chain with max-stable copula under the assumptions of Proposition 3, we now have new general results from Corollary 1 on the hidden tail chain, which here is identical to the tail chain, and a and K_A have been derived. To help interpret these results, given the generality of the terms for a and K_A in Proposition 3, we now investigate two well-established multivariate extreme-value distribution dependence models which satisfy the assumptions of Section 2 and Proposition 3.

Logistic dependence: The exponent function of the $(k+1)$-dimensional max-stable distribution with logistic dependence is given in (39), where $\alpha \in (0, 1)$ controls the strength of dependence, with stronger dependence as α decreases. The case $\alpha = 1$ is excluded as that corresponds to independence. [9] show that the initial limiting distribution (41) is

$$G(z_{1:k-1}) = \{1 + \|\exp(-z_{1:k-1}/\alpha)\|\}^{\alpha-1}, \qquad z_{1:k-1} \in \mathbb{R}^{k-1},$$

in addition to having $a_i(v) = v$, $b_i(v) = 1$ for $i = 1, \ldots, k-1$. Appendix A.8 shows that Assumption 3 holds with normalizing functionals

$$a(\boldsymbol{u}) = -\alpha \log(\|\exp(-\boldsymbol{u}/\alpha)\|), \qquad b(\boldsymbol{u}) = 1, \qquad \boldsymbol{u} = (u_1, \ldots, u_k) \in \mathbb{R}_+^k,$$

and the transition probability kernel of the renormalized Markov chain converges weakly to the distribution

$$K_A(x) = \{1 + \exp(-x/\alpha)\}^{\alpha-k}, \qquad x \in \mathbb{R}.$$

Corollary 1 then asserts that the suitable normalization is $a_t(v) = v$, $b_t(v) = 1$ for $t \geq k$, which leads to the hidden tail chain

$$Z_t = -\alpha \, \log \|\exp\left(-\mathbf{Z}_{t-k:t-1}/\alpha\right)\| + \varepsilon_t, \qquad t = k, k+1, \ldots, \tag{45}$$

where $\{\varepsilon_t\}_{t=k}^{\infty}$ is a sequence of i.i.d. random variables with distribution K_A. Note that the hidden tail chain can also be expressed as

$$Z_t = Z_{t-k} - \alpha \, \log \|\exp\{-(0, Z_{t-k}\mathbf{1}_{k-1} - \mathbf{Z}_{t-k+1:t-1})/\alpha\}\| + \varepsilon_t, \qquad t = k, k+1, \ldots.$$

Here the hidden tail chain is identical to the tail chain, as $a_t(x) = x$ and $b_t(x) = 1$ for all $t = 1, 2, \ldots$. When $k = 1$, the tail chain can be seen to reduce to the random-walk results of [22, 32], but when $k > 1$, the tail chain behaves like a random walk with an extra additive factor which depends in a nonlinear way on the 'profile' $Z_{t-k}\mathbf{1}_{k-1} - \mathbf{Z}_{t-k+1:t-1}$ of the previous $k-1$ values.

Hüsler–Reiss dependence: The exponent function of the $(k+1)$-dimensional max-stable distribution with Hüsler–Reiss dependence is

$$V(\mathbf{y}_{0:k}) = \sum_{i=0}^{k} \frac{1}{y_i} \, \Phi_k\left[\left\{ \log\left(y_j/y_i\right) + \sigma^2 - \sigma_{ij}\right\}_{j\neq i}; \boldsymbol{\Sigma}^{(i)}\right], \qquad \mathbf{y}_{0:k} \in \mathbb{R}_+^{k+1},$$

where $\Phi_k(\,\cdot\,; \boldsymbol{\Sigma}^{(i)})$ denotes the multivariate normal distribution function with mean zero and covariance matrix $\boldsymbol{\Sigma}^{(i)} = \boldsymbol{T}_i \, \boldsymbol{\Sigma} \, \boldsymbol{T}_i^{\top}$, for $\boldsymbol{\Sigma} = (\sigma_{ij})_{i,j=0}^{k}$ a positive definite Toeplitz covariance matrix with common diagonal elements $\sigma_{ii} = \sigma^2$, and $\boldsymbol{T}_i$ is a $k \times (k+1)$ matrix with the $(i+1)$th column having -1 for each entry and the other columns being the k standard orthonormal basis vectors of $\mathbb{R}^k$, that is,

$$\boldsymbol{T}_i = \begin{pmatrix} 1 & 0 & \cdots & 0 & -1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & -1 & 0 & \cdots & 1 \end{pmatrix}, \qquad i = 0, \ldots, k.$$

The matrix $\boldsymbol{\Sigma}$ controls the strength of dependence, with larger values for σ_{ij} indicating stronger dependence between the associated elements of the random vector. The initial limiting distribution (41) is $G(\boldsymbol{x}) = \Phi_{k-1}[\boldsymbol{x} - \{-\text{diag}(\boldsymbol{\Sigma}^{(0)})/2\}; \boldsymbol{\Sigma}^{(0)}]$, $\boldsymbol{x} \in \mathbb{R}^{k-1}$, in addition to having $a_i(v) = v$, $b_i(v) = 1$ for $i = 1, \ldots, k-1$ [7]. Appendix A.9 shows that Assumption 3 holds with normalizing functionals

$$a(\boldsymbol{u}) = -\tau \boldsymbol{K}_{01}^{\top} \boldsymbol{C} \boldsymbol{K}_{10} \cdot \boldsymbol{u}, \qquad b(\boldsymbol{u}) = 1, \qquad \boldsymbol{u} = (u_1, \ldots, u_k) \in \mathbb{R}_+^k,$$

where the quantities τ, $\boldsymbol{q}$, $\boldsymbol{C}$, $\boldsymbol{K}_{10}$, and $\boldsymbol{K}_{01}$ are constants with respect to $\boldsymbol{u}$ but depend on the parameters of Σ; they are defined in Appendix A.9. The transition probability kernel of the renormalized Markov chain converges weakly to the distribution

$$K_A(x) = \Phi\left\{(x/\tau) + \left(\boldsymbol{K}_{01}^{\top} \boldsymbol{\Sigma}^{-1} \mathbf{1}_{k+1}^{\top}\right)/\left(\mathbf{1}_{k+1}^{\top} \boldsymbol{q}\right)\right\}, \qquad x \in \mathbb{R}.$$

Corollary 1 asserts that a suitable normalization is $a_t(v) = v$, $b_t(v) = 1$ for $t \geq k$, which leads to the hidden tail chain (identical to the tail chain)

$$Z_t = -\tau \boldsymbol{K}_{01}^{\top} \boldsymbol{C} \boldsymbol{K}_{10} \cdot \boldsymbol{Z}_{t-k:t-1} + \varepsilon_t, \qquad t = k, k+1, \ldots,$$

where $\{\varepsilon_t\}_{t=k}^{\infty}$ is a sequence of i.i.d. random variables with distribution K_A. Note that the tail chain can also be expressed as

$$Z_t = Z_{t-k} + \tau K_{01}^{\top} C K_{10} \cdot (0, Z_{t-k} \mathbf{1}_{k-1} - Z_{t-k+1:t-1}) + \varepsilon_t, \qquad t = k, k+1, \ldots,$$

which shows, similarly to the case of the logistic copula, that the tail chain behaves like a random walk with an extra factor which depends linearly on the 'profile' $Z_{t-k} \mathbf{1}_{k-1} - Z_{t-k+1:t-1}$ of the previous $k-1$ values; in this respect it differs from the previous example.

Panels (c) and (d) in Figure 1 show almost linear behaviour for two special cases of the tail chains presented. Although the copulas used to derive both (hidden) tail chains have the same extremal coefficient see [28], ensuring that the core level of extremal dependence is common to both, the decay rates of the two processes are markedly different. This shows that the type of drift function and distribution for the innovation term ε_t affect the characteristics of the transition from an extreme state to the main body of the process.

Example 4. (Multivariate extreme value copula with asymmetric logistic structure [33].) This is a second-order Markov process for which Assumptions 2 and 3 fail to hold, and it has a more complicated structure than we have covered so far in studying weak convergence on $\mathbb{R}^k$. In this example, the weak convergences in Assumptions 2 and 3 no longer hold on $\mathbb{R}^{k-1}$ and $\mathbb{R}$ (cf. Remark 1), but on $\overline{\mathbb{R}}^{k-1}$ and $\overline{\mathbb{R}}$, respectively. The example is a special case of a stationary Markov chain with transition probability kernel (33) and exponent function given by

$$\begin{aligned}
V(x_0, x_1, x_2) = {} & \theta_0 x_0^{-1} + \theta_1 x_1^{-1} + \theta_2 x_2^{-1} \\
& + \theta_{01} \left\{ \left(x_0^{-1/\nu_{01}} + x_1^{-1/\nu_{01}} \right)^{\nu_{01}} + \left(x_1^{-1/\nu_{01}} + x_2^{-1/\nu_{01}} \right)^{\nu_{01}} \right\} \\
& + \theta_{01} \left(x_0^{-1/\nu_{02}} + x_2^{-1/\nu_{02}} \right)^{\nu_{02}} \\
& + \theta_{012} \left(x_0^{-1/\nu_{012}} + + x_1^{-1/\nu_{012}} + x_2^{-1/\nu_{012}} \right)^{1/\nu_{012}},
\end{aligned} \tag{46}$$

where $\nu_A \in (0, 1)$ for any $A \in 2^{\{0,1,2\}} \setminus \emptyset$, and

$$\theta_0 + \theta_{01} + \theta_{02} + \theta_{012} = 1, \quad \theta_1 + 2\theta_{01} + \theta_{012} = 1, \quad \theta_2 + \theta_{01} + \theta_{02} + \theta_{012} = 1$$

with $\theta_0, \theta_1, \theta_2, \theta_{01}, \theta_{02}, \theta_{012} > 0$.

Although this distribution does not satisfy the assumptions of Section 2, the strategy that is implemented to find the normalizing functions and hidden tail chain is similar to the strategy presented in Section 5.1. In particular, the initial distribution of the Markov process is $F_{01}(x_0, x_1) = F_{012}(x_0, x_1, \infty) = \exp\{-V(y_0, y_1, \infty)\}$, with (y_0, y_1) defined in (30). It can be seen that the transition probability kernel $\pi(x_0, x_1) = -y_0^2 V_0(y_0, y_1, \infty) \exp(y_0^{-1} - V(y_0, y_1, \infty))$ associated with the conditional distribution of $X_1 \mid X_0$ converges with two distinct normalizations, that is, $\pi(v, dx) \xrightarrow{w} K_0(dx)$ and $\pi(v, v+dx) \xrightarrow{w} K_1(dx)$ as $v \to \infty$, to the distributions $K_0 = (\theta_0 + \theta_{02}) F_E + (\theta_{01} + \theta_{012}) \delta_{+\infty}$ and $K_1 = (\theta_0 + \theta_{02}) \delta_{-\infty} + \theta_{01} G_{01} + \theta_{012} G_{012}$, respectively, where $F_E(x) = (1 - \exp(-x))_+$, $G_A(x) = (1 + \exp(-x/\nu_A))^{\nu_A - 1}$, and δ_x is a point mass at $x \in [-\infty, \infty]$ [cf. Example 5 in 20]. The distributions K_0 and K_1 have their entire mass on $(0, \infty]$ and $[-\infty, \infty)$, respectively. In the first and second normalizations, a mass of size $(1 - \theta_{01} - \theta_{012})$ escapes to $+\infty$ and a mass of size $(\theta_0 + \theta_{02})$ escapes to $-\infty$, respectively. As explained by [20], the reason for this behaviour is that the separate

normalizations are related to two different modes of the conditional distribution of $X_1 \mid X_0$. This phenomenon also manifests in the conditional distribution of $X_2 \mid \{X_0, X_1\}$, which is given by

$$\pi(\mathbf{x}_{0:1},\ x_2) = \frac{(V_0 V_1 - V_{01})(\mathbf{y}_{0:2})}{(V_0 V_1 - V_{01})(\mathbf{y}_{0:1}, \infty)} \ \exp{(V(\mathbf{y}_{0:1}, \infty) - V(\mathbf{y}_{0:2}))},$$

where $g(f_1, f_2, f_3)(x) := g(f_1(x), f_2(x), f_3(x))$ for maps g and f_i, $i = 1, 2, 3$. Here the problem is more complex, with this transition probability kernel converging with $2\,(2^k - 1) = 6$ distinct normalizations. Letting

$$
\begin{aligned}
a_{11,1}(v_1, v_2) &= -v_{012} \log\{\exp{(-v_1/v_{012})} + \exp{(-v_2/v_{012})}\}, \\
a_{10,1}(v_1, v_2) &= v_1,\ a_{01,1}(v_1, v_2) = v_2, \\
a_{11,0}(v_1, v_2) &= a_{01,0}(v_1, v_2) = a_{10,0}(v_1, v_2) = 0,
\end{aligned}
\tag{47}
$$

one can show that for $(x_0, x_1) \in \mathbb{R}^2$ and as $v \to \infty$,

$$
\begin{aligned}
\pi((v + x_0, v + x_1), a_{11,1}(v + x_0, v + x_1) + dy) &\xrightarrow{\text{w}} K_{\{1,1\},\{1\}}(dy; x_0, x_1) &&\text{on } [-\infty, \infty), \\
\pi((v + x_0, v + x_1), a_{11,0}(v + x_0, v + x_1) + dy) &\xrightarrow{\text{w}} K_{\{1,1\},\{0\}}(dy; x_0, x_1) &&\text{on } (0, \infty], \\
\pi((v + x_0, x_1), a_{10,1}(v + x_0, x_1) + dy) &\xrightarrow{\text{w}} K_{\{1,0\},\{1\}}(dy; x_0, x_1) &&\text{on } [-\infty, \infty), \\
\pi((v + x_0, x_1), a_{10,0}(v + x_0, x_1) + dy) &\xrightarrow{\text{w}} K_{\{1,0\},\{0\}}(dy; x_0, x_1) &&\text{on } (0, \infty], \\
\pi((x_0, v + x_1), a_{01,1}(x_0, v + x_1) + dy) &\xrightarrow{\text{w}} K_{\{0,0\},\{1\}}(dy; x_0, x_1) &&\text{on } [-\infty, \infty), \\
\pi((x_0, v + x_1), a_{01,0}(x_0, v + x_1) + dy) &\xrightarrow{\text{w}} K_{\{0,1\},\{0\}}(dy; x_0, x_1) &&\text{on } (0, \infty],
\end{aligned}
$$

where the limiting measures are given by

$$K_{A,\{1\}} = m_A\,\delta_{-\infty} + (1 - m_A)\,G_{A,\{1\}} \text{ and } K_{A,\{0\}} = m_A\,G_{A,\{0\}} + (1 - m_A)\,\delta_\infty,$$

for $A \in \{0, 1\}^2 \backslash \{0, 0\}$, with

$$
m_A(x_0, x_1) =
\begin{cases}
\left\{ 1 + \left(\frac{\kappa_{012}}{\kappa_{01}}\right) e^{\lambda(x_0 + x_1)}\, W_{v_{01}-1}(e^{x_0}, e^{x_1}\,;\, v_{01}) / W_{v_{012}-1}(e^{x_0}, e^{x_1}\,;\, v_{012}) \right\}^{-1}, \\
\theta_0/(\theta_0 + \theta_{02}), \\
\theta_1/(\theta_1 + \theta_{01}),
\end{cases}
$$

for $A = \{1, 1\}$, $A = \{1, 0\}$, and $A = \{0, 1\}$, respectively, and

$$
\begin{aligned}
G_{\{1,1\},\{1\}}(y\,;\, x_0, x_1) &= W_2\{1, \exp{(y)}\,;\, v_{012}\}, \\
G_{\{1,1\},\{0\}}(y\,;\, x_0, x_1) &= F_E(y), \\
G_{\{1,0\},\{1\}}(y\,;\, x_0, x_1) &= W_1\{1, \exp{(y)}\,;\, v_{02}\}, \\
G_{\{1,0\},\{0\}}(y\,;\, x_0, x_1) &= [\theta_1 + \theta_{01}\{1 + W_1(1, T(y)/T(x_1); v_{01})\} \\
&\qquad\qquad + \theta_{012} W_1(1, T(y)/T(x_1)\,;\, v_{012})] g_{10}(x, y), \\
G_{\{0,1\},\{1\}}(y\,;\, x_0, x_1) &= W_1\{1, \exp{(y)}\,;\, v_{01}\}, \\
G_{\{0,1\},\{0\}}(y\,;\, x_0, x_1) &= [\theta_0 + \theta_{01} + \theta_{02} W_1(1, T(y)/T(x_0)\,;\, v_{02}) \\
&\qquad\qquad + \theta_{012} W_1(1, T(y)/T(x_0)\,;\, v_{012})] g_{01}(x, y).
\end{aligned}
$$

Here, the function T is defined in (30), $\kappa_A = \theta_A (\nu_A - 1)/\nu_A$, $W_p(x, y; \nu) = (x^{-1/\nu} + y^{-1/\nu})^{\nu-p}$, with $x, y > 0$, $p \in \mathbb{R}$, $\nu \in (0, 1)$,

$$\log g_{10}(x, y) = V(\infty, T(x), \infty) - V(\infty, T(x), T(y)), \quad \text{and}$$
$$\log g_{01}(x, y) = V(T(x), \infty, \infty) - V(T(x), \infty, T(y)).$$

To help explain the necessity for requiring the normalizing functionals (47) to describe the evolution of an extreme episode after witnessing an extreme event in this second-order Markov process, it is useful to consider the behaviour of the spectral measure H, defined in Equation (31), for the initial distribution F_{012} of this process. Here, the spectral measure H places a mass of size $|A| \theta_A$ on each subface $A \in \mathscr{P}([2])$ of Δ^2 [4], which implies that different subsets of the variables (X_{t-2}, X_{t-1}, X_t) can take their largest values simultaneously; see for example [30]. Hence, if the Markov process is in an extreme episode at time $t - 1$, $t \geq 3$, then it follows that there are four possibilities for the states (X_{t-2}, X_{t-1}); that is, either the variables X_{t-2} and X_{t-1} are simultaneously extreme or just one of them is. Consequently, there are two possibilities for the state of the process at time t; that is, the variable X_t can be either extreme or not, and this is demonstrated by bimodality in the transition probability kernel under all four distinct possibilities for the states X_{t-2} and X_{t-1}. In total, this gives rise to six distinct possibilities which necessitate an 'event-specific' normalizing functional to guarantee the weak convergence of the transition probability kernel. This justifies the labelling of the functionals in (47), where the label (A,b) appearing in the subscript, with $A \in \{0, 1\}^2$ and $b \in \{0, 1\}$, indicates transitioning from one of four possible configurations (A) at times $t - 2$ and $t - 1$ into two possible configurations (b) at time t—with 1 indicating that the state is extreme and 0 otherwise. The case where the Markov process is in an extreme episode at time $t - 1$, for $t = 2$, is handled similarly. There we note that X_0 is already extreme, by virtue of the conditioning, and hence there are two possibilities for X_0 and X_1; that is, either X_1 is extreme or it is not.

Although complex, these modes can be identified by any line determined by the loci of points ($\max(x_{t-2}, x_{t-1})$, $\zeta \max(x_{t-2}, x_{t-1})$), where $x_{t-2}, x_{t-1} \in \mathbb{R}$, for some $\zeta \in (0, 1)$, in the distribution of $X_t \mid \max\{X_{t-2}, X_{t-1}\} > v$; see panels (b), (c), and (d) of Figure 2, where v is taken equal to 9. This facilitates accounting for the identification of the normalizing functionals by introducing the stopping times, $T_0^X = 0$ a.s., and

$$T_j^X = \inf \left\{ t \in \left(T_{j-1}^X, T^X \right] : X_t \leq \zeta \max(X_{t-2}, X_{t-1}) \right\}, \qquad j \geq 1,$$

where

$$T^X = \inf\{ t \geq 2 : X_{t-1} \leq \zeta \max(X_{t-3}, X_{t-2}), X_t \leq \zeta \max(X_{t-2}, X_{t-1}) \},$$

subject to the convention $X_{-s} = 0$ for $s \in \mathbb{N} \setminus \{0\}$; that is, T_j^X, with $j \geq 1$, is the jth time that ζ multiplied by the maximum of the previous two states is not exceeded after time 0, and the termination time T^X is the first time after time 0 where two consecutive states did not exceed ζ times the maximum of their respective two previous states. Define

$$a_1(v) = \begin{cases} v, & T_1^X > 1, \\ 0, & T_1^X = 1, \end{cases} \quad \text{and} \quad b_t(v) = 1 \quad \text{for all } t \geq 1.$$

Then, for $t \in \left(T_{j-1}^X, T_j^X \right]$, letting

$$a_t(v_1, v_2) = \begin{cases} a_{11,1}(v_1, v_2) & \text{if } t \neq T_j^X, t - 1 \neq T_{j-1}^X, t - 2 \neq T_{j-1}^X, \\ a_{11,0}(v_1, v_2) & \text{if } t = T_j^X, t - 1 \neq T_{j-1}^X, t - 2 \neq T_{j-1}^X, \\ a_{10,1}(v_1, v_2) & \text{if } t \neq T_j^X, t - 1 = T_{j-1}^X, t - 2 \neq T_{j-1}^X, \\ a_{10,0}(v_1, v_2) & \text{if } t = T_j^X, t - 1 = T_{j-1}^X, t - 2 \neq T_{j-1}^X, \\ a_{01,1}(v_1, v_2) & \text{if } t \neq T_j^X, t - 1 \neq T_{j-1}^X, t - 2 = T_{j-1}^X, \\ a_{01,0}(v_1, v_2) & \text{if } t = T_j^X, t - 1 \neq T_{j-1}^X, t - 2 = T_{j-1}^X \end{cases}$$

yields the hidden tail chain of this process. Specifically, let $\{B_t : t = 0, 1, \ldots\}$ be a sequence of latent Bernoulli random variables. Define the hitting times $T_j^B = \inf\{T_{j-1}^B < t \leq T^B : B_t = 0\}$ with $T_0^B = 0$ a.s. and $T^B = \inf\{t \geq 2 : B_{t-1} = 0, B_t = 0\}$. Then the hidden tail chain process $\{Z_t\}$ together with the latent Bernoulli process $\{B_t\}$ forms a second-order Markov process with initial distribution $(B_0, Z_0) = (1, 0)$ a.s., $B_1 \sim \text{Bern}(\theta_{01} + \theta_{02})$, and

$$\mathbb{P}(Z_1 \leq y \mid Z_0, \boldsymbol{B}_{0:1}) = \begin{cases} \theta_{01} \, G_{01}(y) + \theta_{012} \, G_{012}(y), & B_1 = 1, \\ F_E(y), & B_1 = 0. \end{cases}$$

The transition mechanism is given by

$$B_t \mid \left(\boldsymbol{B}_{t-2 : t-1}, \boldsymbol{Z}_{t-2 : t-1}, \left\{ t \leq T^B \right\} \right) \sim \text{Bern}(m_{\{\boldsymbol{B}_{t-2 : t-1}\}}(\boldsymbol{Z}_{t-2 : t-1})),$$

and

$$\mathbb{P}\left(Z_t \leq z \mid \boldsymbol{B}_{t-2 : t}, \boldsymbol{Z}_{t-2 : t-1}, \left\{ t \leq T^B \right\} \right) = G_{\{\boldsymbol{B}_{t-2 : t-1}\}, \{B_t\}}\left(z - a_{\boldsymbol{B}_{t-2 : t-1}, B_t}(\boldsymbol{Z}_{t-2 : t-1}) \right).$$

Panel (*a*) in Figure 2 illustrates a realization from a special case of this second-order Markov process. This realized path shows that after witnessing an extreme event at time $t = 0$ the process transitions to the body of the process at time $t = 1$ and then has two extreme states at $t = 2$ and $t = 3$ and two non-extreme states at $t = 4$ and $t = 5$. After two non-extreme values the process has permanently transitioned to its equilibrium, that is, for $t = 6, \ldots$ in this realization. The sampling distribution of the average termination time T_B of the hidden tail chain is presented in panel (*e*), whereas the behaviour of the hidden tail chain conditioned on its terminating after eight steps, that is, $T_B = 8$, is shown in panel (*f*). This shows that whilst at an extreme state, the average value of Z_t is stable through time.

Appendix A. Proofs

A.1. Preparatory results for Theorems 1 and 2

The proofs of Theorems 1 and 2 are based on Lemmas 1 and 2 below, whose proofs are similar to those of Lemmas 1 and 2 in [20] and are therefore omitted for brevity.

Lemma 1. *Let $\{X_t : t = 0, 1, \ldots\}$ be a homogeneous k th-order Markov chain satisfying Assumption 3. Then, for any $g \in C_b(\mathbb{R})$ and for each time step $t = k, k + 1, \ldots$, as $v \to \infty$*

$$\int_{\mathbb{R}} g(x) \pi \left(A_t(v, z), a_t(v) + b_t(v) \, dx \right) \to \int_{\mathbb{R}} g(\psi_t^a(z) + \psi_t^b(z) \, x) \, K(dx),$$

and the convergence holds uniformly on compact sets in the variable $z \in \mathbb{R}^k$.

Lemma 2. *Let* $\{X_t : t = 0, 1, \ldots\}$ *be a homogeneous k th-order Markov chain satisfying Assumption* 5. *Then, for any* $g \in C_b([0, \infty))$ *and for each time step* $t = k, k+1, \ldots$, *as* $v \to \infty$

$$\int_{[0,\infty)} g(y)\pi\left(\boldsymbol{B}_t(v, \boldsymbol{x}), b_{t+1}(v)\, dy\right) \to \int_{[0,\infty)} g(\psi_t^b(\boldsymbol{x})\, y)\, K(dy),$$

and the convergence holds uniformly on compact sets in the variable $\boldsymbol{x} \in [\delta_1, \infty) \times \cdots \times [\delta_k, \infty)$ *for any* $(\delta_1, \ldots, \delta_k) \in (0, \infty)^k$.

Lemma 3. (*Slight variant of* [14].) *Let (E,d) be a complete locally compact separable metric space and* μ_n *a sequence of probability measures which converges weakly to a probability measure* μ *on E as* $n \to \infty$.

   (i) *Let* φ_n *be a uniformly bounded sequence of measurable functions which converges uniformly on compact sets of E to a continuous function* φ. *Then* φ *is bounded on E and* $\lim_{n\to\infty} \mu_n(\varphi_n) \to \mu(\varphi)$.

   (ii) *Let F be a topological space. If* $\varphi \in C_b(F \times E)$, *then the sequence of functions* $F \ni x \mapsto \int_E \varphi(x, y)\mu_n(dy) \in \mathbb{R}$ *converges uniformly on compact sets of F to the (necessarily continuous) function* $F \ni x \mapsto \int_E \varphi(x, y)\mu(dy) \in \mathbb{R}$.

A.2. Proofs of Theorems 1 and 2

Preliminaries. Let $a_0(v) \equiv v$ and $b_0(v) \equiv 1$, and define

$$v_u(z_0) = u + \sigma(u)z_0, \qquad A_t(v, x) = a_t(v) + b_t(v)x, \text{ and}$$
$$A_{t-k:t-1}(v, \boldsymbol{x}_{t-k:t-1}) = (A_{t-k}(v, x_{t-k}), \ldots, A_{t-1}(v, x_{t-1})). \tag{48}$$

We note that in our notation, when $k = 1$, the initial conditional distribution of the rescaled Markov chain is

$$\frac{F_0(v_u(dz_0))}{\overline{F}_0(u)} = \mathbb{P}\left(\frac{X_0 - u}{\sigma(u)} \in dz_0 \mid X_0 > u\right), \tag{49}$$

whereas when $k > 1$ it equals the product of the right-hand side of Equation (49) with

$$\pi_0(v_u(z_0), \boldsymbol{A}_{1:k-1}(v_u(z_0), dz_{1:k-1})) := \mathbb{P}\left(\bigcap_{j=1}^{k-1}\left\{\frac{X_j - a_j(X_0)}{b_j(X_0)} \in dz_j\right\} \,\middle|\, \frac{X_0 - u}{\sigma(u)} = z_0\right).$$

For $j \in \{k, \ldots, t\}$ with $t \geq k \geq 1$ the transition kernels of the rescaled Markov chain can be written as

$$\pi(\boldsymbol{A}_{j-k:j-1}(v_u(z_0), z_{j-1,k}), A_j(v_u(z_0), dz_j))$$
$$= \mathbb{P}\left(\frac{X_j - a_j(X_0)}{b_j(X_0)} \in dz_j \,\middle|\, \left\{\frac{X_{j-i} - a_{j-i}(X_0)}{b_{j-i}(X_0)} = z_{j-i}\right\}_{i=1,\ldots,k}\right).$$

Proof of Theorem 1. Consider, for $t \geq k \geq 1$, the measures

$$\mu_t^{(u)}(dz_0, \ldots, dz_t) = \prod_{j=k}^{t} \pi(\boldsymbol{A}_{j-k:j-1}(v_u(z_0), \boldsymbol{z}_{j-k:j-1}), A_j(v_u(z_0), dz_j))$$

$$\times [\pi_0(v_u(z_0), \boldsymbol{A}_{1:k-1}(v_u(z_0), dz_{1:k-1}))]^{1(k>1)} \frac{F_0(v_u(dz_0))}{\overline{F}_0(u)}$$

and

$$\mu_t(dz_0, \ldots, dz_t) = \prod_{j=k}^{t} K\left(\frac{dz_j - \psi_j^a(z_{j-k:j-1})}{\psi_j^b(z_{j-k:j-1})}\right) [G(dz_1 \times \cdots \times dz_{k-1})]^{1(k>1)} H_0(dz_0)$$

on $[0, \infty) \times \mathbb{R}^t$, where $1(k > 1)$ denotes the indicator function of $\{k > 1\}$. For $f \in C_b([0, \infty) \times \mathbb{R}^t)$, we may write

$$\mathbb{E}\left[f\left(\frac{X_0 - u}{\sigma(u)}, \frac{X_1 - a_1(X_0)}{b_1(X_0)}, \ldots, \frac{X_t - a_t(X_0)}{b_t(X_0)}\right) \,\bigg|\, X_0 > u\right] =$$

$$= \int_{[0,\infty)\times\mathbb{R}^t} f(z_{0:t})\mu_t^{(u)}(dz_0, \ldots, dz_t)$$

and

$$\mathbb{E}[f(E_0, Z_1, \cdots, Z_t)] = \int_{[0,\infty)\times\mathbb{R}^t} f(z_{0:t})\mu_t(dz_0, \ldots, dz_t).$$

We need to show that $\mu_t^{(u)}$ converges weakly to μ_t. Let $g_0 \in C_b([0, \infty))$ and $g \in C_b(\mathbb{R}^k)$. The proof is by induction on t. For $t = k$ it suffices to show that

$$\int_{[0,\infty)\times\mathbb{R}^k} g_0(z_0)g(z_1, \ldots, z_k)\mu_k^{(u)}(dz_0, dz_1, \ldots, dz_k)$$

$$= \int_{[0,\infty)} g_0(z_0)\Bigg[\int_{\mathbb{R}^k} g(z_{1:k})\pi((v_u(z_0), \boldsymbol{A}_{1:k-1}(v_u(z_0), z_{1:k-1})), A_k(v_u(z_0), dz_k))$$

$$\pi_0(v_u(z_0), \boldsymbol{A}_{1:k-1}(v_u(z_0), dz_{1:k-1}))^{1(k>1)}\Bigg] \frac{F_0(v_u(dz_0))}{\overline{F}_0(u)} \qquad (50)$$

converges to $\mathbb{E}(g_0(E_0))\,\mathbb{E}(g(Z_1, \ldots, Z_k))$.

By Assumptions 2 and 3, the integrand in the term in square brackets in (50) converges pointwise to a limit and is dominated by

$$\sup\left\{g(z) : z \in \mathbb{R}^k\right\} \times \pi((v_u(z_0), \boldsymbol{A}_{1:k-1}(v_u(z_0), z_{1:k-1})), A_k(v_u(z_0), dz_k)).$$

Lebesgue's dominated convergence theorem yields that the term in square brackets in (50) is bounded and converges to $\mathbb{E}[g(\boldsymbol{Z}_{1:k})]$ for $u \to \infty$, since $v_u(z_0) \to \infty$ as $u \to \infty$. The convergence holds uniformly in the variable $z_0 \in [0, \infty)$ since $\sigma(u) > 0$. Therefore Lemma 3 applies, which guarantees convergence of the entire term (50) to $\mathbb{E}(E_0)\,\mathbb{E}(g(\boldsymbol{Z}_{1:k}))$ by Assumption 1.

Next, assume that the statement is true for some $t > k$. It suffices to show that for any $g_0 \in C_b([0, \infty) \times \mathbb{R}^t)$, $g \in C_b(\mathbb{R})$,

$$\int_{[0,\infty)\times\mathbb{R}^{t+1}} g_0(z_{0:t})g(z_{t+1})\mu_{t+1}^{(u)}(dz_0, dz_1, \ldots, dz_{t+1})$$

$$= \int_{[0,\infty)\times\mathbb{R}^t} g_0(z_{0:t})\Bigg[\int_{\mathbb{R}} g(z_{t+1})\pi(\boldsymbol{A}_{t-k+1:t}(v_u(z_0), z_{t-k+1:t}), A_t(v_u(z_0), dz_{t+1}))\Bigg]$$

$$\mu_t^{(u)}(dz_0, dz_1, \ldots, dz_t) \qquad (51)$$

converges to

$$\int_{[0,\infty)\times\mathbb{R}^{t+1}} g_0(\mathbf{z}_{0\,:\,t})g(z_{t+1})\mu_{t+1}(dz_0, dz_1, \ldots, dz_{t+1})$$

$$= \int_{[0,\infty)\times\mathbb{R}^t} g_0(\mathbf{z}_{0\,:\,t})\left[\int_{\mathbb{R}} g(z_{t+1})K\left(\frac{dz_{t+1} - \psi_t^a(\mathbf{z}_{t-k+1\,:\,t})}{\psi_t^b(\mathbf{z}_{t-k+1\,:\,t})} \right) \right]\mu_t(dz_0, dz_1, \ldots, dz_t).$$

(52)

The term in square brackets in (51) is bounded, and by Lemma 1 and Assumptions 2 and 3, it converges uniformly on compact sets in both variables $(z_0, \mathbf{z}_{t-k+1\,:\,t}) \in [0, \infty) \times \mathbb{R}^k$ jointly, since $\sigma(u) > 0$. Hence the induction hypothesis and Lemma 3 imply the desired result.

Proof of Theorem 2. Define

$$\mathbf{b}_{t-k\,:\,t-1}(v, \mathbf{x}_{t-k\,:\,t-1}) = (b_{t-k}(v)\,x_{t-k}, \ldots, b_{t-1}(v)\,x_{t-1}).$$

Consider the measures

$$\mu_t^{(u)}(dz_0, \ldots, dz_t) = \left[\prod_{j=k}^t \pi(\mathbf{b}_{j-k\,:\,j-1}(v_u(z_0), \mathbf{z}_{j-k\,:\,j-1}), b_j(v_u(z_0))\,dz_j) \right]$$

$$\times [\pi_0(v_u(z_0), \mathbf{b}_{1\,:\,k-1}(v_u(z_0), d\mathbf{z}_{1\,:\,k-1}))]^{\mathbf{1}(k>1)} \frac{F_0(v_u(dz_0))}{\overline{F}_0(u)} \qquad (53)$$

and

$$\mu_t(dz_0, \ldots, dz_t) = \left[\prod_{j=k}^t K\left(\frac{dz_j}{\psi_j^b(\mathbf{z}_{j-k\,:\,j-1})} \right) \right][G(dz_1, \ldots, dz_{k-1})]^{\mathbf{1}(k>1)} H_0(dz_0) \qquad (54)$$

on $[0, \infty) \times [0, \infty)^t$. We may write

$$\mathbb{E}\left[f\left(\frac{X_0 - u}{\sigma(u)}, \frac{X_1}{b_1(X_0)}, \ldots, \frac{X_t}{b_t(X_0)} \right) \,\bigg|\, X_0 > u \right] = \int_{[0,\infty)\times[0,\infty)^t} f(\mathbf{z}_{0\,:\,t})\mu_t^{(u)}(dz_0, \ldots, dz_t)$$

and

$$\mathbb{E}[f(E_0, Z_1, \ldots, Z_t)] = \int_{[0,\infty)\times[0,\infty)^t} f(\mathbf{z}_{0\,:\,t})\mu_t(dz_0, \ldots, dz_t)$$

for $f \in C_b([0, \infty) \times [0, \infty)^t)$. Note that $b_j(0)$, $j = 1, \ldots, t$, need not be defined in (53), since $v_u(z_0) \geq u > 0$ for $z_0 \geq 0$ and sufficiently large u, whereas (54) is well-defined, since the measures G and K put no mass at any half-plane $C_j = \{(\mathbf{z}_{1\,:\,k-1}) \in [0, \infty)^{k-1} : z_j = 0\} \in [0, \infty)^{k-1}$ and at $0 \in [0, \infty)$, respectively. Formally, we may set $\psi_j^b(\mathbf{0}_k) = 1$, $j = 1, \ldots, t$, in order to emphasize that we consider measures on $[0, \infty)^{t+1}$ instead of $[0, \infty) \times (0, \infty)^t$. To prove the theorem, we need to show that $\mu_t^{(u)}(dz_0, \ldots, dz_t)$ converges weakly to $\mu_t(dz_0, \ldots, dz_t)$. The proof is by induction on t. We show two statements by induction on t:

(**I**) $\mu_t^{(u)}(dz_0, \ldots, dz_t)$ converges weakly to $\mu_t(dz_0, \ldots, dz_t)$ as $u \uparrow \infty$;

(**II**) for all $\varepsilon > 0$ there exists $\delta_t > 0$ such that $\mu_t([0, \infty) \times [0, \infty)^{t-1} \times [0, \delta_t]) < \varepsilon$.

We start by proving the case $t = k$.

(I) <u>for $t = k$</u>: It suffices to show that for any $g_0 \in C_b([0, \infty))$ and $g \in C_b([0, \infty)^{k-1})$,

$$\int_{[0,\infty)\times[0,\infty)^{k-1}\times[0,\infty)} g_0(z_0)g(z_{1:k})\mu_1^{(u)}(dz_0, \ldots, dz_k)$$

$$= \int_{[0,\infty)} g_0(z_0)$$

$$\times \left[\int_{[0,\infty)^{k-1}} \int_{[0,\infty)} g(z_1, \ldots, z_k)\pi((v_u(z_0), \boldsymbol{b}_{0:k-1}(v_u(z_0), z_{0:k-1})), b_k(v_u(z_0))dz_k) \right]$$

$$\times \pi_0(v_u(z_0), \boldsymbol{b}_{1:k-1}(v_u(z_0), dz_{1:k-1}))\frac{F_0(v_u(dz_0))}{\overline{F}_0(u)} \tag{55}$$

converges to

$$\int_{[0,\infty)\times[0,\infty)^k} g_0(z_0)g(z_{1:k})\mu_1(dz_0, \ldots, dz_k) = \mathbb{E}(g_0(E_0))\mathbb{E}(g(Z_1, \ldots, Z_k)).$$

By Assumptions 1 and 5, the integrand in the term in square brackets converges pointwise to a limit and is dominated by

$$\sup\{g(\boldsymbol{z}) : \boldsymbol{z} \in \mathbb{R}_+^k\} \times \pi((v_u(z_0), \boldsymbol{b}_{1:k-1}(v_u(z_0), z_{1:k-1})), b_k(v_u(z_0))\, dz_k).$$

Lebesgue's dominated convergence theorem yields that the term in square brackets in (55) is bounded and converges to $\mathbb{E}(g(\boldsymbol{Z}_{1:k}))$ for $u \uparrow \infty$, since $v_u(z_0) \to \infty$ for $u \uparrow \infty$. The convergence is uniform in the variable z_0, since $\sigma(u) > 0$. Therefore, Lemma 3(i) applies, which guarantees convergence of the entire term (55) to $\mathbb{E}(g_0(E_0))\mathbb{E}[g(\boldsymbol{Z}_{1:k})]$ by Assumption 1.

(II) <u>for $t = k$</u>: Since $K(\{0\}) = 0$, there exists $\delta > 0$ such that $K([0, \delta]) < \varepsilon$, which immediately implies $\mu_k([0, \infty)^k \times [0, \delta]) = H_0([0, \infty)) [G([0, \infty)^{k-1})]^{1(k>1)} K([0, \delta]) < \varepsilon$.

Now, let us assume that both statements ((I) and (II)) are proved for some $t \in \mathbb{N}$.

(I) <u>for $t + 1$</u>: It suffices to show that for any $g_0 \in C_b([0, \infty) \times [0, \infty)^t)$, $g \in C_b([0, \infty))$,

$$\int_{[0,\infty)\times[0,\infty)^{t+1}} g_0(z_{0:t}) \times g(z_{t+1})\mu_{t+1}^{(u)}(dz_0, dz_1, \ldots, dz_t, dz_{t+1})$$

$$= \int_{[0,\infty)\times[0,\infty)^t} g_0(z_{0:t})$$

$$\times \left[\int_{[0,\infty)} g(z_{t+1})\pi(\boldsymbol{b}_{t-k+1:t}(v_u(z_0), z_{t-k+1:t}), b_{t+1}(v_u(z_0))\, dz_{t+1}) \right]$$

$$\times \mu_t^{(u)}(dz_0, dz_1, \ldots, dz_t) \tag{56}$$

converges to

$$\int_{[0,\infty)\times[0,\infty)^{t+1}} g_0(z_{0:t})g(z_{t+1})\mu_{t+1}(dz_0, dz_1, \ldots, dz_t, dz_{t+1})$$

$$= \int_{[0,\infty)\times[0,\infty)^t} g_0(z_{0:t})\left[\int_{[0,\infty)} g(z_{t+1})K\big(dz_{t+1}/\psi_t^b\big(z_{t-k+1:t}\big)\big) \right]$$

$$\times \mu_t(dz_0, dz_1, \ldots, dz_t). \tag{57}$$

From Lemma 2 and Assumptions 4 and 5 we know that, for any $\delta > 0$, the (bounded) term in square brackets in (56) converges uniformly on compact sets in the variable $z_{t-k+1:t} \in \prod_{i=1}^{k} [\delta_i, \infty)$ to the continuous function

$$\int_{[0,\infty)} g(\psi_t^b(z_{t-k+1:t}) z_{t+1}) K(dz_{t+1})$$

(the term in square brackets in (57)). This convergence holds even uniformly on compact sets in both variables $(z_0, z_{t-k+1:t}) \in [0, \infty) \times \prod_{i=1}^{k} [\delta_i, \infty)$ jointly, since $\sigma(u) > 0$. Hence, the induction hypothesis (I) and Lemma 3(i) imply that for any $\delta > 0$ the integral in (56) converges to the integral in (57) if the integrals with respect to μ_t and $\mu_t^{(u)}$ were restricted to $A_\delta :=$ $[0, \infty) \times [0, \infty)^{t-1} \times [\delta, \infty)$ (instead of being taken over $[0, \infty) \times [0, \infty)^{t-1} \times [0, \infty)$).

Since g_0 and g are bounded, it suffices to control the mass of μ_t and $\mu_t^{(u)}$ on the complement $A_\delta^c = [0, \infty) \times [0, \infty)^{t-1} \times [0, \delta)$. For some prescribed $\varepsilon > 0$ it is possible to find some sufficiently small $\delta > 0$ and sufficiently large u so that $\mu_t(A_\delta^c) < \varepsilon$ and $\mu_t^{(u)}(A_\delta^c) < 2\varepsilon$. Because of the induction hypothesis (II), we indeed have $\mu_t(A_{\delta_t}) < \varepsilon$ for some $\delta_t > 0$. Choose $\delta = \delta_t/2$ and note that the sets of the form A_δ are nested. Let C_δ be a continuity set of μ_t with $A_\delta^c \subset C_\delta \subset A_{2\delta}^c$. Then the value of μ_t on all three sets $A_\delta^c, C_\delta, A_{2\delta}^c$ is smaller than ε, and by the induction hypothesis (I), the value $\mu_t^{(u)}(C_\delta)$ converges to $\mu_t(C_\delta) < \varepsilon$. Hence, for sufficiently large u, we also have $\mu_t^{(u)}(A_\delta^c) < \mu_t^{(u)}(C_\delta) < \mu_t(C_\delta) + \varepsilon < 2\varepsilon$, as desired.

(II) for $t + 1$: We have for any $\delta > 0$ and any $c > 0$

$$\mu_{t+1}\big([0, \infty) \times [0, \infty)^t \times [0, \delta]\big) = \int_{[0,\infty) \times [0,\infty)^t} K\big([0, \delta/\psi_t^b(z_{t-k+1:t})]\big) \mu_t(dz_0, \dots, dz_t).$$

Splitting the integral according to $\big\{\psi_t^b(z_{t-k+1:t}) > c\big\}$ or $\big\{\psi_t^b(z_{t-k+1:t}) \le c\big\}$ yields

$$\mu_{t+1}([0, \infty) \times [0, \infty)^t \times [0, \delta]) \le K([0, \delta/c]) + \mu_t\big([0, \infty) \times [0, \infty)^{t-1} \times (\psi_t^b)^{-1}([0, c])\}\big).$$

By Assumption 5(i) and the induction hypothesis (II), we may choose $c > 0$ sufficiently small so that the second summand $\mu_t([0, \infty) \times [0, \infty)^{t-1} \times (\psi_t^b)^{-1}([0, c])\})$ is smaller than $\varepsilon/2$. Next, since $K(\{0\}) = 0$, it is possible to choose $\delta_{t+1} = \delta > 0$ accordingly small so that the first summand $K\big([0, \frac{\delta}{c}]\big)$ is smaller than $\varepsilon/2$, which shows (II) for $t + 1$.

A.3. Proofs of Propositions

Proof of Proposition 1. We start by proving that (*i*) implies (*ii*). Let $\mathbf{0} := \mathbf{0}_k$ and suppose there exist a_t, b_t, ψ_t^a, and ψ_t^b such that (*i*) holds. Then, for $t = k, k+1 \dots$,

$$\frac{a(A_t(v, z_v)) - a(A_t(v, \mathbf{0}))}{b(A_t(v, \mathbf{0}))}$$

$$= \frac{a(A_t(v, z_v)) - a_t(v)}{b_t(v)} \frac{b_t(v)}{b(A_t(v, \mathbf{0}))} - \frac{a(A_t(v, \mathbf{0})) - a_t(v)}{b_t(v)} \frac{b_t(v)}{b(A_t(v, \mathbf{0}))}$$

$$\to \frac{\psi_t^a(z)}{\psi_t^b(\mathbf{0})} - \frac{\psi_t^a(\mathbf{0})}{\psi_t^b(\mathbf{0})} = \frac{\psi_t^a(z) - \psi_t^a(\mathbf{0})}{\psi_t^b(\mathbf{0})} \quad \text{whenever } z_v \to z \text{ as } v \to \infty,$$

and

$$\frac{b(\boldsymbol{A}_t(v, z_v))}{b(\boldsymbol{A}_t(v, \boldsymbol{0}))} = \frac{b(\boldsymbol{A}_t(v, z_v))}{b_t(v)} \frac{b_t(v)}{b(\boldsymbol{A}_t(v, \boldsymbol{0}))} \rightarrow \frac{\psi_t^b(z)}{\psi_t^b(\boldsymbol{0})} \quad \text{whenever } z_v \rightarrow z \text{ as } v \rightarrow \infty.$$

Next we prove (*ii*) implies (*i*). Let $a_t(v) = a(\boldsymbol{A}_t(v, 0)) - c\, b(\boldsymbol{A}_t(v, \boldsymbol{0}))$ and $b_t(v) = d\, b(\boldsymbol{A}_t(v, \boldsymbol{0}))$ for arbitrary constants $c \in \mathbb{R}$, $d \in \mathbb{R}_+$. Then, for $t = k, k+1 \ldots,$

$$\frac{a(\boldsymbol{A}_t(v, z_v)) - a_t(v)}{b_t(v)} = \frac{b(\boldsymbol{A}_t(v, \boldsymbol{0}))}{b_t(v)} \left[\frac{a(\boldsymbol{A}_t(v, z_v)) - a(\boldsymbol{A}_t(v, \boldsymbol{0}))}{b(\boldsymbol{A}_t(v, \boldsymbol{0}))} + c \right]$$

$$\rightarrow \frac{\lambda_t^b(z)}{d}[\lambda_t^a(z) + c] \quad \text{whenever } z_v \rightarrow z \in \mathbb{R}^k \text{ as } v \rightarrow \infty.$$

Define $\psi_t^a(z) = (\lambda_t^a(z) + c)/d$ and $\psi_t^b(z) = \lambda_t^b(z)/d$. By assumption, $\lambda_t^a(\boldsymbol{0}) = 0$ and $\lambda_t^b(\boldsymbol{0}) = 1$. Hence, $\lambda_t^a(z) = [\psi_t^a(z) - \psi_t^a(\boldsymbol{0})]/\psi_t^b(\boldsymbol{0})$ and $\lambda_t^b(z) = \psi_t^b(z)/\psi_t^b(\boldsymbol{0})$, which completes the proof.

Proof of Proposition 2. The recurrence relation

$$\alpha_t = c \left[\sum_{i=1}^k \gamma_i (\gamma_i \alpha_{t-k+i-1})^\delta / (\gamma_1^{1+\delta} + \cdots + \gamma_k^{1+\delta}) \right]^{1/\delta}$$

can be converted to the homogeneous linear recurrence relation $y_t = \sum_{i=1}^k c_i y_{t-k+i-1}$, where $\{y_t\} = \{\alpha_t^\delta\}$ and $c_i = c^\delta \gamma_i^{1+\delta} / (\gamma_1^{1+\delta} + \cdots + \gamma_k^{1+\delta})$. Solving the recurrence relation and transforming the solution to the original sequence $\{\alpha_t\}$ leads to the claim.

Proof of Proposition 3. (*i*) Because a_P is 1-homogeneous, a satisfies the property (19). By the definition of the right-inverse $V_{0:k-1}^{\leftarrow}$ and Assumption 8, we have that, for all $(\boldsymbol{y}_{0:k-1}, q)$ in the domain of $V_{0:k-1}^{\leftarrow}$,

$$V_{0:k-1}\big\{\boldsymbol{y}_{0:k-1}, V_{0:k-1}^{\leftarrow}(q;\boldsymbol{y}_{0:k-1})\big\}$$
$$= V_{0:k-1}(\boldsymbol{y}_{0:k-1}, \infty)\, K_P\big\{V_{0:k-1}^{\leftarrow}(q;\boldsymbol{y}_{0:k-1})/a_P(\boldsymbol{y}_{0:k-1})\big\} = q.$$

Hence $V_{0:k-1}^{\leftarrow}(q;\boldsymbol{y}_{0:k-1}) = a_P(\boldsymbol{y}_{0:k-1})\, K_P^{\leftarrow}\big\{q/V_{0:k-1}(\boldsymbol{y}_{0:k-1}, \infty)\big\}$. Taking logarithms, setting $\boldsymbol{y}_{0:k-1} = e^{\boldsymbol{x}_{0:k-1}}$, and letting $q = p^\star V_{0:k-1}(\boldsymbol{y}_{0:k-1}, \infty)$ where $p^\star$ satisfies $K_P^{\leftarrow}(p^\star) = 1$ gives

$$a(\boldsymbol{x}_{0:k-1}) = \log V_{0:k-1}^{\leftarrow}(p^\star V_{0:k-1}(e^{\boldsymbol{x}_{0:k-1}}, \infty); e^{\boldsymbol{x}_{0:k-1}}) - \log K_P^{\leftarrow}(p^\star).$$

(*ii*) Since a satisfies the property (19), Appendix A.7 shows that Assumption 3 holds with limit distribution $K(x;z_{0:k-1})$ given by (42). Using Assumption 8, we further have that under the normalizing functionals a and b, (42) simplifies to $K(x;z_{0:k-1}) = K_P(e^x)$ for all $z_{0:k-1} \in \mathbb{R}^k$.

(*iii*) The limit distribution $K(x;z_{0:k-1})$ in (42) does not depend on $z_{0:k-1}$. Because a_P is positive, 1-homogeneous, and continuous, 1 is in the image of a_P and thus 0 is in the image of a. Hence there exists $z_{0:k-1}^\star \in \mathbb{R}^k$ such that $a(z_{0:k-1}^\star) = 0$ and $K_P(e^x) = K(x;z_{0:k-1}^\star)$, which proves the claim.

A.4. Proof of Corollaries

Proof of Corollary 1. Since a is continuous, we have that $a(v\mathbf{1}_k + z_v) - v = a(z_v) \rightarrow a(z)$ whenever $z_v \rightarrow z \in \mathbb{R}^k$. Hence the convergence (8) holds true with $\psi_t^a(z) = a(z)$ and $\psi_t^b(z) = 1$.

For any $s \in \mathbb{N}$, [1] show that under the assumptions of Corollary 1, the random vector $X_P = \exp(X_{0:s})$ is multivariate regularly varying; that is, for any $A \subset \Delta^s$,

$$\mathbb{P}\left(\frac{X_P}{\|X_P\|} \in A, \|X_P\| > v\,r \,\middle|\, \|X_P\| > v\right) \to H(A)\,r^{-1}, \qquad r \geq 1,$$

where H is a Radon measure on Δ^s satisfying $H(\Delta^s) = s+1$ and $\int_{\Delta^s} w_j dH(\boldsymbol{w}_{0:s}) = 1$ for $j = 0, \ldots, s$. Theorem 1 and Proposition 4 of [8] imply that $Z_t \sim G_t$ where

$$G_t(z) = \int_0^{q(z)} (1-w)H_t(dw) \quad \text{with } q(z) = e^z/(1+e^z), \qquad z \in \mathbb{R}.$$

Here H_t denotes the lag-t bivariate spectral measure associated with H; that is, for every $t \geq 1$, H_t is a Radon measure on Δ^1 that satisfies $H_t(\Delta^1) = 2$ and $\int_{\Delta^1} w dH_t(w) = 1$. Thus, we have that the expected value of Z_t satisfies

$$\mathbb{E}(Z_t) = \int_{\mathbb{R}} z\,dG_t(z) = \int_{\mathbb{R}} \int_0^z du\; dG_t(z) = -\int_{-\infty}^0 G_t(z)\,dz + \int_0^\infty [1 - G_t(z)]\,dz$$

$$= -\int_0^{1/2} \int_0^u (1-w)\,dH_t(w)\,d\log\left(\frac{u}{1-u}\right) + \int_{1/2}^1 \int_u^1 (1-w)\,dH_t(w)\,d\log\left(\frac{u}{1-u}\right)$$

$$= -\int_0^{1/2} \int_w^{1/2} \left[d\log\left(\frac{u}{1-u}\right)\right] (1-w)\,dH_t(w) + \int_{1/2}^1 \int_{1/2}^w \left[d\log\left(\frac{u}{1-u}\right)\right] (1-w)\,dH_t(w)$$

$$= \int_{\Delta^1} \log\{w/(1-w)\}\,(1-w)\,dH_t(w) < \log\left(\int_{\Delta^1} w\,dH_t(w)\right) = 0.$$

The strict inequality follows from Jensen's inequality, the strict concavity of the log function, and Assumption 2 (*ii*), which requires G_t to be a non-degenerate distribution. The latter ensures that $H_t \neq 2\,\delta_{1/2}$, where δ_x denotes the Dirac measure at $\{x\}$.

Proof of Corollary 2. First we prove the statement that $\alpha_i < 1$ for all $i \geq k$ if the recurrence relation (20) holds. Let $\alpha_{t-k:t-1}^\star := \max \boldsymbol{\alpha}_{t-k:t-1}$; then when $t = k$ we have $\alpha_{0:k-1}^\star = 1$ by the conditions of Corollary 2, as $\alpha_0 = 1$. From the relation (20) with $t = k$ we have that $\alpha_k = a(\boldsymbol{\alpha}_{0:k-1}) \leq a(\alpha_{0:k-1}^\star \mathbf{1}_k) = \alpha_{0:k-1}^\star a(\mathbf{1}_k) < 1 \times 1 = 1$. Here, the first inequality comes from the order-preserving property of a, and the second-to-last equality comes from a being 1-homogeneous. The result $\alpha_t < 1$ for all $t \geq k$ follows from induction over $t \geq k$, noting that all $\alpha_{t-k:t-1}^\star < 1$ for $t \geq k+1$.

For the remainder of the proof it suffices to show that for $a_t(x) = \alpha_t x$ and $b_t(x) = x^\beta$, with α_t given by (20), the convergence (8) holds true with $\psi_t^a(z) = \nabla a(\boldsymbol{\alpha}_{t-k:t-1}) \cdot z$ and $\psi_t^b(z) = b(\boldsymbol{\alpha}_{t-k:t-1})$, and that $\alpha_t \to 0$ as $t \to \infty$. Since a is twice continuously differentiable, we have

$$v^{-\beta}[a(\boldsymbol{\alpha}_{t-k:t-1}\,v + v^\beta\,z_v) - a(\boldsymbol{\alpha}_{t-k:t-1}\,v)] =$$
$$= \nabla a(\boldsymbol{\alpha}_{t-k:t-1}\,v) \cdot z_v + v^\beta z_v^\top \nabla\nabla^\top a(\boldsymbol{\alpha}_{t-k:t-1}v)z_v + O(v^{2(\beta-1)})$$
$$= \nabla a(\boldsymbol{\alpha}_{t-k:t-1}) \cdot z_v + v^{\beta-1} z_v^\top \nabla\nabla^\top a(\boldsymbol{\alpha}_{t-k:t-1})z_v + o(v^{\beta-1}) \quad \text{as } v \to \infty,$$

where the last equality follows because a is 1-homogeneous, which gives that

$$\nabla a(\boldsymbol{\alpha}_{t-k:t-1} v) = \nabla a(\boldsymbol{\alpha}_{t-k:t-1}) \quad \text{and} \quad \nabla \nabla^\top a(\boldsymbol{\alpha}_{t-k:t-1} v) = v^{-1} \nabla \nabla^\top a(\boldsymbol{\alpha}_{t-k:t-1}).$$

Similarly, because b is continuous and β-homogeneous with $\beta \in [0, 1)$, this gives $v^{-\beta} b(\boldsymbol{\alpha}_{t-k:t-1} v + v^\beta z_v) = b(\boldsymbol{\alpha}_{t-k:t-1} + v^{\beta-1} z_v) \to b(\boldsymbol{\alpha}_{t-k:t-1})$, as $v \to \infty$. Hence the convergence (8) holds true with $\psi_t^a(z) = \nabla a(\boldsymbol{\alpha}_{t-k:t-1}) \cdot z$ and $\psi_t^b(z) = b(\boldsymbol{\alpha}_{t-k:t-1})$.

Lastly, we show that $\alpha_t \to 0$ as $t \to \infty$. Let $f: \mathbb{R}_+^k \to \mathbb{R}_+^k$ with $f(x_{0:k-1}) = (x_{1:k-1}, a(x_{0:k-1}))$ for all $x_{0:k-1} \in \mathbb{R}_+^k$. Let $r^t = f^{kt} := f \circ f \circ \cdots \circ f$, $f^0 := \text{id}$, $t = 0, 1, \ldots$, denote the $(k\,t)$-fold composition of f with itself. Since a is assumed order-preserving and 1-homogeneous, with $a(\mathbf{1}_k) < 1$, it follows that the function r is also an order-preserving 1-homogeneous function, mapping $\mathbb{R}_+^k$ into $\mathbb{R}_+^k$, with $r(\mathbf{1}_k) < \mathbf{1}_k$ componentwise. Since $\boldsymbol{\alpha}_{0:k-1} \in [0, 1]^k$ and r is order-preserving, we have $r(\boldsymbol{\alpha}_{0:k-1}) < \mathbf{1}_k$ componentwise too. The latter inequality implies that there exists $v \in (0, 1)$ such that $r(\boldsymbol{\alpha}_{0:k-1}) \leq v \mathbf{1}_k$. Similarly, the 2-fold composition of r gives $r^2(\boldsymbol{\alpha}_{0:k-1}) = r(r(\boldsymbol{\alpha}_{0:k-1})) \leq r(v \mathbf{1}_k) = v^2 \mathbf{1}_k$, where the latter equality follows from the homogeneity of the map r. Likewise, iterating forward yields $r^t(\boldsymbol{\alpha}_{0:k-1}) \leq v^t \mathbf{1}_k \to \mathbf{0}_k$ as $t \to \infty$. The claim is proved upon noting that the function r satisfies $\{r^t(\boldsymbol{\alpha}_{0:k-1}) : t = 0, 1, \ldots\} = \{\alpha_t : t = 0, 1 \ldots\}$.

Proof of Corollary 3. Let $\log \beta_t = \log \beta + \log(\max_{i=1,\ldots,k} \beta_{t-i})$, and consider the convergence (12). For $t = k$ we have

$$v^{-\beta} b(vz_0, v^\beta z_1, \ldots, v^\beta z_{k-1}) = b(z_0, v^{\beta-1} z_1, \ldots, v^{\beta-1} z_{k-1}) \to b(z_0, \mathbf{0}_{k-1}) \quad \text{as } v \to \infty.$$

For $t = k + 1$ we have

$$v^{-\beta^2} b(v^\beta z_0, v^\beta z_1, \ldots, v^\beta z_{k-1}) = b(z_0, z_1, \ldots, z_{k-1}),$$

and for $t = k + j$ with $j \in \{2, \ldots, k-1\}$ we have

$$v^{-\beta^2} b(v^\beta z_0, \ldots v^\beta z_{k-j}, v^{\beta^2} z_{k-j+1} \ldots, v^{\beta^2} z_{k-1})$$
$$= b(z_0, \ldots z_{k-j}, v^{\beta(\beta-1)} z_{k-j+1}, \ldots, v^{\beta(\beta-1)} z_{k-1}) \to b(z_0, \ldots, z_{k-j}, \mathbf{0}_{j-1}) \quad \text{as } v \to \infty.$$

Iterating forward for $t = 2k, 2k + 1, \ldots$, we see that the convergence (12) holds with

$$\psi_t^b(z) = \begin{cases} b(z_0, \mathbf{0}_{k-1}) & \text{when } \text{mod}_k(t) = 0, \\ b(z_0, \ldots, z_{k-1}) & \text{when } \text{mod}_k(t) = 1, \\ b(z_0, \ldots, z_{k-j}, \mathbf{0}_{j-1}) & \text{when } \text{mod}_k(t) = j \in \{2, \ldots, k-1\}, \end{cases}$$

which completes the proof.

A.5. Convergence of multivariate normal full conditional distribution

Let $\boldsymbol{X}_N = (X_{N,0}, \ldots, X_{N,k}) \sim \mathcal{N}(\mathbf{0}_{k+1}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} \in \mathbb{R}^{(k+1) \times (k+1)}$ is a positive definite correlation matrix with $(i+1, j+1)$ element ρ_{ij}, $i, j = 0, \ldots, k$. Let $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$ and write q_{ij} for its $(i+1, j+1)$ element, $i, j = 0, \ldots, k$. For $k \geq 1$ and $z_{0:k-1} \in \mathbb{R}^k$, the conditional distribution of $X_{N,k}$ given $X_{N,0:k-1} = z_{0:k-1}$ is normal with mean $-q_{kk}^{-1} \sum_{i=0}^{k-1} q_{ik} z_i$ and variance q_{kk}^{-1}. Let $\boldsymbol{X} = (X_0, \ldots, X_k)$ with $X_i = -\log\{1 - \Phi(X_{N,i})\}$, so that $X_i \sim \text{Exp}(1)$ for $i = 0, \ldots, k$.

Following the strategy outlined in Section 5.1, we have that for any $t \geq k \geq 1$,

$$\mathbb{P}(X_k < a(\boldsymbol{X}_{0:k-1}) \mid \boldsymbol{X}_{0:k-1} = \boldsymbol{A}_t(v, \boldsymbol{z}_{0:k-1}))$$
$$= \Phi\left[q_{kk}^{1/2}\left\{\Phi_{t,v}^{-1} - \sum_{i=0}^{k-1}\left(-\frac{q_{ik}}{q_{kk}}\right)\Phi_{t-k+i,v}^{-1}\right\}\right], \tag{58}$$

where

$$\Phi_{i,v}^{-1} := \Phi^{\leftarrow}[1 - \exp\{-A_i(v, z_i)\}], \quad \text{for } i = t-1, \ldots, t-k,$$

and

$$\Phi_{t,v}^{-1} := \Phi^{\leftarrow}\{1 - \exp\left(-[a\{\boldsymbol{A}_t(v, \boldsymbol{z}_{0:k-1})\}]\right)\}.$$

Now, let $t = k$. First, we seek a function a such that the conditional probability in Equation (58) converges to a number $p \in (0, 1)$. Suppose that this function satisfies $a(\boldsymbol{a}_t - k : t - 1(v)) \to \infty$ as $v \to \infty$. Using standard asymptotic series for the cumulative distribution function of the standard normal distribution, we have

$$\Phi_{i,v}^{-1} = \{2\,A_i(v, z_i)\}^{1/2} - \frac{\log A_i(v, z_i) + \log 4\pi}{2\{2\,A_i(v, z_i)\}^{1/2}} + o(A_i(v, z_i)^{-1/2}), \quad i = t-k, \ldots, t-1,$$

$$\Phi_{t,v}^{-1} = \{2\,a(\boldsymbol{A}_t(v, \boldsymbol{z}_{0:k-1}))\}^{1/2} - \frac{\log a(\boldsymbol{A}_t(v, \boldsymbol{z}_{0:k-1})) + \log 4\pi}{2\{2\,a(\boldsymbol{A}_t(\boldsymbol{z}_{0:k-1}))\}^{1/2}}$$
$$+ o(a(\boldsymbol{A}_t(v, \boldsymbol{z}_{0:k-1}))^{-1/2}),$$

as $v \to \infty$. Therefore,

$$\Phi_{t,v}^{-1} - \sum_{i=0}^{k-1}\left(-\frac{q_{ik}}{q_{kk}}\right)\Phi_{t-k+i,v}^{-1}$$
$$= (2[a\{\boldsymbol{A}_t(v, \boldsymbol{z}_{0:k-1})\}])^{1/2} - \sum_{i=0}^{k-1}\left(-\frac{q_{ik}}{q_{kk}}\right)\{2\,A_{t-k+i,v}(z_i)\}^{1/2} + o(1), \tag{59}$$

as $v \to \infty$. Substituting in (58), we observe that for the choice of a being $a(\boldsymbol{y}_{0:k-1}) = \left\{\sum_{i=0}^{k-1}(-q_{ik}/q_{kk})\,|y_i|^{1/2}\right\}^2$, for $\boldsymbol{y}_{0:k-1} \in \mathbb{R}^k$, the conditions set out in Section 5.1 are met. In particular, since (59) converges to zero and Φ is continuous, we have the conditional probability (58) converging to $p = 1/2$; that is,

$$\lim_{v \to \infty} \Phi\left[q_{kk}^{1/2}\left\{\Phi_{t,v}^{-1} - \sum_{i=0}^{k-1}\left(-\frac{q_{ik}}{q_{kk}}\right)\Phi_{t-k+i,v}^{-1}\right\}\right] = 1/2.$$

Using similar asymptotic series, we have that for $b(\boldsymbol{y}_{0:k-1}) = a(\boldsymbol{y}_{0:k-1})^{1/2}$ and any $x_k \in \mathbb{R}$,

$$\lim_{v \to \infty} \mathbb{P}\left(X_k < a(\boldsymbol{X}_{0:k-1}) + b(\boldsymbol{X}_{0:k-1})\,x_k \mid \boldsymbol{X}_{0:k-1} = \boldsymbol{A}_t(v, \boldsymbol{z}_{0:k-1})\right)$$
$$= \Phi\{(q_{kk}/2)^{1/2}\,x_k\}. \tag{60}$$

The convergence in (60) holds uniformly on compact sets in the variable $z_{0:k-1}$ by continuous convergence (see Section 0.1 in [25]). That is, (60) holds true if we replace $z_{0:k-1}$ by $z_{0:k-1}(v)$ satisfying $z_{0:k-1}(v) \to z_{0:k-1}$ as $v \to \infty$, and since the limit function is continuous in $z_{0:k-1}$ (constant function), the argument follows. Additionally, we have that for any $\boldsymbol{\alpha}_{t-k:t-1} \in (0, 1]^k$,

$$\lim_{v\to\infty} v^{-1/2} \left[a\big(\boldsymbol{\alpha}_{t-k:t-1} v + v^{1/2} z_{0:k-1}\big) - a(\boldsymbol{\alpha}_{t-k:t-1} v)\right] = \nabla a(\boldsymbol{\alpha}_{t-k:t-1}) \cdot z_{0:k-1},$$

$$\lim_{v\to\infty} v^{-1/2} b\big(\boldsymbol{\alpha}_{t-k:t-1} v + v^{1/2} z_{0:k-1}\big) = b(\boldsymbol{\alpha}_{t-k:t-1}),$$

where both convergences hold uniformly on compact sets in the variable $z_{0:k-1}$, since monotone increasing functions (in every argument) converge pointwise to a continuous limit. Thus, Assumption 3 holds true for the special case $t = k$ with

$$a_t(v) = \left\{ \sum_{i=0}^{k-1} \left(-\frac{q_{ik}}{q_{kk}} \right) \rho_{t-k+i}^{1/2} \right\}^2 v \quad \text{and} \quad b_t(v) = v^{1/2}.$$

Finally, observe that the entire argument after (58) remains unchanged if we change $t = k$ to $t = k + 1$. The claim is proved through iteration.

A.6. Convergence of multivariate inverted logistic full conditional distribution

The transition probability kernel of this process is given by (35) with $V(\boldsymbol{x}) = \|\boldsymbol{x}^{-1/\alpha}\|^\alpha$, $\boldsymbol{x} \in \mathbb{R}_+^{k+1}$. For $t \geq k \geq 1$ and $z_{0:k-1} \in \mathbb{R}_+^k$, we have that $\mathbb{P}(X_k/b(X_{0:k-1}) < 1 \mid X_{0:k-1} = \boldsymbol{B}_t(v, z_{0:k-1}))$ is equal to

$$\mathscr{L}(v, z_{0:k-1}) \exp \left\{ \left\| \boldsymbol{B}_t(v, \boldsymbol{x}_{0:k-1})^{1/\alpha} \right\|^\alpha - \left\| (\boldsymbol{B}_t(v, z_{0:k-1}), b\{\boldsymbol{B}_t(v, z_{0:k-1})\})^{1/\alpha} \right\|^\alpha \right\}, \quad (61)$$

where $\boldsymbol{B}_t(v, z_{0:k-1}) = (B_{t-k}(v, z_0), \ldots, B_{t-1}(v, z_{k-1}))$ and $\mathscr{L}(v, z_{0:k-1}) = 1 + o(1)$ for all $z_{0:k-1} \in \mathbb{R}_+^k$ as $v \to \infty$.

Now let $t = k$ and set $z_0 = 1$. First, we seek a function b such that the conditional probability in Equation (61) converges to a number $p \in (0, 1)$. Suppose that this function satisfies $b(\boldsymbol{b}_{0:k-1}(v)) \to \infty$ as $v \to \infty$ with $b(\boldsymbol{b}_{0:k-1}(v)) = o(v)$. Under this assumption, we have that as $v \to \infty$,

$$\log \overline{\pi}^{\text{inv}}[\boldsymbol{B}_t(v, z_{0:k-1}), b\{\boldsymbol{B}(v, z_{0:k-1})\}]$$

$$= \|\{\boldsymbol{B}_t(v, z_{0:k-1})\}^{1/\alpha}\|^\alpha - \|[\{\boldsymbol{B}_t(v, z_{0:k-1})\}^{1/\alpha}, \{b(\boldsymbol{B}_t(v, z_{0:k-1}))\}^{1/\alpha}]\|^\alpha + o(1)$$

$$= \left(\sum_{i=t-k}^{t-1} B_i(v, z_i)^{1/\alpha} + [b\{\boldsymbol{B}_t(v, z_{0:k-1})\}]^{1/\alpha} \right)^\alpha - \left(\sum_{i=t-k}^{t-1} B_i(v, z_i)^{1/\alpha} \right)^\alpha + o(1)$$

$$= \left(\sum_{i=t-k}^{t-1} B_i(v, z_i)^{1/\alpha} \right)^\alpha \left(1 + \frac{[b\{\boldsymbol{B}_t(v, z_{0:k-1})\}]^{1/\alpha}}{\sum_{i=t-k}^{t-1} B_i(v, z_i)^{1/\alpha}} \right)^\alpha - \left(\sum_{i=t-k}^{t-1} B_i(v, z_i)^{1/\alpha} \right)^\alpha + o(1)$$

$$= \alpha \left(\sum_{i=t-k}^{t-1} B_i(v, z_i)^{1/\alpha} \right)^{-(1-\alpha)} [b\{\boldsymbol{B}_t(v, z_{0:k-1})\}]^{1/\alpha} + o(1).$$

This expression converges to a positive constant provided

$$b(\boldsymbol{B}(v, z_{0:k-1})) = \mathcal{O}\left[\left\{ \sum_{i=0}^{k-1} B_t(v, z_i)^{1/\alpha} \right\}^{\alpha(1-\alpha)} \right] \quad \text{as } v \to \infty.$$

Hence, choosing b equal to $b(\boldsymbol{y}) = \|\boldsymbol{y}^{1/\alpha}\|^{\alpha(1-\alpha)}, \boldsymbol{y} \in \mathbb{R}_+^k$, gives the conditional probability (61) converging to $p = 1 - \exp(-\alpha)$, that is,

$$\pi^{\text{inv}}[\boldsymbol{B}_t(v, z_{0:k-1}), b\{\boldsymbol{B}(v, z_{0:k-1})\}] \to 1 - \exp(-\alpha), \qquad \alpha \in (0, 1),$$

and generally, we also have that for any $x_k \in \mathbb{R}_+$,

$$\lim_{v \to \infty} \pi^{\text{inv}}(\boldsymbol{B}(v, z_{0:k-1}), b(\boldsymbol{B}(v, z_{0:k-1})) z_k) = 1 - \exp\left(-\alpha x_k^{1/\alpha}\right). \tag{62}$$

Lastly, we note that the convergence in (62) holds uniformly on compact sets in the variable $\boldsymbol{x}_{0:k-1} \in [\delta_1, \infty) \times \cdots \times [\delta_k, \infty)$ by continuous convergence (see Section 0.1 in [25]). That is, (62) holds true if we replace $z_{0:k-1}$ by $z_{0:k-1}(v)$ satisfying $z_{0:k-1}(v) \to z_{0:k-1} \in [\delta_1, \infty) \times \cdots \times [\delta_k, \infty)$ as $v \to \infty$, and since the limit function is continuous in $z_{0:k-1}$ (constant function), the argument follows.

Let β_t satisfy the recurrence relation $\log \beta_t = \log(1 - \alpha) + \log(\max_{i=1,\dots,k} \beta_{t-i})$ subject to $\beta_i = 1 - \alpha$ for $i = 1, \dots, k-1$. For all $\delta_1, \dots, \delta_k > 0$ and $z_{0:k-1} \in [\delta_1, \infty) \times \dots \times [\delta_k, \infty)$,

$$v^{-\beta_t} b(\boldsymbol{B}_t(v, z_v)) \to \psi_t^b(z) \quad \text{whenever } z_v \to z \text{ as } v \to \infty,$$

where $\psi_t^b > 0$ is continuous and has the same form as in Corollary 3. Thus, Assumption 5 holds for the special case $t = k$ with $b_t(v) = v^{\beta_t}$.

Finally, observe that the entire argument after (61) remains unchanged if we change $t = k$ to $t = k+1$. The claim is proved through iteration.

A.7. Convergence of max-stable full conditional distribution—no mass on boundary

Suppose that $a(v\boldsymbol{1}_k) \to \infty$ as $v \to \infty$. Let Π_{k-1} denote the set of partitions of $[k-1] = \{0, \dots, k-1\}$. Then, for $z_{0:k-1} \in \mathbb{R}^k$ and with some rearrangement, $\mathbb{P}(X_k < a(X_{0:k-1}) \mid X_{0:k-1} = \boldsymbol{A}_t(v, z_{0:k-1}))$ is equal to

$$\frac{1 + \sum_{p \in \Pi_{k-1} \setminus [k-1]} (-1)^{|p|} \{\prod_{J \in p} V_J(\boldsymbol{y}_{0:k})\} / V_{0:k-1}(\boldsymbol{y}_{0:k})}{1 + \sum_{p \in \Pi_{k-1} \setminus [k-1]} (-1)^{|p|} \{\prod_{J \in p} V_J(\boldsymbol{y}_{0:k-1}, \infty)\} / V_{0:k-1}(\boldsymbol{y}_{0:k-1}, \infty)}$$

$$\times \frac{V_{0:k-1}(\boldsymbol{y}_{0:k})}{V_{0:k-1}(\boldsymbol{y}_{0:k-1}, \infty)} \times \exp\{V(\boldsymbol{y}_{0:k-1}, \infty) - V(\boldsymbol{y}_{0:k})\}, \tag{63}$$

where

$$\boldsymbol{y}_{0:k-1} = -1/\log[1 - \exp\{-\boldsymbol{A}_t(v, z_{0:k-1})\}] \quad \text{and}$$
$$y_k = -1/\log[1 - \exp\{-a(\boldsymbol{A}_t(v, z_{0:k-1}))\}].$$

Since V_J is a $-(|J| + 1)$-homogeneous function [4], it follows that

$$\frac{\prod_{J \in p} V_J(\boldsymbol{y}_{0:k})}{V_{[k-1]}(\boldsymbol{y}_{0:k})} = \mathcal{O}(\exp\{(1 - |p|) v\}), \quad \frac{\prod_{J \in p} V_J(\boldsymbol{y}_{0:k-1}, \infty)}{V_{[k-1]}(\boldsymbol{y}_{0:k-1}, \infty)} = \mathcal{O}(\exp\{(1 - |p|) v\}), \tag{64}$$

as $v \to \infty$. Because $|p| \geq 2$ for any $p \in \Pi_{k-1} \setminus [k-1]$, it follows that the first fraction in (63) converges to unity as $v \to \infty$, whereas the homogeneity property of the exponent function V

also guarantees that the last term in (63) converges to unity since $V(\mathbf{y}_{0:k-1}, \infty) = \mathcal{O}\{\exp(-v)\}$ and $V(\mathbf{y}_{0:k}) = \mathcal{O}\{\exp(-v)\}$. This leads to

$$\mathbb{P}(X_k < a(\mathbf{X}_{0:k-1}) \mid \mathbf{X}_{0:k-1} = \mathbf{A}_t(v, \mathbf{z}_{0:k-1})) = \frac{V_{[0:k-1]}(\mathbf{y}_{0:k-1}, y_k)}{V_{[0:k-1]}(\mathbf{y}_{0:k-1}, \infty)}(1 + o(1)), \qquad (65)$$

as $v \to \infty$. Therefore, for any functional a satisfying the property (19), we have

$$\lim_{v \to \infty} \frac{V_{0:k-1}(\mathbf{y}_{0:k-1}, y_k)}{V_{0:k-1}(\mathbf{y}_{0:k-1}, \infty)} = \frac{V_{0:k-1}[\exp(\mathbf{z}_{0:k-1}), \exp\{a(\mathbf{z}_{0:k-1})\}]}{V_{0:k-1}[\exp\{\mathbf{z}_{0:k-1}, \infty\}]}. \qquad (66)$$

Similarly, for $x_k \in \mathbb{R}$, $\mathbb{P}(X_k < a(\mathbf{X}_{0:k-1}) + x_k \mid \mathbf{X}_{0:k-1} = \mathbf{A}_t(v, \mathbf{z}_{0:k-1}))$ converges to K given by (42). The convergence in (66) holds uniformly on compact sets in the variable $\mathbf{z}_{0:k-1}$ by continuous convergence (see Section 0.1 in [25]). That is, (66) holds true if we replace $\mathbf{z}_{0:k-1}$ by $\mathbf{z}_{0:k-1}(v)$ satisfying $\mathbf{z}_{0:k-1}(v) \to \mathbf{z}_{0:k-1}$ as $v \to \infty$, and since the limit function is continuous in $\mathbf{z}_{0:k-1}$, the argument follows.

A.8. Convergence of logistic full conditional distribution

Under Assumption 3, (65) implies that for all $\mathbf{z}_{0:k-1} \in \mathbb{R}^k$,

$$\mathbb{P}(X_k < a(\mathbf{X}_{0:k-1}) \mid \mathbf{X}_{0:k-1} = \mathbf{A}_t(v, \mathbf{z}_{0:k-1})) = \left[1 + \frac{\exp\{-a(\mathbf{y}_{0:k-1})/\alpha\}}{\|\exp(-\mathbf{y}_{0:k-1}/\alpha)\|}\right]^{\alpha-k}$$
$$\times (1 + o(1))$$

as $v \to \infty$, where $\mathbf{y}_{0:k-1} = -1/\log[1 - \exp\{-\mathbf{A}_t(v, \mathbf{z}_{0:k-1})\}]$. Choosing a to be $a(\mathbf{z}_{0:k-1}) = -\alpha \log\{\|\exp(-\mathbf{z}_{0:k-1}/\alpha)\|\}$, we see that

$$\lim_{v \to \infty} \mathbb{P}(X_k < a(\mathbf{X}_{0:k-1}) \mid \mathbf{X}_{0:k-1} = \mathbf{A}_v(\mathbf{z}_{0:k-1})) = 2^{\alpha-k} \in (0, 1),$$

and more generally, for any $x_k \in \mathbb{R}$,

$$\lim_{v \to \infty} \mathbb{P}(X_k < a(\mathbf{X}_{0:k-1}) + x_k \mid \mathbf{X}_{0:k-1} = \mathbf{A}_t(v, \mathbf{z}_{0:k-1})) = \{1 + \exp(-x_k/\alpha)\}^{\alpha-k}.$$

The limit distribution does not depend on $\mathbf{z}_{0:k-1}$ since a satisfies the property (19).

A.9. Convergence of Hüsler–Reiss full conditional distribution

Equation (15) of [34] and (65) imply that for all $\mathbf{z}_{0:k-1} \in \mathbb{R}^k$,

$$\mathbb{P}(X_k < a(\mathbf{X}_{0:k-1}) \mid \mathbf{X}_{0:k-1} = \mathbf{A}_v(\mathbf{z}_{0:k-1}))$$
$$= \Phi[\tau^{-1}\{a(\mathbf{A}_v(\mathbf{z}_{0:k-1})) - \mu(\mathbf{A}_v(\mathbf{z}_{0:k-1}))\}](1 + o(1))$$

as $v \to \infty$. Here Φ denotes the cumulative distribution function of the standard normal distribution, and $\mu(\mathbf{y}_{0:k-1}) = -\tau\left(\mathbf{K}_{01}^\top \mathbf{C} \mathbf{K}_{10} \cdot \mathbf{y}_{0:k-1} + \mathbf{K}_{01}^\top \mathbf{\Sigma}^{-1} \mathbf{1}_{k+1}^\top / \mathbf{1}_{k+1}^\top \mathbf{q}\right)$, where $\tau^{-1} = \mathbf{K}_{01}^\top \mathbf{C} \mathbf{K}_{01}$, $\mathbf{C} = \left(\mathbf{\Sigma}^{-1} - \mathbf{q}\mathbf{q}^\top / \mathbf{1}_{k+1}^\top \mathbf{q}\right)$ is a $(k+1) \times (k+1)$ matrix of rank k, $\mathbf{q} = \mathbf{\Sigma}^{-1} \mathbf{1}_{k+1}$, and

$$\mathbf{K}_{10} = \begin{pmatrix} \mathbf{I}_k \\ \mathbf{0}_{1,k} \end{pmatrix}, \quad \mathbf{K}_{01} = \begin{pmatrix} \mathbf{0}_{k,1} \\ 1 \end{pmatrix}.$$

Choosing a to be $a(z_{0:k-1}) = -\tau K_{01}^{\top} C K_{10} \cdot z_{0:k-1}$, we see that for any $x_k \in \mathbb{R}$,

$$\lim_{v \to \infty} \mathbb{P}(X_k < a(X_{0:k-1}) + x_k \mid X_{0:k-1})$$
$$= A_t(v, z_{0:k-1})) = \Phi\big[\big\{x_k + \tau\big(K_{01}^{\top} \Sigma^{-1} \mathbf{1}_{k+1}^{\top}/\mathbf{1}_{k+1}^{\top} q\big)\big\}/\tau\big].$$

The limit distribution does not depend on $z_{0:k-1}$, since a satisfies the property (19). The latter follows from the properties $K_{10} \cdot \mathbf{1}_k = \mathbf{1}_{k+1} - K_{01}$ and $C \cdot \mathbf{1}_{k+1} = \mathbf{0}_{k+1,1}$, which give $-\tau K_{01}^{\top} C K_{10} \cdot \mathbf{1}_k = 1$.

Acknowledgements

Funding information

Competing interests

References

[1] ASENOVA, S. AND SEGERS, J. (2023). Extremes of Markov random fields on block graphs: max-stable limits and structured Hüsler–Reiss distributions. *Extremes* **26**, 433–468.

[2] BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J. AND TEUGELS, J. (2004), *Statistics of Extremes, Theory and Applications*. John Wiley, New York.

[3] CHIAPINO, M., SABOURIN, A. AND SEGERS, J. (2019). Identifying groups of variables with the potential of being large simultaneously. *Extremes* **22**, 193–222.

[4] COLES, S. G. AND TAWN, J. A. (1991). Modelling extreme multivariate events. *J. R. Statist. Soc. B [Statist. Methodology]* **53**, 377–392.

[5] ENGELKE, S. AND HITZ, A. S. (2020). Graphical models for extremes (with discussion). *J. R. Statist. Soc. B [Statist. Methodology]* **82**, 871–932.

[6] ENGELKE, S. AND IVANOVS, J. (2021). Sparse structures for multivariate extremes. *Ann. Rev. Statist. Appl.* **8**, 241–270.

[7] ENGELKE, S., MALINOWSKI, A., KABLUCHKO, Z. AND SCHLATHER, M. (2015). Estimation of Hüsler-Reiss distributions and Brown-Resnick processes. *J. R. Statist. Soc. B. [Statist. Methodology]* **77**, 239–265.

[8] HEFFERNAN, J. E. AND RESNICK, S. I. (2007). Limit laws for random vectors with an extreme component. *Ann. Appl. Prob.* **17**, 537–571.

[9] HEFFERNAN, J. E. AND TAWN, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *J. R. Statist. Soc. B [Statist. Methodology]* **66**, 1–34.

[10] HUSER, R. AND DAVISON, A. C. (2013). Composite likelihood estimation for the Brown–Resnick process. *Biometrika* **100**, 511–518.

[11] JANSSEN, A. AND SEGERS, J. (2014). Markov tail chains. *J. Appl. Prob.* **51**, 1133–1153.

[12] JOE, H. (2015). *Dependence Modeling with Copulas*. CRC Press, Boca Raton, FL.

[13] KEEF, C., PAPASTATHOPOULOS, I. AND TAWN, J. A. (2013). Estimation of the conditional distribution of a multivariate variable given that one of its components is large: additional constraints for the Heffernan and Tawn model. *J. Multivariate Anal.* **115**, 396–404.

[14] KULIK, R. AND SOULIER, P. (2015). Heavy tailed time series with extremal independence. *Extremes* **18**, 1–27.

[15] LEADBETTER, M. R., LINDGREN, G. AND ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Series*. Springer, New York.

[16] LEDFORD, A. W. AND TAWN, J. A. (1997). Modelling dependence within joint tail regions. *J. R. Statist. Soc. B [Statist. Methodology]* **59**, 475–499.

[17] LEMMENS, B. AND NUSSBAUM, R. (2012). *Nonlinear Perron–Frobenius Theory*. Cambridge University Press.

[18] MAULIK, K., RESNICK, S. AND ROOTZÉN, H. (2002). Asymptotic independence and a network traffic model. *J. Appl. Prob.* **39**, 671–699.

[19] O'BRIEN, G. L. (1987). Extreme values for stationary and Markov sequences. *Ann. Prob.* **15**, 281–291.

[20] PAPASTATHOPOULOS, I., STROKORB, K., TAWN, J. A. AND BUTLER, A. (2017). Extreme events of Markov chains. *Adv. Appl. Prob.* **49**, 134–161.

[21] PAPASTATHOPOULOS, I. AND TAWN, J. A. (2016). Conditioned limit laws for inverted max-stable processes. *J. Multivariate Anal.* **150**, 214–228.

[22] PERFEKT, R. (1994). Extremal behaviour of stationary Markov chains with applications. *Ann. Appl. Prob.* **4**, 529–548.

[23] PERFEKT, R. (1997). Extreme value theory for a class of Markov chains with values in $\mathbb{R}^d$. *Adv. Appl. Prob.* **29**, 138–164.

[24] PICKANDS, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3**, 119–131.

[25] RESNICK, S. I. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer, New York.

[26] RESNICK, S. I. AND ZEBER, D. (2013). Asymptotics of Markov kernels and the tail chain. *Adv. Appl. Prob.* **45**, 186–213.

[27] ROOTZÉN, H. (1988). Maxima and exceedances of stationary Markov chains. *Adv. Appl. Prob.* **20**, 371–390.

[28] SCHLATHER, M. AND TAWN, J. A. (2003). A dependence measure for multivariate and spatial extreme values: properties and inference. *Biometrika* **90**, 139–156.

[29] SEGERS, J. (2020). One- versus multi-component regular variation and extremes of Markov trees. *Adv. Appl. Prob.* **52**, 855–878.

[30] SIMPSON, E. S., WADSWORTH, J. L. AND TAWN, J. A. (2020). Determining the dependence structure of multivariate extremes. *Biometrika* **107**, 513–532.

[31] SMITH, R. L. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statist. Sci.* **4**, 367–377.

[32] SMITH, R. L. (1992). The extremal index for a Markov chain. *J. Appl. Prob.* **29**, 37–45.

[33] TAWN, J. A. (1990). Modelling multivariate extreme value distributions. *Biometrika* **77**, 245–253.

[34] WADSWORTH, J. L. AND TAWN, J. A. (2014). Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika* **101**, 1–15.

[35] WALKER, G. (1931). On periodicity in series of related terms. Proc. R. Soc. London A 131, 518–532.

[36] WINTER, H. C. AND TAWN, J. A. (2015). Modelling heatwaves in central France: a case-study in extremal dependence. *J. R. Statist. Soc. C [Appl. Statist.]* **65**, 345–365.

[37] WINTER, H. C. AND TAWN, J. A. (2017). kth-order Markov extremal models for assessing heatwave risks. *Extremes* **20**, 393–415.

[38] YULE, G. U. (1927). On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Phil. Trans. R. Soc. London A* **226**, 267–298.

[39] YUN, S. (1998). The extremal index of a higher-order stationary Markov chain. *Ann. Appl. Prob.* **8**, 408–437.