Extreme events of higher-order Markov chains: hidden tail chains and extremal Yule–Walker equations

Ioannis Papastathopoulos* and Jonathan A. Tawn†

* School of Mathematics and Maxwell Institute, University of Edinburgh, Edinburgh, EH9 3FD
† Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF
i.papastathopoulos@ed.ac.uk j.tawn@lancaster.ac.uk

Abstract

We derive some key extremal features for $k$th order Markov chains, which can be used to understand how the process moves to and fro between the body of the process and an extreme state. The chains are studied given that there is an exceedance of a threshold, as the threshold tends to the upper endpoint of the distribution. The extremal properties of the Markov chain at lags up to $k$ are determined by the kernel of the chain, through a joint initialisation distribution, with the subsequent values determined by the conditional independence structure through a transition behaviour. We study the extremal properties of each of these elements under weak assumptions for broad classes of extremal dependence structures. We find that it is possible to find a simple affine normalization, dependent on the threshold excess, such that non-degenerate limiting behaviour of the process is assured for all lags. These normalization functions have an interesting structure that has a striking parallel to the Yule-Walker equations. Furthermore, the limiting process is always linear in the innovations. We illustrate the results with the study of $k$th order stationary Markov chains based on widely studied families of $k+1$ dimensional copula.

Key-words: conditional extremes; conditional independence; Markov chains; tail chains; extremal Yule–Walker equations

MSC subject classifications: 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 generalised to include the situation where the variables are no longer independent, as for example in the monograph of Leadbetter et al. (1983) 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 Markov chains which provide the backbone of a broad range of statistical models and automatically satisfy the Leadbetter et al. (1983) long-range dependence conditions. Such models have attracted considerable interest in the analysis of extremes of stochastic processes, by considering the behaviour of the process when it is extreme, i.e., when it exceeds a high threshold. Rootzén (1988) 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, can be derived as the threshold converges to the upper endpoint of the marginal distribution.

Although powerful, this approach only reveals the behaviour of the chain whilst it remains at the same level of marginal extremity as the threshold, and therefore is only informative about clustering for asymptotically dependent processes. For example, for any asymptotically independent Markov process, e.g., for a Gaussian Markov process, this limit theory describes each extreme event as a single observation. It is critical to understand better the behaviour of a Markov chain within an extreme event under less restrictive conditions through using more sophisticated limiting mechanisms, which allows us to characterise the event as it moves to and fro between the body of the distribution and an extreme state. In the case of first-order Markov chains, Papastathopoulos et al. (2017) treat both asymptotically dependent and asymptotically independent chains in a unified theory. The focus of this paper is similar, but this time on higher-order Markov chains.

To help illustrate the complexity in higher-order Markov chains, consider standard measures of extremal dependence (Ledford and Tawn, 1997). When analysing the extremal behaviour of real-valued process \( \{X_t : t = 0, 1, 2, \ldots \} \) with marginal distribution \( F_t \), one has to distinguish between two classes of extremal dependence. Let \( D = \{1, \ldots, k\} \), where \( k \in \mathbb{N} \) is the order of the Markov chain, and define \( M = 2^D \setminus \{\emptyset\} \). The two classes can be characterized through the quantities

\[
\chi_A = \lim_{u \to \infty} \mathbb{P}(\{F_j(X_j) > u\}_{j \in A} | F_0(X_0) > u) \quad \text{for } A \in M, \tag{1}
\]

assuming that such limits exist. When \( \chi_A > 0 \) for some \( A \in M \) (\( \chi_A = 0 \) for all \( A \in M \)) the process is said to be **asymptotically dependent** (asymptotically independent). Bounds can be obtained for the coefficients, e.g., for any \( A_1, A_2 \in M \), \( \chi_{A_1 \cup A_2} \leq \min(\chi_{A_1}, \chi_{A_2}) \). However the coefficients can admit apparently complicated structure, for instance, if \( \chi_A > 0 \) for any \( A \in M \) then it is possible that \( \chi_B = 0 \) for any \( B \supseteq A \). Similar structure manifests in asymptotically independent chains through related coefficients of asymptotic independence (Ledford and Tawn, 2003). In a \( k \)th order Markov chain initialized from an extreme state, there are \( 2^k \) combinations of joint extremal states of the chain over the next \( k \) consecutive values. Hence there is a combinatorial explosion of the number of types of extreme events that are possible. Papastathopoulos et al. (2017) needed to only consider two types when \( k = 1 \).

Previous work on \( k \)th order Markov chains considers only the case where \( \chi_D > 0 \). Although we cover both \( \chi_D > 0 \) and \( \chi_D = 0 \), here we primarily focus on the latter case where \( \chi_A = 0 \) for all \( A \in M \) and also consider cases where \( \chi_A > 0 \) and \( \chi_B = 0 \), for at least one \( A, B \in M \setminus D \). To the best of our knowledge, important limiting characteristics of the limiting tail chain of higher-order Markov chains have not been dealt with in depth, yet these are crucial for understanding the evolution of random processes and for providing well-founded parametric models that can be used for statistical inference, prediction and assessment of risk.

To derive greater detail about the behaviour within extreme events for Markov chains we need to explore the properties of the **hidden tail chain** where a hidden tail chain describes the nature of the Markov chain after witnessing an extreme event. This is expressed in the limit as the state tends to the upper endpoint of the marginal distribution of \( X_t \). The distinction between the hidden tail chain and the usual tail chain (Resnick and Zeber, 2013; Janßen and Segers, 2014) is explained below. For higher-order chains there are few results, e.g., Perfekt (1997); Yun (1998); Janßen and Segers (2014), and these are restricted to asymptotically dependent processes and tail chains only.

Almost all the above mentioned results have been derived under stationarity and regular variation assumptions on the marginal distribution, rescaling the Markov chain by the extreme observation, resulting in the tail chain being a multiplicative random walk. More recently, Papastathopoulos et al. (2017) treat asymptotically independent first-order Markov chains with marginal distributions with exponential-like tails, in the Gumbel domain of attraction as this reveals structure, for asymptotically independent processes, i.e., not apparent through regularly varying marginals.
Let the stochastic process \( \{Y_t := (X_t, \ldots, X_{t+k-1}) : t = 0, 1, \ldots \} \), then \( Y_t \) is a homogeneous \( \mathbb{R}^k \)-valued Markov chain. The stochastic process \( Y_t \) can be represented as a stochastic difference equation of the form

\[
Y_t = g(Y_{t-1}; w_t) \quad t = 1, 2, \ldots
\]

(2)

where \( \{w_t : t = 0, 1, \ldots \} \) is a sequence of univariate i.i.d. random elements of a measurable space \((E, \mathcal{E})\) that is independent of \( Y_0 \), and \( g \) is a measurable function from \( \mathbb{R}^k \times E \to \mathbb{R}^k \) (Kifer, 2012). If there exist 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

\[
\left\{ \frac{X_t - a_t(X_0)}{b_t(X_0)} : t = 1, \ldots, k - 1 \right\} \bigg| X_0 > u
\]

converges weakly to a process that is non-degenerate in each component, as \( u : F_0(u) \to 1 \), then our aim is to find conditions under which, additional functions \( a_t : \mathbb{R} \to \mathbb{R} \) and \( b_t : \mathbb{R} \to \mathbb{R}_+ \) for all \( t = k, k + 1, \ldots \) such that

\[
\left\{ \frac{X_t - a_t(X_0)}{b_t(X_0)} : t = 1, 2, \ldots \right\} \bigg| X_0 > u \overset{w}{\to} \{M_t : t = 1, 2, \ldots \}
\]

where each \( M_t \) is non-degenerate and \( \{M_t : t = 1, 2, \ldots \} \) is termed the hidden tail chain. Note this is not the tail chain studied by Janßen and Segers (2014), as in that treatment, the norming functions are restricted there to \( a_t(x) = x \) and \( b_t(x) = 1 \) for all \( t \), and any \( M_t \) can be degenerate at \( \{-\infty\} \). Our target is to find how the first \( k - 1 \) norming functions \( a_t(\cdot) \) and \( b_t(\cdot) \), control those where \( t \geq k \) and to identify the transition dynamics of the hidden tail chain. It is important to characterise how the dynamics of the hidden tail chain encode information about how the process changes along its index and state space. Under weak conditions, we make the surprising finding that, whatever the form of \( g \), in equation (2), we can always express \( M_{t+1} \) in the form

\[
M_{t+1} = \psi_t^a(M_{t,k}) + \psi_t^b(M_{t,k}) \varepsilon_t \quad \text{for } t > k,
\]

(3)

where \( M_{t,k} = \{	ext{M}_{t-k}, \ldots, M_{t-1}\} \), \( \psi_t^a : \mathbb{R}^k \to \mathbb{R} \), \( \psi_t^b : \mathbb{R}^k \to \mathbb{R}_+ \) are update functions and \( \{\varepsilon_t : t = 1, 2, \ldots\} \) is a sequence of non-degenerate i.i.d. innovations. This simple structure for the hidden tail chain is controlled through the update functions which we show take particular classes of forms. The parallels between the extremal properties of the norming functions and the Yule–Walker equations, used in standard time-series analysis (Yule, 1927; Walker, 1931), will be shown to be striking.

**Organization of the paper.** In Section 2, we state our main theoretical results for higher-order tail chains with affine update functions under rather broad assumptions on the extremal behaviour of both asymptotically dependent and asymptotically independent Markov chains. As in previous accounts (Perfekt (1994); Resnick and Zeber (2013); Janßen and Segers (2014), Kulik and Soulier (2015) and Papastathopoulos et al. (2017), our results only need the homogeneity (and not the stationarity) of the Markov chain and therefore, we state our results in terms of homogeneous Markov chains with initial distribution \( F_0 \). In Section 3 we study hidden tail chains of asymptotically independent and asymptotically dependent Markov chains with standardized marginal distributions. Subsequently, in Section 4 we characterise 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 joint distributions. All proofs are postponed to Section 6.
Notation. Throughout this text, we use the following notation. For a topological space $E$ we denote its Borel-$\sigma$-algebra by $\mathcal{B}(E)$. The set of bounded continuous functions on $E$ is $C_b(E)$. If $f_n, f$ are real-valued functions on $E$, we say that $f_n$ (resp. $f_n(x)$) converges uniformly on compact sets (in the variable $x \in E$) 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, $f_n$ (resp. $f_n(x)$) is said to converge uniformly on compact sets to $\infty$ (in the variable $x \in E$) if $\inf_{x \in C} f_n(x) \to \infty$ for compact sets $C \subset E$. Weak convergence of measures on $E$ is abbreviated by $\xrightarrow{w} \infty$, yet, by abuse of language, we also use $X_n \xrightarrow{w} X$ when the corresponding measures of the random variables converge in the weak topology. When $K$ is a distribution on $\mathbb{R}$, we write $K(x)$ instead of $K((-\infty, x])$. By saying that a distribution is supported on $A \subseteq \mathbb{R}^n$, for some $n \in \mathbb{N}$, we do not allow the distribution to have mass at the boundary $\partial A$ of $A$. If $F_0$ is a distribution function, we abbreviate its survivor function by $\overline{F}_0 = 1 - F_0$. The quantile function of a univariate distribution $F$ is denoted by $F^{-1}(u) = \{ x : F(x) \geq u \}$. The relation $\sim$ stands for “is distributed like”.

Vectors of length greater than one are typeset in bold, usually $x \in \mathbb{R}^k$, $k > 1$. Vector algebra is used throughout the paper. For a sequence of measurable functions $\{g_t\}$, the notation $g_{t,k}(x)$, for $k \in \mathbb{N}$, is used to denote $(g_{t-1}(x), \ldots, g_{t-k}(x))$. We denote by $\Pi_k$ the set of partitions of the set $\{1, \ldots, k\}$. We use the standard notation $\|x\|_p$ for the $L_p$ norm of a vector $x$ in the $k$-dimensional Euclidean space. For $J \subseteq \{0, 1, \ldots, k\}$ and a differentiable function $V : \mathbb{R}^{k+1}_+ \to \mathbb{R}_+$, we denote by $V_J$ the higher order partial derivative $\partial^{|J|} V(x) / \prod_{j \in J} \partial x_j$. For a $k$-dimensional real vector $x = (x_0, \ldots, x_{k-1})$ and real valued $y$ we define

$$V(x, \infty) := \lim_{y \to \infty} V(x, y) \quad \text{and} \quad V_{k-1,k}(x, \infty) := \frac{\partial^k}{\prod_{j=0}^{k-1} \partial x_j} \lim_{y \to \infty} V(x, y),$$

assuming that the limit exists for all $x \in \mathbb{R}^k_+$. By saying that a distribution is supported on $A \subseteq \mathbb{R}^n$, for some $n \in \mathbb{N}$, we do not allow the distribution to have mass at the boundary $\partial A$ of $A$. For a Cartesian coordinate system $\mathbb{R}^k$ with coordinates $x_1, \ldots, x_k$, $\nabla$ is defined by the partial derivative operators as

$$\nabla = \sum_{i=1}^k \frac{\partial}{\partial x_i} e_i$$

for an orthonormal basis $\{e_1, \ldots, e_k\}$. The standard infix operator $x \cdot y$ is used for the scalar product of two vectors $x, y \in \mathbb{R}^k$, i.e., $x \cdot y = \sum_{i=1}^k x_i y_i$. By convention, univariate functions on vectors are applied componentwise, e.g., if $f : \mathbb{R} \to \mathbb{R}, x \in \mathbb{R}^k$, then $T(x) = (T(x_1), \ldots, T(x_k))$. Lastly, the symbols $0$ and $1$ are used to denote the vectors $(0, \ldots, 0) \in \mathbb{R}^p$ and $(1, \ldots, 1) \in \mathbb{R}^p$, for some $p \in \mathbb{N}$.

2 Theory

2.1 Real valued chains with location and scale norming

Conditional extreme value theory requires standardization of the distribution of $X_0$, see for example Heffernan and Resnick (2007). In practice, standardization is performed on all marginal distributions of the process and this is typically achieved via the probability integral transform. To facilitate the generality of theoretical developments, our first assumption concerns only the extremal behaviour of $X_0$ and is the same throughout Section 2.

Assumption $A_0$. $F_0$ has upper end point $\infty$ and there exists a non-degenerate probability distribution $H_0$ on $[0, \infty)$ and a measurable norming function $\sigma(u) > 0$, such that

$$\frac{F_0(u + \sigma(u)dx)}{1 - F_0(u)} \xrightarrow{w} H_0(dx) \quad \text{as } u \to \infty.$$
However, in Sections 3 and 5 we assume that all one-dimensional marginal distributions of the Markov process are standardized to unit-rate exponential random variables, as this is the case with the clearest mathematical formulation, see Papastathopoulos et al. (2017). For \( \{Y_t := (X_t, \ldots, X_{t+k-1}) : t = 0, 1, \ldots\} \), let \( F_{Y_0}(y) = \Pr(Y_0 \leq y), y \in \mathbb{R}^k \) be the initial distribution of the process, and
\[
\pi(y, A) = \Pr(Y_1 \in A \mid Y_0 = y) \quad y \in \mathbb{R}^k, \quad A \in \mathcal{B}(\mathbb{R}^k),
\]
its transition kernel. The next assumption guarantees that the initial distribution of the process \( F_{Y_0} \), conditionally on an extreme state at time zero, converges weakly under appropriate location and scale norms.

**Assumption A_1.** *(behaviour of initial states in the presence of an extreme event)* If \( k > 1 \), there exist measurable functions \( a_t : \mathbb{R} \to \mathbb{R} \) and \( b_t : \mathbb{R} \to \mathbb{R}_+ \) for \( t = 1, \ldots, k - 1 \), and a distribution \( G \) supported on \( \mathbb{R}^{k-1} \), with non-degenerate margins, such that, as \( u \to \infty \), \( a_t(u) + b_t(u)x \to \infty \) for all \( t \) and \( x \in \mathbb{R} \) and
\[
P \left( \frac{X_0 - u}{\sigma(u)} \in dx, \frac{X_t - a_t(X_0)}{b_t(X_0)} \in dx_t : t = 1, \ldots, k - 1 \mid X_0 > u \right) \xrightarrow{w} H_0(dx) G \left( \prod_{t=1}^{k-1} dx_t \right).
\]

**Remark 1.** Assumption A_1 is a reformulation of the assumptions in Heffernan and Tawn (2004) who show that if \( X_0, \ldots, X_{k-1} \) have identical distributions and \( H_0 \) is the exponential distribution, then it holds widely and for all standard copula models considered in Joe (1997).

Assumption A_1 is required for Markov processes on \( \mathbb{R}^k \) with \( k > 1 \) and becomes inconsequential if \( k = 1 \) since then, all information about the initial distribution \( F_{Y_0} \) is contained in the conditioning event \( \{X_0 = u\} \). For \( k > 1 \), the assumption guarantees, in the distributional sense, initial conditions for the first \( k - 1 \) elements of the process after \( t = 0 \). Informally, for \( u \) sufficiently close to \( \sup \{x : F_0(x) < 1\} \), such initial conditions take the form \( X_t = a_t(X_0) + b_t(X_0) M_t \), for \( t = 1, \ldots, k - 1 \) and \( X_0 > u \), where \( (M_1, \ldots, M_{k-1}) \) is a random vector with distribution \( G \) that is independent of \( X_0 \). Subsequently, we observe that if the states of the process \( X_0, \ldots, X_{k-1} \) are initialised, then the stochastic difference equation (2) can be solved by repeated substitution. The difference equation determines the one step transition probability of the time series process since
\[
\pi(x_{t,k}, y) = \Pr \left( y(X_{t,k}; w_t) \leq y \mid X_{t,k} = x_{t,k} \right),
\]
so that in order to describe the behaviour of the next state in the Markov process, a condition on the transition kernel is required. This is guaranteed by our next assumption which asserts that the transition kernel converges weakly to a non-degenerate limiting distribution under appropriate location-scale functionals.

**Assumption A_2.** *(behaviour of the next state as the initial states become extreme)* There exist measurable functions \( a : \mathbb{R}^k \to \mathbb{R} \) and \( b : \mathbb{R}^k \to \mathbb{R}_+ \), and a non-degenerate distribution function \( K \) supported on \( \mathbb{R} \), such that as \( u \to \infty \), \( a(u 1) + b(u 1)x \to \infty \) for all \( x \in \mathbb{R} \), and, for all \( (x_1, \ldots, x_{k-1}) \in \mathbb{R}^k \)
\[
P \left( \frac{X_k - a(X_0, \ldots, X_{k-1})}{b(X_0, \ldots, X_{k-1})} \in dx \mid X_0 = u, \left\{ \frac{X_t - a_t(u)}{b_t(u)} = x_t \right\}_{t=1,\ldots,k-1} \right) \xrightarrow{w} K(dx).
\]
By conditioning on an extreme state at \( t = 0 \) and the initial conditions implied by assumption A_1, then assumption A_2 guarantees that the next state behaves as a random variable which is bounded in
probability and translated and dilated by location and scale functions respectively, of the $k$ previous states.

To motivate assumption $A_2$, it is instructive to consider how a solution for all time steps can be obtained. First, we make an ansatz that there exist sequences of norming functions $a_t$ and $b_t$ such that

$$\lim_{u \to \infty} \mathbb{P} \left( \frac{X_t - a_t(X_0)}{b_t(X_0)} \leq x \mid X_0 = u \right) = \mathbb{P}(M_t \leq x) \quad t = 1, 2, \ldots$$ \hspace{1cm} (4)

where $\{M_t : t = 1, 2, \ldots\}$ is independent of $X_0$ a.s. and non-degenerate in all components. For measurable functions $a_t : \mathbb{R} \mapsto \mathbb{R}$, $b_t : \mathbb{R} \mapsto \mathbb{R}_+$, for $t = k, k + 1, \ldots$, and $a : \mathbb{R}^k \mapsto \mathbb{R}$, $b : \mathbb{R}^k \mapsto \mathbb{R}_+$, we can always write

$$\frac{X_t - a_t(X_0)}{b_t(X_0)} = \frac{a(X_{t,k}) - a_t(X_0)}{b_t(X_0)} + \frac{b(X_{t,k})}{b_t(X_0)} \frac{X_t - a(X_{t,k})}{b(X_{t,k})}, \quad t \geq k.$$ \hspace{1cm} (5)

Here we will assume that the functions $a_t$ and $b_t$ are such that if limit (4) holds then, as $u \to \infty$, then all of the quotient terms in the right hand side of expression (5) converge weakly to random elements that are bounded in probability. This convergence is guaranteed when there is sufficient regularity in $a_t$ and $b_t$ and the functionals $a$ and $b$, which we make explicit below.

**Assumption $A_3$ (norming functions and update functions for the tail chain)**

(a) There exist measurable functions $a_t : \mathbb{R} \mapsto \mathbb{R}$ and $b_t : \mathbb{R} \mapsto \mathbb{R}_+$ such that $a_t(v) + b_t(v) \to \infty$, as $v \to \infty$, for all $x \in \mathbb{R}$ and for each time step $t = k, k + 1, \ldots$.

(b) Let $a_{t,k}(v) = (a_{t-k}(v), \ldots, a_{t-1}(v))$ and $b_{t,k}(v) = (b_{t-k}(v), \ldots, b_{t-1}(v))$. There exist continuous update functions

$$\psi_t^a(x) = \lim_{v \to \infty} \frac{a(a_{t,k}(v) + b_{t,k}(v)x) - a_t(v)}{b_t(v)} \in \mathbb{R}$$

$$\psi_t^b(x) = \lim_{v \to \infty} \frac{b(a_{t,k}(v) + b_{t,k}(v)x)}{b_t(v)} > 0$$

defined for $t = k, k + 1, \ldots$, such that the remainder terms

$$r_t^a(x, v) = \frac{a_t(v) - a(a_{t,k}(v) + b_{t,k}(v)x)}{b_t(v)}$$

$$r_t^b(x, v) = 1 - \frac{b_t(v)\psi_t^b(x)}{b(a_{t,k}(v) + b_{t,k}(v)x)}$$

converge to 0 as $v \to \infty$ and both convergences hold uniformly on compact sets in the variable $x \in \mathbb{R}^k$.

The functions $a$ and $b$ can be characterised to some extent from the conditional extreme value model of Resnick and Zeber (2014). In particular, up to isomorphic additive form, we can identify the pair of functions $(a, b)$ as those that belong to the class of multivariate extended regularly varying functions since, the requirements set up by assumption $A_3$, imply the existence of functions $\psi^a : \mathbb{R}^k \mapsto \mathbb{R}$ and $\psi^b : \mathbb{R}^k \mapsto \mathbb{R}_+$ such that, for $x \in \mathbb{R}^k$, the maps

$$\mathbb{R}^k \ni x \mapsto a(x) \quad \text{and} \quad \mathbb{R}^k \ni x \mapsto b(x),$$

satisfy

$$\lim_{u \to \infty} \frac{a(u 1 + x) - a(u 1)}{b(u 1)} = \psi^a(x) \quad \text{and} \quad \lim_{u \to \infty} \frac{b(u 1 + x)}{b(u 1)} = \psi^b(x) \hspace{1cm} (6)$$
It can be easily seen that this class contains the pairs of functions that are homogeneous of degree one and less than one respectively, but more importantly embodies a much wider range of possibilities which we explore in later sections.

With the above assumptions in place we are in position to state our main theorem.

**Theorem 1.** Let \( \{X_t : t = 0, 1, \ldots \} \) be a homogeneous \( k \)th order Markov chain satisfying assumptions \( A_0, A_1, A_2 \) and \( A_3 \). Then as \( v \to \infty \)

\[
\left( \frac{X_0 - v}{\sigma(v)}, \frac{X_1 - a_1(X_0)}{b_1(X_0)}, \ldots, \frac{X_t - a_t(X_0)}{b_t(X_0)} \bigg| X_0 > v \right)
\]

converges weakly to \((E_0, M_1, \ldots, M_t)\), where

(i) \( E_0 \sim H_0 \) and \((M_1, M_2, \ldots, M_t)\) are independent,

(ii) \((M_1, \ldots, M_{k-1}) \sim G\) and

\[
M_s = \psi^a_s(M_{s,k}) + \psi^k_s(M_{s,k}) \varepsilon_s, \quad s = k, k + 1, \ldots
\]

where \(M_{s,k} = (M_{s-k}, \ldots, M_{s-1})\), for an i.i.d. sequence of random variables \(\varepsilon_s \sim K\).

### 2.2 Nonnegative chains with only scale norming

In this section, we extend the results of Section 2.1 to nonnegative chains which require additional care due to \(a_t \equiv 0\), for \(t = 1, \ldots, k - 1\). Here, we still assume \(A_0\) but, similarly to Papastathopoulos et al. (2017), we introduce additional conditions that control the mass of the limiting renormalized initial distribution and the limiting renormalized transition kernel of the Markov process.

**Assumption B₁.** (behaviour of the next state as the previous states becomes extreme) There exist measurable functions \(b_t(v) > 0\) for \(t = 1, \ldots, k - 1\), and a distribution function \(G\) supported on \([0, \infty)^{k-1}\), with no mass at any of the open half-planes \(C_j = \{ (x_1, \ldots, x_{k-1}) \in [0, \infty)^{k-1} : x_j = 0 \}\), i.e., \(G(C_j) = 0\) for \(j = 1, \ldots, k - 1\), such that, as \(u \to \infty\),

\[
P \left( \frac{X_0 - u}{\sigma(u)} \in dx, \frac{X_t}{b_t(X_0)} \in dx_t : t = 1, \ldots, k - 1 \bigg| X_0 > u \right) \xrightarrow{w} H_0(dx) G \left( \prod_{t=1}^{k-1} dx_t \right).
\]

**Assumption B₂.** (behaviour of the next state as the initial states become extreme) There exists a measurable function \(b : \mathbb{R}^k \to \mathbb{R}_+\), and a non-degenerate distribution function \(K\) supported on \([0, \infty)\) with no mass at 0, i.e., \(K(\{0\}) = 0\), such that, as \(u \to \infty\), \(b(u 1) \to \infty\), and for all \((x_0, \ldots, x_{k-1}) \in \mathbb{R}_+^k\),

\[
P \left( \frac{X_k}{b(X_0, \ldots, X_{k-1})} \in dx \bigg| X_0 = u, \left\{ \frac{X_t}{b_t(u)} = x_t \right\}_{t=1, \ldots, k-1} \right) \xrightarrow{w} K(dx).
\]

**Assumption B₃.** (scaling function and update function for the tail chain)

(a) There exist measurable \(b_t(v) > 0\) for each time step \(t = k, k + 1 \ldots\) such that \(b_t(v) \to \infty\) as \(v \to \infty\) for all \(t = k, k + 1, \ldots\).
(b) Let \( b_{t,k}(v) = (b_{t-k}(v), \ldots, b_{t-1}(v)) \). There exist continuous update functions
\[
\psi^b_t(x_{t,k}) = \lim_{v \to \infty} \frac{b_t(b_{t,k}(v) x_{t,k})}{b_t(v)} > 0
\]
defined for \( x_{t,k} = (x_{t-k}, \ldots, x_{t-1}) \in (0, \infty)^k \) such that the remainder terms
\[
r^b_t(x_{t,k}, v) = 1 - \frac{b_t(v) \psi^b_t(x_{t,k})}{b_t(b_{t,k}(v) x_{t,k})} > 0
\]
converge to 0 as \( v \to \infty \) and both convergences hold uniformly on compact sets in the variable \( x_{t,k} \in [\delta_1, \infty) \times \cdots \times [\delta_k, \infty) \), for any \( \delta_1, \ldots, \delta_k > 0 \).

(c) Finally, we assume that \( \sup\{||x||_\infty : x \in A_c\} \to 0 \) as \( c \downarrow 0 \), where \( A_c = \{ x \in (0, \infty)^k : \psi^b_t(x) \leq c \} \) with the convention that \( \sup(\emptyset) = 0 \).

**Theorem 2.** Let \( \{X_t : t = 0, 1, \ldots\} \) be a homogeneous Markov chain satisfying assumptions \( A_0, B_1, B_2 \) and \( B_3 \). Then as \( u \to \infty \)
\[
\left( \frac{X_0 - u}{\sigma(u)}, \frac{X_1}{b_t(X_0)}, \ldots, \frac{X_t}{b_t(X_0)} \mid X_0 > u \right)
\]
converges weakly to \((E_0, M_1, \ldots, M_t)\), where

(i) \( E_0 \sim H_0 \) and \((M_1, M_2, \ldots, M_t)\) are independent,

(ii) \((M_1, \ldots, M_{k-1}) \sim G\) and
\[
M_s = \psi^b_t(M_{s,k}) \varepsilon_s, \quad s = k, k+1, \ldots
\]
where \( M_{s,k} = (M_{s-k}, \ldots, M_{s-1}) \), for an i.i.d. sequence of random variables \( \varepsilon_s \sim K \).

### 3 Stochastic difference equations of tail chains

#### 3.1 Assumptions

Here we explore the form of the results of Theorems 1 and 2 when the Markov chain is stationary with
\[
H_0(x) = (1 - \exp(-x))_+
\]
i.e., the marginal distributions \( F_t \) are identical and the tail belongs to the domain of attraction of the exponential distribution. Furthermore, we restrict the norming functions \( a \) and \( b \) in assumption \( A_2 \) from their general form (6) to the subclass of continuously differentiable functions that are homogeneous of order 1 and \( \beta \in [0, 1) \), respectively, i.e.
\[
(a(tx_1, \ldots, tx_k) = ta(x_1, \ldots, x_k) > 0, \quad \text{and} \quad b(tx_1, \ldots, tx_k) = t^\beta b(x_1, \ldots, x_k),
\]
holds identically in \((x_1, \ldots, x_k)\) and for \( t > 0 \). Note here that we rule out the case \( \beta < 0 \), corresponding to the case where location only normalization gives limits that are degenerate, with all limiting mass at \( \{0\} \). Throughout this section, when we say that a random vector \( M_{k,k-1} \) and a random element \( \varepsilon_k \) follow distributions supported on \( \mathbb{R}^{k-1} \) and \( A \subseteq \mathbb{R} \), respectively, we mean the distributions that are associated to the renormalized initial distribution and the renormalized transition kernel of the Markov process, as specified by assumptions \( A_2 \) and \( B_2 \) in Section 2.

In Section 3.2 we consider the asymptotic dependence case, when \( \chi_D > 0 \), which has been previously studied, before considering in Section 3.3 the case where the process is asymptotically independent at all lages, i.e., \( \chi_A = 0 \) for all \( A \subseteq M \). Intermediate cases are discussed in Example 4 of Section 5.
3.2 Asymptotically dependent Markov chains

**Corollary 1** (Asymptotically dependent Markov chains). Let \( \{X_t : t = 0, 1, \ldots\} \) be a \( k \)-th order stationary Markov chain with \( H_0 \) given by (7) and suppose that as \( u \to \infty \),
\[
\{X_t - X_0 : t = 1, \ldots, k - 1\} \mid X_0 > u \xrightarrow{w} \{M_t : t = 1, \ldots, k - 1\},
\]
where \((M_1, \ldots, M_{k-1})\) is a \((k - 1)\)-dimensional random vector with distribution supported on \( \mathbb{R}^{k-1} \). Suppose further, that as \( u \to \infty \),
\[
X_k + \log g(\exp(-X_0), \ldots, \exp(-X_{k-1})) \mid X_0 > u \xrightarrow{w} \varepsilon_k,
\]
where \( \varepsilon_k \) is a non-degenerate random variable with distribution supported on \( \mathbb{R} \) and \( g \) is a homogeneous function of order 1. Subject to the initial conditions (8) and the transition behaviour determined by (9), then as \( u \to \infty \), the solution of the stochastic difference equation (2) is
\[
X_t - X_0 \mid X_0 > u \xrightarrow{w} M_t, \quad \text{for } t = k, k + 1, \ldots
\]
where
\[
M_t = -\log g(\exp(-M_{t-k}), \ldots, \exp(-M_{t-1})) + \varepsilon_t,
\]
and \( \{\varepsilon_t\}_{t=k}^{\infty} \) is a sequence of i.i.d. random variables supported on \( \mathbb{R} \).

Note that due to the homogeneity of function \( g \), the recursion (10) can also be written as
\[
M_t = M_{t-k} - \log g(1, \exp(M_{t-k} - M_{t-k+1}), \ldots, \exp(M_{t-k} - M_{t-1})) + \varepsilon_t
\]
So when \( k = 1 \) it can be seen to reduce to the random walk results of Smith (1992) and Perfekt (1994), but when \( k > 1 \), the tail chain (not hidden here as \( a_j(x) = x \) and \( b_j(x) = 1 \) for \( j = 1, \ldots, k - 1 \)) behaves like a random walk with an additional factor which depends on the “profile” \( M_{t,k} - M_{t-k} \), of the \( k \) previous values.

3.3 Asymptotically independent Markov chains

**Corollary 2** (Asymptotically independent Markov chains with location and scale norming). Let \( \{X_t : t = 0, 1, \ldots\} \) be a \( k \)-th order stationary Markov chain with \( H_0 \) given by (7) and suppose that there exist functions \( a_t : \mathbb{R} \to \mathbb{R}, t = 1, \ldots, k - 1 \) that are homogeneous of order 1 and functions \( b_t : \mathbb{R} \to \mathbb{R}_+, t = 1, \ldots, k - 1 \), that are homogeneous of order \( \beta \in [0, 1) \), such that as \( u \to \infty \)
\[
\left\{ \frac{X_t - a_t(X_0)}{b_t(X_0)} : t = 1, \ldots, k - 1 \right\} \mid X_0 > u \xrightarrow{w} \{M_t : t = 1, \ldots, k - 1\},
\]
where \((M_1, \ldots, M_{k-1})\) is a \((k - 1)\)-dimensional random vector with distribution supported on \( \mathbb{R}^{k-1} \). Suppose that, as \( u \to \infty \),
\[
\frac{X_k - a(X_0, \ldots, X_{k-1})}{b(X_0, \ldots, X_{k-1})} \mid X_0 > u \xrightarrow{w} \varepsilon_k,
\]
where \( \varepsilon_k \) is a random variable with distribution supported on \( \mathbb{R} \) and \( a \) and \( b \) satisfying the assumptions laid out in Section 3.1. Subject to the initial conditions (11) and the transition behaviour determined by (12), then as \( u \to \infty \), the solution of the stochastic difference equation (2) is
\[
\frac{X_t - a_t(X_0)}{b_t(X_0)} \mid X_0 > u \xrightarrow{w} M_t, \quad \text{for } t = k, k + 1, \ldots
\]
where \( a_t(x) = a_t(1) x \) and \( b_t(x) = x^\beta \), with

\[
a_t(1) = a(a_t(k(1)), \quad (13)
\]

and

\[
M_t = \nabla a(a_t(k(1))) \cdot M_t + b(a_t(k(1))) \varepsilon_t, \quad (14)
\]

for \( \{\varepsilon_t\}_{t=1}^{\infty} \) being a sequence of i.i.d. random variables supported on \( \mathbb{R} \).

Here we find that the norming functions \( a_t \), \( t = k, k+1, \ldots \), have a particularly neat structure, in particular \( a_t(X_0) = a_t(1) X_0 \), where \( a_t(1) \) is determined by the difference equation \((13)\) of the \( k \) previous values \( a_t(k) \) through the homogeneous function \( a(\cdot) \). For a flexible parametric class of the function \( a(\cdot) \), in Section 4 we are able to explicitly solve the difference equations \((13)\).

In general we can view the recurrence relation in expression \((13)\) as the parallel of the Yule–Walker equations and hence, 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. In particular, let \( \phi_1, \ldots, \phi_k \) be 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 \( D = \{ z \in \mathbb{C} : |z| \leq 1 \} \). For a \( k \)th order linear Markov process \( Z_t = \sum_{i=1}^{k} \phi_i Z_{t-k+i-1} + \eta_t \) with \( \{\eta_t\}_{t=0}^{\infty} \) a sequence of zero mean, common finite variance and uncorrelated random variables, the Yule–Walker equations relate the autocorrelation function of the process \( \rho_t = \text{cor}(Z_{s-t}, Z_s) \) with the regression parameters \( \phi_1, \ldots, \phi_k \) and the \( k \) lagged autocorrelations

\[
\rho_t = \sum_{i=1}^{k} \phi_i \rho_{t-i}, \quad t \in \mathbb{Z}.
\]

The sequence \( a_t(1) \) has a similar structure for extremes via recurrence \((13)\).

**Corollary 3** (Asymptotically independent Markov chains with only scale norming). Let \( \{X_t : t = 0, 1, \ldots \} \) be a \( k \)-th order stationary Markov chain with \( H_0 \) given by \((7)\) and suppose there exists functions \( b_t : \mathbb{R} \mapsto \mathbb{R}_+ \) that are homogeneous of order \( \beta \in [0, 1) \), such that as \( u \to \infty \),

\[
\begin{align*}
\left\{ \frac{X_t}{b_t(X_0)} : t = 1, \ldots, k - 1 \right\} & \quad \mid X_0 > u \overset{w}{\to} M_t
\end{align*}
\]

where \( (M_1, \ldots, M_{k-1}) \) is a \( (k-1) \)-dimensional random vector with distribution function supported on \( \mathbb{R}^{k-1} \). Suppose further that there exist \( b : \mathbb{R}^k \mapsto \mathbb{R}_+ \) that is homogeneous of order \( \beta \in (0, 1) \), such that as \( u \to \infty \),

\[
\frac{X_k}{b(X_0, \ldots, X_{k-1})} \quad \mid X_0 > u \overset{w}{\to} \varepsilon_k,
\]

where \( \varepsilon_k \) is random variable with distribution supported on \( [0, \infty) \) and \( b \) satisfying the assumptions laid out in Section 3.1. Subject to the initial conditions \((15)\) and the transition behaviour determined by \((16)\), then as \( u \to \infty \), the solution of the stochastic difference equation \((2)\) is

\[
\frac{X_t}{b_t(X_0)} \quad \mid X_0 > u \overset{w}{\to} M_t, \quad \text{for } t = k+1, k+2, \ldots
\]

where the functions \( b_t : \mathbb{R} \mapsto \mathbb{R}_+, t = k, k+1, \ldots, \) are homogeneous of order \( \beta_t \), where

\[
\log \beta_t = \log \beta + \log \left( \frac{1}{\sqrt{\beta_{t-1}}} \right) = (\left\lfloor (t-1)/k \right\rfloor) \log \beta, \quad (17)
\]
\[ [\cdot] \] denotes the floor function and
\[
M_{nk+j} = b(0, M_{(n-1)k+j}, \ldots, M_{nk}) \varepsilon_{nk+j},
\]
for \( n \in \mathbb{N}, j \in \{1, \ldots, k\} \) and \( \varepsilon_t \) a sequence of i.i.d. random variables supported on \( \mathbb{R}_+ \).

Here we see a similar structure to the special case of Corollary 2, with \( a_t(1) \) decaying in \( t \), but here, \( \beta_t \) does this with the hidden tail chain at time \( t \) depending on only the last \( j \) values with \( j \equiv t \pmod{k} \). So \( M_t \) is independent of the previous values of the hidden tail chain when \( t \pmod{k} = 0 \).

**Remark 2.** With the same assumptions, normalizing functions \( a_t, b_t \), and functionals \( a, b \), as in Corollary 2, an approach is to consider the variance stabilizing transformation (Lamperti, 1962)

\[
Y_t = \int_{X_t}^{\infty} \frac{1}{b(s1)} \, ds.
\] (18)

The map \( \mathbb{R}_+ \ni s \mapsto b(s1) \) is regularly varying at infinity with order \(-\beta \in (-1, 0)\). Therefore, conditionally on \( X_0 > u \) and under the assumptions of Corollary 2 we have \( Y_t \overset{a.a.s.}{\rightarrow} X_t^{1-\beta}/\{b(1)(1-\beta)\} \) for \( t = 0, 1, \ldots, \) as \( u \to \infty \). From continuous mapping theorem (Billingsley, 1999), we obtain

\[
\{Y_t - a_t(Y_0) : t = 1, \ldots, k - 1\} \mid X_0 > u \overset{w}{\rightarrow} \{M_t^Y : t = 1, \ldots, k - 1\}
\]

where \( M_t^Y = (a_t(1)^{-\beta} b_t(1)/b(1)) M_t, \) for \( t = 1, \ldots, k - 1, \) and

\[
Y_k - a_t(Y_0) \mid Y_0 > u \overset{w}{\rightarrow} \frac{1}{b(1)} a(a_t(k))^{-\beta} b(a_t(k)) \varepsilon_t
\]

This results in the solution of the stochastic difference equation given by

\[
Y_t - a_t(Y_0) \mid Y_0 > u \overset{w}{\rightarrow} M_t^Y, \quad t = k, k + 1, \ldots,
\]

where

\[
M_t^Y = \nabla a(a_t(k)^{1/(1-\beta)})^{1-\beta} \cdot M_t^Y + \frac{b(a_t(k))}{b(1)} a(a_t(k))^{-\beta} \varepsilon_t, \quad t = k, k + 1, \ldots \quad (19)
\]

Observe that for the special case \( b(x) = a(x)^{\beta} \), the hidden tail chain has location only dynamics since the scaling function in (19) reduces to \( 1/b(1) \). This case can be seen as a conditional analogue of the stretched exponential distribution (Laherrère and Sornette, 1998).

With the same assumptions, normalizing functions \( b_t \), and functional \( b \), as in Corollary 3, an alternative is to consider the transformation

\[
Y_t = \log X_t.
\] (20)

Conditionally on \( X_0 > u \) and under the assumptions of Corollary 3, we have

\[
\{Y_t - \log b_t(X_0) : t = 1, \ldots, k - 1\} \mid X_0 > u \overset{w}{\rightarrow} M_t^Y
\]
as \( u \to \infty \), where \( M_t^Y = \log M_t, \) for \( t = 1, \ldots, k - 1, \) and

\[
\{Y_k - \log b(\exp(Y_{k,k})) : Y_0 > u \overset{w}{\rightarrow} \log \varepsilon_t
\]

This results in the solution of the stochastic difference equation given by

\[
\{Y_t - \log b_t(X_0) : t = 1, \ldots, k - 1\} \mid X_0 > u \overset{w}{\rightarrow} M_t^Y, \quad t = k, k + 1, \ldots
\]

where

\[
M_{nk+j}^Y = \log b(0, \exp(M_{(n-1)k+j}^Y), \ldots, \exp(M_{nk}^Y)) + \log \varepsilon_{nk+j}.
\]
4 A class of closed form solutions for asymptotically independent chains with homogeneous normalization

Although the results of Section 3 provide much more insight into the form of the norming and updating functions of Theorems 1 and 2, as they hold for arbitrary homogeneous functions \(a(\cdot)\) and \(b(\cdot)\), the precise formulation is opaque. Motivated by examples considered in Section 5, we now restrict the class of homogeneous functions of order 1 to a parsimonious yet flexible parametric class, and see what this means for the features observed in Section 3. Observe that if a function \(f_1(\cdot)\) is homogeneous of order 1 then \(f_\beta(\cdot) = |f_1(\cdot)|^\beta\) is homogeneous of order \(\beta\). So we only need to consider a parametric family of homogeneous functions of order 1.

Consider the function \(f : \mathbb{R}^k \mapsto \mathbb{R}_+\) given by

\[
f(x_1, \ldots, x_k; \gamma, \delta) = c \left\{ \gamma_1 (\gamma_1 x_1)^\delta + \cdots + \gamma_k (\gamma_k x_k)^\delta \right\}^{1/\delta}, \quad \gamma = (\gamma_1, \ldots, \gamma_k) \tag{21}\]

where \(c > 0, \delta \in [-\infty, \infty]\) and \(\gamma \in S_{k-1} = \{ \gamma \in \mathbb{R}^k_+ : \|\gamma\|_1 = 1 \}\). Then \(f\) is homogeneous of order 1, i.e., \(f(tx_1, \ldots, tx_k) = t f(x_1, \ldots, x_k)\) for \((x_1, \ldots, x_k) \in \mathbb{R}_+^k\) and any \(t > 0\). Additionally, \(f\) is continuous in \(\delta \in [-\infty, \infty]\) with

\[
\lim_{\delta \to 0} f(x_1, \ldots, x_k) = c_0 x_1^{\gamma_1} x_2^{\gamma_2} \cdots x_k^{\gamma_k}, \quad c_0 = \prod_{i=1}^n \gamma^\delta_i. \tag{22}\]

and

\[
\lim_{\delta \to \pm \infty} f(x_1, \ldots, x_k) = c_{\pm \infty} f_{\pm \infty} (\gamma_1 x_1, \gamma_2 x_2, \ldots, \gamma_k x_k)\]

where \(f_{\infty}(x) = \sqrt[\delta]{\prod_{i=1}^k x_i}, f_{-\infty}(x) = \wedge_{i=1}^k x_i\), and \(c_{\pm \infty} = f_{\pm \infty}(\gamma)\), with \(x = (x_1, \ldots, x_k)\).

**Proposition 1.** Consider the recurrence relation (13) with \(a(x) = f(x)\), for all \(x \in \mathbb{R}_+^k\), and \(f\) defined by equation (21). Suppose that the \(s \in \mathbb{N}\) distinct roots of the characteristic polynomial

\[
x^k - c^\delta \gamma_1^{1+\delta} x^{k-1} - \cdots - c^\delta \gamma_1^{1+\delta} = 0 \tag{23}\]

are \(r_1, \ldots, r_s\) with multiplicities \(m_1, \ldots, m_s\), \(\sum_i m_i = k\). Then the solution of (13) for \(t = k, k + 1, \ldots\), subject to the initial conditions \((a_1(1), \ldots, a_{k-1}(1)) \in (0, 1)^k\) is

\[
a_t(x) = x \left( \sum_{i=1}^s (C_{i0} + C_{i1} t + \cdots + C_{i,m_i-1} t^{m_i-1}) r_i^t \right)^{1/\delta} \tag{24}\]

where the constants \(C_{i0}, \ldots, C_{i,m_i-1}, i = 1, \ldots, s\), are uniquely determined by the system of equations, for \(t = 0, \ldots, k - 1\)

\[
a_t(1) = \left( \sum_{i=1}^s (C_{i0} + C_{i1} t + \cdots + C_{i,m_i-1} t^{m_i-1}) r_i^t \right)^{1/\delta}. \]

**Remark 3.** Consider the recurrence relation (13) for \(\delta \to 0\). A logarithmic transformation in limit (22) results in the linear nonhomogeneous recurrence

\[
\log a_t(x) - \gamma_1 \log a_{t-1}(x) - \cdots - \gamma_k \log a_{t-k}(x) = \log c - I(\gamma) \tag{25}\]

and \(I(\gamma) = -\sum_{i=1}^k \gamma_i \log \gamma_i\). Suppose that the \(s \in \mathbb{N}\) distinct roots of the characteristic polynomial

\[
x^k - \gamma_k x^{k-1} - \cdots - \gamma_1 = 0, \tag{26}\]
are \( r_1, \ldots, r_s \) with multiplicities \( m_1, \ldots, m_s, \sum m_i = k \). Then the solution of (13), subject to the initial conditions \( a_0(1) = 1 \) and \( a_1(1), \ldots, a_{k-1}(1) \in (0, 1) \), is, for \( t = k, k+1, \ldots \)

\[
a_t(x) = x \exp \left\{ \sum_{i=1}^{s} \left( C_{i0} + C_{i1} t + \cdots + C_{i,m_i-1} t^{m_i-1} \right) r_i^t + \frac{c - I(\gamma)}{\gamma_1 + 2\gamma_2 + \cdots + k\gamma_k} t \right\}
\]

(27)

where the constants \( C_{i0}, \ldots, C_{i,m_i-1}, i = 1, \ldots, s \), are uniquely determined by the system of equations, for \( t = 0, \ldots, k-1 \)

\[
a_t(1) = \exp \left\{ \sum_{i=1}^{s} \left( C_{i0} + C_{i1} t + \cdots + C_{i,m_i-1} t^{m_i-1} \right) r_i^t + \frac{c - I(\gamma)}{\gamma_1 + 2\gamma_2 + \cdots + k\gamma_k} t \right\}.
\]

5 Results for kernels based on important copula classes

5.1 Preliminaries

To illustrate the results in Theorems 1 and 2, we study the extremal behaviour of \( k \)th order stationary Markov chains with unit exponential margins, for two important classes of transition kernels for the copula of \( k+1 \) consecutive values and for the Gaussian copula. These cover both asymptotic dependence and asymptotic independence. For these classes we derive the norming functions and the hidden tail chain. We also consider a case which exhibits a mixture of asymptotic independence and asymptotic dependence over different lags.

Let \( F_F \) denote the joint distribution function of a random vector \( X = (X_0, \ldots, X_k) \), assumed to absolutely continuous w.r.t. Lebesgue measure and standardized to unit Fréchet margins, i.e., \( F_i(x) = \exp(-1/x), x > 0, \) for \( i = 0, \ldots, k \). The construction of all Markov processes studied in this section, is summarised as follows. Write \( C : [0, 1]^{k+1} \rightarrow [0, 1] \) for the copula of \( X \), i.e., \( C(u) = F_F(F_U^{-1}(u_0), \ldots, F_k^{-1}(u_k)), u = (u_0, \ldots, u_k) \in [0, 1]^{k+1} \). Define the Markov kernel, \( \pi_F : \mathcal{B}(R^k) \rightarrow [0, 1] \), by

\[
\pi_F(x_{k,k}, x_k) = \left( \frac{\partial^k}{\partial y_0 \cdots \partial y_{k-1}} C(y_{k,k}, y_k) \right) \bigg/ \left( \frac{\partial^k}{\partial y_0 \cdots \partial y_{k-1}} C(y_{k,k}, \infty) \right),
\]

where \( y_{k+1,k+1} = \exp(-1/x_{k+1,k+1}) \). Then subject to appropriate restrictions on the copula function explained in Section 5.2, the initial distribution \( F_F(x_{k,k}, \infty) \) is the invariant distribution of a Markov process with unit Fréchet margins and kernel \( \pi_F \). To facilitate comparison between the hidden tail chains we standardize the marginal scale to unit rate exponential, i.e., we study stationary Markov processes with the initial distributions and transition kernels given by

\[
F(x_{k,k}) := F_F(T(x_{k,k}), \infty) \quad \text{and} \quad \pi(x_{k,k}, x_k) := \pi_F(T(x_{k,k}), T(x_k))
\]

respectively, where \( T(x) = -1/ \log (1 - \exp(-x)) \).

The first two classes of transition kernels are obtained from the class of multivariate extreme value distributions (Resnick, 1987) and inverted multivariate extreme value distribution (Ledford and Tawn, 1997). The \( k+1 \) dimensional distribution function of the multivariate extreme value distribution with Fréchet margins is given by

\[
F_F(x_{k+1,k+1}) = \exp(-V(x_{k+1,k+1}))
\]

where \( V : \mathbb{R}^{k+1} \rightarrow \mathbb{R}_+ \) is known as the exponent measure which is related with a surjective map to a Radon measure \( H \) on \( S_k = \{ \omega \in \mathbb{R}_+^{k+1} : ||\omega||_1 = 1 \} \) with total mass \( k+1 \), satisfying the moment
constraints \( \int_{S_{k-1}} \omega_i H(d\omega) = 1 \), for all \( i \in \{1, \ldots, k+1\} \), and is given by

\[
V(x_{k+1,k+1}) = \int_{S_k} \sqrt[k]{x_i} H(d\omega).
\]  

(28)

Throughout this section, we assume \( V \) is differentiable (Coles and Tawn, 1991) and denote by \( V_J \) the higher order partial derivative \( \partial^{|J|} V(x) / \prod_{j \in J} \partial x_j \). Let \( \Pi_{k-1} \) denote the set of partitions of \( \{0, \ldots, k-1\} \) and \( \Pi^*_k = \Pi_{k-1} \setminus \{\{0\}, \ldots, \{k-1\}\} \).

The transition kernel induced by the multivariate extreme value copula, in Fréchet margins, is

\[
\pi_F(x_{k,k}, x_k) = \frac{\left[ \sum_{p \in \Pi_{k-1}} (-1)^{|p|} \prod_{J \in p} V_J(x) \right]}{\sum_{p \in \Pi_{k-1}} (-1)^{|p|} \prod_{J \in p} V_J(x, k, \infty)} \exp \{ V(x_{k,k}, \infty) - V(x) \}.
\]

(29)

The second class of kernels is obtained as follows. The \( k+1 \) dimensional survivor function of the multivariate inverted max-stable distribution in exponential margins, is given by

\[
F(x_{k+1,k}, x_k) = \exp\left(-V\left(1/x_{k+1,k+1}\right)\right),
\]

and this induces the transition kernel

\[
\pi^m(x_{k,k}, x_k) = \pi_F\left(1/x_{k,k}, 1/x_k\right),
\]

where \( \pi_F \) is given by equation (29).

### 5.2 Stationary processes and constraints on exponent measure

A \( k \)th order Markov chain with any given margin can be constructed from a \( (k+1) \)-variate copula. However, to ensure stationarity some additional structure needs to be imposed (Joe, 2015). Formally, we achieve this by requiring that the distributions of \( \{X_i : i \in A\} \) and \( \{X_i : i \in B\} \), are identical for any \( A, B \subseteq \mathbb{N} \) such that \( B = A + \tau, \tau \in \mathbb{Z} \). Specifically, let \( [k] = \{0, 1, \ldots, k\} \) and \( \mathcal{P}([k]) = 2^{[k]} \setminus \{\emptyset\} \). Consider the equivalence relation \( \equiv \) on \( \mathcal{P}([k]) \) defined by

\[
A \equiv B : \text{there exists } \tau \in \mathbb{Z} \text{ such that } \{i + \tau : i \in A\} = B.
\]

(30)

The relation \( \equiv \) partitions \( \mathcal{P}([k]) \) into

\[
1 + \sum_{j=1}^{k} \sum_{i=1}^{k-j} \frac{(k-i-1)}{2} \cdot \frac{(k-i-(j-2))}{j-2}
\]

distinct equivalence classes. In what follows, for the \( (k+1) \)-variate exponent measures associated to multivariate extreme value and inverted max-stable copula models, we assume

\[
\lim_{x_{[k] \backslash A} \to \infty} V(x) \bigg|_{x_A=y} = \lim_{x_{[k] \backslash B} \to \infty} V(x) \bigg|_{x_B=y}, \quad y \in \mathbb{R}^{|A|}_+
\]

whenever \( A \equiv B \).
In Fréchet margins, the transition kernel

\[ \pi \]

where

\[ \rho = \alpha(1-k) - |J| \] and \( J \) the set of all partitions of \( k-1 \) excluding \( \{0, \ldots, k-1\} \).

Then assumption \( A \) holds with norming functions \( \alpha(r) = v, b(r) = 1 \), for \( r = 1, \ldots, k-1 \), and

\[ \zeta_k(a_k + 1) = 1 + \sum_{p \in J} (-1)^{|p|} \prod_{j \in p} (1 + |J| / \alpha) \prod_{j \in p} (|J|-|p|) \exp \left( \frac{-v}{\alpha} \right) \]

(1)

where

\[ \tau_k(a_k + 1) = \frac{\zeta_k(a_k + 1)}{\zeta_k(a_k) \tau_k(a_k)} \]

In Fréchet margins, the transition kernel \( \pi \) is given by

\[ \pi_t(x) = \frac{\zeta_k(a_k + t)}{\zeta_k(a_k)} \]

(2)

5.3 Examples

Example 1 (Multivariate extreme value copula with logistic dependence). The multivariate extreme value distribution on Fréchet margins with logistic dependence structure has joint distribution function is

\[ F(x) = \frac{\zeta_k(a_k + 1)}{\zeta_k(a_k) \tau_k(a_k)} \]

(3)

Example 2 (Stationary Gaussian autoregressive process—positive dependence). Let \( \rho = (\rho_0, \rho_1, \ldots, \rho_k) \) with \( |\rho_1| > 0 \), and \( \theta, \alpha > 0 \) such that

\[ \sum_{j=0}^{k-1} \theta_j^2 = 1 \]

\[ \rho_j = \alpha \theta_j \]

and is seen for a more general walk tail chain

\[ \bar{v} = \rho \bar{v} \]

where the limiting distributions \( G \) and \( K \) are respectively

\[ \phi(x) = 1 + \sum_{k=1}^{\infty} \exp(-x/\alpha) (\alpha x/\alpha)^{k-1} \exp(-x/\alpha), \quad b(u) = 1, \quad a = (a_0, a_1, \ldots, a_k) \]

Note that under the initial norms \( \theta, \alpha > 0 \) such that the matrix \( \Sigma \in \mathbb{R}^{k \times k} \) is positive definite, the transition kernel is order \( \mathcal{O}(t) \) as \( t \to \infty \). Also, the spectral measure \( H \) as defined in Section 5.1, places all its mass in the interior of the unit simplex \( S \). Then assumption \( A \) holds with norming functions \( \alpha = v, b(r) = 1 \), for \( r = 1, \ldots, k-1 \), and

\[ \zeta_k(a_k + 1) = 1 + \sum_{p \in J} (-1)^{|p|} \prod_{j \in p} (1 + |J| / \alpha) \prod_{j \in p} (|J|-|p|) \exp \left( \frac{-v}{\alpha} \right) \]

(1)

where

\[ \tau_k(a_k + 1) = \frac{\zeta_k(a_k + 1)}{\zeta_k(a_k) \tau_k(a_k)} \]

(2)

In Fréchet margins, the transition kernel \( \pi \) is given by

\[ \pi_t(x) = \frac{\zeta_k(a_k + t)}{\zeta_k(a_k) \tau_k(a_k)} \]

(3)

The joint distribution function is

\[ F(x) = \frac{\zeta_k(a_k + 1)}{\zeta_k(a_k) \tau_k(a_k)} \]

(4)
\(i, j = 1, \ldots, k\), is positive definite. Consider a stationary \(k\)th order Markov chain with a \(k\)-variate Gaussian distribution

\[
F_G(x_{k-1,k}) = \int_{\prod_{i=0}^{k-1}(-\infty, \infty]} \frac{|\Sigma|^{-1/2}}{(2\pi)^{k/2}} \exp \left( -\frac{1}{2} s \Sigma^{-1} s^T \right) \, ds \quad s = (s_0, \ldots, s_{k-1}).
\]

Using this joint distribution to construct the transition kernel for a \(k\)th order stationary Markov chain with exponential margins, we have the transition kernel

\[
\pi(x_{k,k}, x_k) = \pi_G \left( \Phi^{-1}(1 - \exp(-x_{k+1,k+1})) \right),
\]

where \(\Phi^{-1} : [0,1] \mapsto \mathbb{R}\) denotes the quantile function of the standard normal distribution and

\[
\pi_G(x_{k,k}, x_k) = \Phi \left( \frac{x_k - \sum_{i=0}^{k-1} \phi_{i+1} x_i}{\sigma_\varepsilon} \right).
\]

Here \(\phi = (\phi_1, \ldots, \phi_k)\) is given by \(\rho = \phi \Sigma\) where the elements are real valued scalar parameters with \(\phi_k \neq 0\), and to ensure stationarity, they further satisfy \(1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_k z^k \neq 0\) on \(D = \{z \in \mathbb{C} : |z| \leq 1\}\). Standardization to standard Gaussian margins is achieved by setting

\[
\sigma_\varepsilon^2 = 1 - \sum_{i=1}^k \phi_i^2 - 2 \sum_{j=1}^{k-1} \rho_j \sum_{i=1}^{k-j} \phi_i \phi_{i+j}.
\]

Assumption A1 holds with norming functions \(a_i(v) = \rho_i^2 v, b_i(v) = v^{1/2}\), for \(i = 1, \ldots, k - 1\), and

\[
a(u) = \left( \sum_{i=1}^k \phi_i u_i^{1/2} \right)^2, \quad b(u) = (a(u))^{1/2}, \quad u = (u_1, \ldots, u_k).
\]

where the limiting distributions \(G\) and \(K\) are respectively

\[
G(x) = \Phi\{x; S (\Sigma - \rho_{-0}^T \rho_{-0}) S\} \quad \text{and} \quad K(x) = \Phi\{x/(\sqrt{2} \sigma_\varepsilon)\}.
\]

Here \(S\) is a diagonal matrix with diagonal \(\sqrt{2} \rho_{-0}, \rho_{-0}\) is the first row of \(\Sigma\) with the first element omitted and \(\Sigma_{-0}\) is \(\Sigma\) with first row and first column omitted (Heffernan and Tawn, 2004). Using the same strategy as in Example 1, the functional \(a\) and \(b\) are identified by balancing \(y_t - \sum_{i=0}^{k-1} \phi_{i+1} y_{t-i}\), where \(y_t = (2x_t)^{1/2} - \{\log x_t + \log(4\pi)\}/\{2(2x_t)^{1/2}\}\), for \(t = 0, 1, \ldots, k\), see also Heffernan and Tawn (2004).

A suitable normalization after \(t\) steps is \(a_t(v) = \rho_t^2 v, b_t(v) = v^{1/2}\), where \(\rho_t\) is the correlation function of the stationary autoregressive Gaussian process at lag \(t\), i.e., \(\rho_t = \sum_{i=1}^{k} \phi_i \rho_{t-i}\), which leads to the scaled autoregressive tail chain

\[
M_t = \rho_t \sum_{i=1}^{k} \frac{\phi_i}{\rho_{t-i}} M_{t-i} + \sqrt{2} \rho_t \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2).
\]

Remark 4. For \(c = \left( \sum_{i=1}^{k} \phi_i^{2/3} \right)^3\) and \(\gamma_i = \phi_i^{2/3} / \sum_{i=1}^{k} \phi_i^{2/3}\), the location functional \(a\) can be written in form (21).
Example 3 (Inverted multivariate extreme value copula with logistic dependence). Consider a stationary $k$th order Markov chain with a $k$-variate distribution function

$$F(x_{k,k}) = \exp \left\{ - \frac{1}{\alpha} \right\} \quad \alpha \in (0, 1),$$

Using this joint distribution to construct the transition kernel for a $k$th order stationary Markov chain with exponential margins, we have the transition kernel

$$\pi^{\text{inv}}(x_{k,k}, x_k) = \mathcal{L}^{\text{inv}}(x_{k+1,k+1}) \left( 1 + \frac{x_{k,k}^{1/\alpha}}{\|x_{k,k}^{1/\alpha}\|_1} \right)^{a-k} \exp \left\{ \|x_{k,k}^{1/\alpha}\|_1 - \|x_{k+1,k+1}^{1/\alpha}\|_1 \right\}$$

where $\mathcal{L}^{\text{inv}}(x_{k+1,k+1}) = \mathcal{L}(1/x_{k+1,k+1})$, with $\mathcal{L}$ as defined by equation (31). Then assumption $B_1$ holds with norming functions $b_i(v) = v^{1-\alpha}$, for $i = 1, \ldots, k - 1$,

$$a(u) = 0, \quad b(u) = \left( \sum_{i=1}^{k} u_i^{1/\alpha} \right)^{(1-\alpha)} \quad u = (u_1, \ldots, u_k),$$

where the limiting distributions are respectively

$$G(x_1, \ldots, x_{k-1}) = \prod_{j=1}^{k-1} \exp (-\alpha x_j^{1/\alpha}), \quad K(x) = \exp (-\alpha x_1^{1/\alpha}).$$

where, by factorization of $G$, we have, in the limit, independence between elements of the initial random vector $X_1, \ldots, X_{k-1}$. Using the same strategy as in Example 1, the functional $b$ is identified by balancing

$$\left( \sum_{j=0}^{k-1} x_j^{1/\alpha} \right)^{\alpha} = \left( \sum_{j=0}^{k} x_j^{1/\alpha} \right)^{\alpha}$$

A suitable normalization after $t$ steps is $a_t(v) = \log b_t(v) = ((1 - \alpha)^{1+[\ell-1]/k]} \log v$, which leads to the scaled random walk tail chain

$$M_{nk+j} = b(0, M_{(n-1)k+j}, \ldots, M_{nk}) \xi_{nk+j}, \quad 0 = (0, \ldots, 0) \in \mathbb{R}^j.$$ 

for $n \in \mathbb{N}, j \in \{1, \ldots, k\}$ and $\{\xi_t\}_{t=k}$ a sequence of i.i.d. random variables.

The next example shows an asymptotically dependent 2nd order Markov process for which assumptions $A_1$ and $A_2$ fail to hold and has apparently more complicated structure than what we have considered so far.

Example 4 (Max-stable dependence with asymmetric logistic structure (Tawn, 1990)). Consider the exponent measure

$$V_{012}(x_0, x_1, x_2) = \theta_0 x_0^{-1} + \theta_1 x_1^{-1} + \theta_2 x_2^{-1} +$$

$$+ \theta_01 \left\{ \left( x_0^{-1/\nu_01} + x_1^{-1/\nu_01} \right)^{1/\nu_01} + \left( x_1^{-1/\nu_01} + x_2^{-1/\nu_01} \right)^{1/\nu_01} \right\} +$$

$$+ \theta_02 \left( x_0^{-1/\nu_02} + x_2^{-1/\nu_02} \right)^{\nu_02} +$$

$$+ \theta_{012} \left( x_0^{-1/\nu_{012}} + x_1^{-1/\nu_{012}} + x_2^{-1/\nu_{012}} \right)^{1/\nu_{012}},$$
where the parameters $\nu_A \in (0, 1)$ for any $A \in \mathcal{P}([2])$, and

\[
\begin{align*}
\theta_0 + \theta_{01} + \theta_{02} + \theta_{012} &= 1 \\
\theta_1 + 2\theta_{01} + \theta_{012} &= 1 \\
\theta_2 + \theta_{01} + \theta_{02} + \theta_{012} &= 1 \\
\theta_0, \theta_1, \theta_2, \theta_{01}, \theta_{02}, \theta_{012} &> 0.
\end{align*}
\]

Here, the spectral measure $H$ of the multivariate extreme value distribution places mass of size $|J|\theta_J$ on subface $J \in \mathcal{P}([2])$ of $S_2$ (Coles and Tawn, 1991). Furthermore the joint initial distribution of $(X_0, X_1)$ is $F_{01}(x_0, x_1) = F_{012}(x_0, x_1, \infty) = \exp(-V_{01}(x_0, x_1))$ where $V_{01}(x_0, x_1) = \lim_{x_2 \to \infty} V(x_0, x_1, x_2)$. It can be seen that the kernel obtained from the conditional distribution of $X_1 \mid X_0$

\[
\pi(x_0, x_1) = -x_0^2 \frac{\partial}{\partial x_0}V(x_0, x_1) \exp \left( \frac{1}{x_0} - V(x_0, x_1) \right)
\]

covers with two distinct normalizations

\[
\pi(v, dx) \xrightarrow{w} K_0(dx) \quad \text{and} \quad \pi(v, v + dx) \xrightarrow{w} K_1(dx)
\]
to the distributions

\[
K_0 = (\theta_0 + \theta_{02}) F_E + (1 - \theta_0 - \theta_{012}) \delta_{+\infty}, \quad F_E(x) = (1 - \exp(-x))_+, \\
K_1 = (\theta_0 + \theta_{02}) \delta_{-\infty} + \theta_0 G_{01} + \theta_{012} G_{012}, \quad G_A(x) = (1 + \exp(-x/\nu_A))^{\nu_A - 1}.
\]

These distributions have entire mass on $[-\infty, \infty)$ and $(0, \infty)$, respectively (cf. Papastathopoulos et al., 2017), where $\delta_x$ is a point mass at $x \in [-\infty, \infty]$. Similarly, the transition kernel obtained from the conditional distribution of $X_2 \mid X_0, X_1$ is

\[
\pi(x_{2,2}, x_2) = \frac{(V_0 V_1 - V_{01})(T(x_{3,3}))}{(V_0 V_1 - V_{01})(T(x_{2,2}))} \exp(V(T(x_{2,2})) - V(T(x_{3,3}))),
\]

where we have used the notation $g(f_1(x), f_2(x), f_3(x)) = g(f_1, f_2, f_3)(x)$ for maps $g$ and $f_i$, $i = 1, 2, 3$. This transition kernel can be shown to converge with $2 (2^k - 1)$ to $6$ distinct normalizations

\[
\begin{align*}
\pi((v, v + x), v + dy) \xrightarrow{w} K_{111}(dy), & \quad \pi((v, v + x), dy) \xrightarrow{w} K_{110}(dy), \\
\pi((v, x), v + dy) \xrightarrow{w} K_{101}(dy), & \quad \pi((v, x), dy) \xrightarrow{w} K_{100}(dy), \\
\pi((x, v), v + dy) \xrightarrow{w} K_{011}(dy), \quad \text{and} \quad \pi((x, v), dy) \xrightarrow{w} K_{010}(dy),
\end{align*}
\]
to the distributions

\[
\begin{align*}
K_{111} &= m_{111} \delta_{-\infty} + (1 - m_{111}) G_{111} \\
K_{110} &= m_{110} G_{110} + (1 - m_{110}) \delta_{+\infty} \\
K_{101} &= \frac{\theta_0}{\theta_0 + \theta_{02}} \delta_{-\infty} + \frac{\theta_{02}}{\theta_0 + \theta_{02}} G_{101} \\
K_{100} &= \frac{\theta_0(1 - \theta_{012})}{\theta_0 + \theta_{02}} G_{100} + \left(1 - \frac{\theta_0(1 - \theta_{012})}{\theta_0 + \theta_{02}}\right) \delta_{+\infty} \\
K_{011} &= \frac{\theta_1}{\theta_1 + \theta_{01}} \delta_{-\infty} + \frac{\theta_{01}}{\theta_1 + \theta_{01}} G_{011} \\
K_{010} &= \frac{\theta_1(1 - \theta_{012})}{\theta_1 + \theta_{01}} G_{010} + \left(1 - \frac{\theta_1(1 - \theta_{012})}{\theta_1 + \theta_{01}}\right) \delta_{+\infty}
\end{align*}
\]
where

\[ G_{111}(y) = \left( 1 + \frac{\exp(-y/\nu_{012})}{\exp(-x/\nu_{012}) + \exp(-y/\nu_{012})} \right)^{\nu_{012}-2} \]
\[ G_{110}(y) = F_E(y) = (1 - \exp(-y))_+ \]
\[ G_{101}(y) = (1 + \exp(-y/\nu_{02}))^{\nu_{02}-1} \]
\[ G_{100}(y) = \frac{1}{1 - \theta_{012}} \left( \theta_2 + \theta_{01} + \theta_{02} + \theta_{12} \left\{ 1 + \exp(-(y-x)/\nu_{012}) \right\}^{\nu_{012}-1} \right) \]
\[ G_{011}(y) = (1 + \exp(-y/\nu_{01}))^{\nu_{01}-1} \]
\[ G_{010}(y) = \frac{1}{1 - \theta_{012}} \left( \theta_0 + \theta_{01} + \theta_{02} + \theta_{12} \left\{ 1 + \exp(-(y-x)/\nu_{012}) \right\}^{\nu_{012}-1} \right) \]

with entire mass on \([-\infty, \infty), (0, \infty), [-\infty, \infty), (-\infty, \infty), [-\infty, \infty), (-\infty, \infty]\) respectively, where

\[ m_{110} = \frac{\theta_{012}^{\nu_{012}-1} x^{-1-1/\nu_{012}} y^{-2/\nu_{012}} + \theta_{01}^{\nu_{01}-1} x^{-1-1/\nu_{01}} (1 + x^{-1/\nu_{01}})^{\nu_{01}-2}}{\theta_{012}^{\nu_{012}-1} x^{-1/\nu_{012}} (1 + y^{-1/\nu_{012}})^{\nu_{012}-2} + \theta_{01}^{\nu_{01}-1} x^{-1-1/\nu_{01}} (1 + x^{-1/\nu_{01}})^{\nu_{01}-2}} \]
\[ m_{111} = \frac{\theta_{012}^{\nu_{012}-1} (xy)^{-1-1/\nu_{012}} (1 + x^{-1/\nu_{012}})^{\nu_{012}-2} + \theta_{01}^{\nu_{01}-1} (xy)^{-1-1/\nu_{01}} (1 + x^{-1/\nu_{01}})^{\nu_{01}-2}}{\theta_{012}^{\nu_{012}-1} (xy)^{-1/\nu_{012}} (1 + y^{-1/\nu_{012}})^{\nu_{012}-2} + \theta_{01}^{\nu_{01}-1} (xy)^{-1-1/\nu_{01}} (1 + x^{-1/\nu_{01}})^{\nu_{01}-2}}. \]

Here we see that for every out of the \(2^k - 1\) possible initial normings, there are two distinct modes of normalization for the next state in the process. This shows a type of transition mechanism that allows jumps between the body of the process and an extreme state, for as long as \(k = 2\) consecutive non-extreme states are observed. The propensity for jumping between extreme and non-extreme states is dictated by the parameters \(\theta_f\).

6 Proofs

6.1 Preparatory results

The proofs of Theorem 1 and Theorem 2 are based on Lemma 1 and Lemma 2 whose proofs are verbatim to Lemma 4 and Lemma 5 in Papastathopoulos et al. (2017) and are omitted for brevity.

**Lemma 1.** Let \(\{X_t : t = 0, 1, \ldots\}\) be a homogeneous \(k\)-th order Markov chain satisfying assumptions \(A_1, A_2\). Let \(h \in C_b(\mathbb{R})\). Then, for \(t = k, k + 1, \ldots\), as \(v \to \infty\)

\[ \int_{\mathbb{R}} h(y) \pi(a_{t,k}(v) + b_{t,k}(v)x_{t,k}, t, v) + b_t(v) dy \to \int_{\mathbb{R}} h(\psi_t^a(x_{t,k}) + \psi_t^b(x_{t,k})) K(dy). \]

and the convergence holds uniformly on compact sets in the variable \(x_{t,k} \in \mathbb{R}^k\)

**Lemma 2.** Let \(\{X_t : t = 0, 1, \ldots\}\) be a homogeneous \(k\)-th order Markov chain satisfying assumptions \(B_1, B_2\). Let \(h \in C_b(\mathbb{R})\). Then, for \(t = k, k + 1, \ldots\), as \(v \to \infty\)

\[ \int_{\mathbb{R}} h(y) \pi(b_{t,k}(v)x_{t,k}, t, v) dy \to \int_{\mathbb{R}} h(\psi_t^b(x_{t,k})) K(dy). \]

and the convergence holds uniformly on compact sets in the variable \(x_{t,k} \in [\delta_1, \infty) \times \cdots \times [\delta_k, \infty)\) for any \((\delta_1, \ldots, \delta_k) \in (0, \infty)^k\).
Lemma 3. Let \((E, d)\) be a complete locally compact separable metric space and \(\mu_n\) be a sequence of probability measures which converges weakly to a probability measure \(\mu\) on \(E\).

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

(ii) Let \(F\) be a topological space. If \(\varphi \in C_0(F \times E)\), 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 \(F \ni x \mapsto \int_E \varphi(x, y) \mu(dy) \in \mathbb{R}\).

6.2 Proof of Theorem 1

Let \(a_0(v) \equiv v\) and \(b_0(v) \equiv 1\) and define
\[
v_u(y_0) = u + \sigma(u)y_0 \\
A_t(v, x) = a_t(v) + b_t(v)x \\
A_{t,k}(v, x_{t,k}) = (A_{t-k}(v, x_{t-k}), \ldots, A_{t-1}(v, x_{t-1}))
\]

**Proof of Theorem 1.** Considering the measures
\[
\mu_{t}^{(u)}(dy_0, \ldots, dy_t) = \pi(A_{t,k}(v_u(y_0), y_{t,k}), A_t(v_u(y_0), dy_t))
\]
\[
\pi(A_{k+1,k}(v_u(y_0), y_{k+1,k}), A_t(v_u(y_0), dy_{k+1}))
\]
\[
\pi((v_u(y_0), A_{k,k}(v_u(y_0), y_{k,k})), A_k(v_u(y_0), dy_k))
\]
\[
\pi(v_u(y_0), A_{k,k-1}(v_u(y_0), dy_{k-1})) \frac{F_0(v_u(dy_0))}{F_0(u)}
\]

and
\[
\mu_t(dy_0, \ldots, dy_t) = K \left( \frac{dy_t - \psi^a(y_{t,k})}{\psi^b(y_{t,k})} \right) \ldots K \left( \frac{dy_{k+1} - \psi^a_k(y_{k+1,k})}{\psi^b_k(y_{k+1,k})} \right) K(dy_k)
\]
\[
G(dy_1 \times \cdots \times dy_{k-1}) H_0(dy_0)
\]
on \([0, \infty) \times \mathbb{R}^t\), we may rewrite
\[
E \left[ f \left( \frac{X_0 - u}{\sigma(u)}, \frac{X_1 - a_1(X_0)}{b_1(X_0)}, \ldots, \frac{X_t - a_t(X_0)}{b_t(X_0)} \right) \left| X_0 > u \right. \right]
\]
\[
= \int_{(0, \infty) \times \mathbb{R}^t} f(y_0, \ldots, y_t) \mu_{t}^{(u)}(dy_0, \ldots, dy_t)
\]
and
\[
E \left[ f(E_0, M_1, \ldots, M_t) \mid X_0 > u \right] = \int_{(0, \infty) \times \mathbb{R}^t} f(y_0, \ldots, y_t) \mu_t(dy_0, \ldots, dy_t)
\]
for $f \in C_b([0, \infty) \times \mathbb{R}^k)$. We need to show that $\mu_t^{(w)}$ converges weakly to $\mu_t$. Let $f_0 \in C_b([0, \infty))$ and $g \in C_b(\mathbb{R}^k)$. The proof is by induction on $t$. For $t = k$ it suffices to show that

$$
\int_{[0,\infty) \times \mathbb{R}^k} f_0(y_0) g(y_1, \ldots, y_k) \mu_k^{(w)}(dy_0, dy_1, \ldots, dy_k)
= \int_{[0,\infty)} f_0(y_0) \left[ \int_{\mathbb{R}^k} g(y_1, \ldots, y_k) \pi((v_u(y_0), A_{k,k-1}(v_u(y_0), y_{k-1})), A_k(v_u(y_0), dy_k)) \right. \\
\left. \pi(v_u(y_0), A_{k,k-1}(v_u(y_0), d\gamma_{k-1})) \right] \frac{F_0(v_u(dy_0))}{F_0(u)}, \tag{33}
$$

converges to $E(f(E_0)) E(g(M_1, \ldots, M_k))$.

The term in square brackets of (33) is bounded and by assumptions $A_1$ and $A_2$, it converges to $E[g(M_1, \ldots, M_k)]$ for $u \to \infty$ since $v_u(y_0) \to \infty$ as $u \to \infty$. The convergence holds uniformly in the variable $y_0 \in [0, \infty)$ since $\sigma(u) > 0$. Therefore Lemma 3 applies, which guarantees convergence of the entire term (33) to $E(E_0) E(g(M_1, \ldots, M_k))$ with regards to assumption $A_0$.

Next, assume that the statement is true for some $t > k$. It suffices to show that for $f_0 \in C_b([0, \infty) \times \mathbb{R}^t, g \in C_b(\mathbb{R})$, 

$$
\int_{[0,\infty) \times \mathbb{R}^{t+1}} f_0(y_0, y_1, \ldots, y_t) g(y_{t+1}) \mu_{t+1}^{(w)}(dy_0, dy_1, \ldots, dy_{t+1})
= \int_{[0,\infty) \times \mathbb{R}^t} f_0(y_0, y_1, \ldots, y_t) \left[ \int_{\mathbb{R}} g(y_{t+1}) \pi(A_{t+1,k}(v_u(y_0), y_{t+1,k}), A_t(v_u(y_0), dy_{t+1})) \right. \\
\left. \mu_t(dy_0, dy_1, \ldots, dy_t) \right]. \tag{34}
$$

converges to

$$
\int_{[0,\infty) \times \mathbb{R}^{t+1}} f_0(y_0, y_1, \ldots, y_t) g(y_{t+1}) \mu_{t+1}(dy_0, dy_1, \ldots, dy_{t+1})
= \int_{[0,\infty) \times \mathbb{R}^t} f_0(y_0, y_1, \ldots, y_t) \left[ \int_{\mathbb{R}} g(y_{t+1}) K \left( \frac{dy_{t+1}}{\psi_{t}^{(u)}(y_{t+1})} \right) \psi_{t}^{(u)}(y_{t+1,k}) \right] \mu_t(dy_0, dy_1, \ldots, dy_t). \tag{35}
$$

The term in square brackets of (34) is bounded, and by Lemma 1 and assumptions $A_1$ and $A_2$, it converges uniformly on compact sets in both variables $(y_0, y_{t+1,k}) \in [0, \infty) \times \mathbb{R}^k$ jointly, since $\sigma(u) > 0$. Hence the induction hypothesis and Lemma 3 imply the desired result. □

6.3 Proof of Theorem 2

Proof of Theorem 2. To simplify the notation, we abbreviate the affine transformation $v_u(y_0) = u + \sigma(u)y_0$ henceforth. Define

$$
v_u(y_0) = u + \sigma(u)y_0 \\
b_{t,k}(v, x_{t,k}) = (b_{t-k}(v) x_{t-k}, \ldots, b_{t-1}(v) x_{t-1})
$$
Considering the measures
\[
\mu_t^{(w)}(dy_0, \ldots, dy_t) = \pi(b_{t,k}(v_u(y_0), y_{t,k}), b_t(v_u(y_0)), dy_t) \ldots
\]

\[
\pi(b_{k+1,t}(v_u(y_0), y_{k+1,t}), b_t(v_u(y_0))dy_{k+1})
\]

\[
\pi((v_u(y_0), b_{k,k}(v_u(y_0), y_{k,k})), b_k(v_u(y_0))dy_k)
\]

\[
\pi(v_u(y_0), b_{k,k-1}(v_u(y_0), dy_{k,k-1})) \frac{F_0(v_u(dy_0))}{F_0(u)}
\]  

(36)

and

\[
\mu_t(dy_0, \ldots, dy_t) = K \left( \frac{dy_t}{\psi_t^k(y_{t,k})} \right) \ldots K \left( \frac{dy_{k+1}}{\psi_{k+1}^k(y_{k+1,k})} \right)
\]

\[
K \left( \frac{dy_k}{\psi_k^k(y_{k,k})} \right) G(dy_1, \ldots, dy_{k-1})H_0(dy_0),
\]  

(37)

on \([0, \infty) \times [0, \infty)^t\), we may write

\[
E \left[ f \left( \frac{X_0 - u}{\sigma(u)}, \frac{X_1}{b_1(X_0)}, \ldots, \frac{X_t}{b_t(X_0)} \right) \bigg| X_0 > u \right] = \int_{[0,\infty)\times[0,\infty)^t} f(y_0, y_1, \ldots, y_t) \mu_t^{(u)}(dy_0, \ldots, dy_t)
\]

and

\[
E \left[ f(E_t, M_1, \ldots, M_t) \right] = \int_{[0,\infty)\times[0,\infty)^t} f(y_0, y_1, \ldots, y_t) \mu_t(dy_0, \ldots, dy_t)
\]

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 (36), since \(v_u(y_0) \geq u > 0\) for \(y_0 \geq 0\) and sufficiently large \(u\), whereas (37) is well-defined, since the measures \(G\) and \(K\) put no mass at any half-plane \(C_j = \{y_j, y_{j-1} \in [0, \infty)^{k-1} : y_j = 0\} \subseteq [0, \infty)^{k-1}\) and at \(0 \in [0, \infty)\) respectively. Formally, we may set \(\psi_j^0(0) = 1, j = 1, \ldots, t\) 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 \(\mu_t^{(w)}(dy_0, \ldots, dy_t)\) converges weakly to \(\mu_t(dy_0, \ldots, dy_t)\). The proof is by induction on \(t\). We show two statements by induction on \(t\):

**I**  \(\mu_t^{(w)}(dy_0, \ldots, dy_t)\) converges weakly to \(\mu_t(dy_0, \ldots, dy_t)\) 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 proving the case \(t = k\).

**I** for \(t = k\): It suffices to show that for \(f_0 \in C_b([0, \infty))\), \(g_1 \in C_b([0, \infty)^{k-1})\) and \(g_2 \in C_b([0, \infty))\)

\[
\int_{[0,\infty)\times[0,\infty)^{k-1} \times [0, \infty)} f_0(y_0)g(y_1, \ldots, y_k)\mu_1^{(w)}(dy_0, \ldots, dy_k)
\]

\[
= \int_{[0,\infty)} f_0(y_0) \int_{[0,\infty)^{k-1}} \int_{[0,\infty)} g(y_1, \ldots, y_k)\pi((v_u(y_0), b_{k,k}(v_u(y_0), y_{k,k})), b_k(v_u(y_0))dy_k)
\]

\[
\pi(v_u(y_0), b_{k,k-1}(v_u(y_0), dy_{k,k-1})) \frac{F_0(v_u(dy_0))}{F_0(u)}
\]  

(38)
converges to
\[ \int_{(0,\infty) \times (0,\infty)} f_0(y_0) g(y_1, \ldots, y_k) \mu_1(dy_0, dy_1, \ldots, dy_k) = E(f_0(E_0)) E(g(M_1, \ldots, M_k)). \]

The term in the curly brackets \([\ldots]\) is bounded and, by assumption \(B_2\), it converges to \(E(g(M_1, \ldots, M_k))\) for \(u \uparrow \infty\), since \(v_u(y_0) \to \infty\) for \(u \uparrow \infty\). The convergence is uniform in the variable \(y_0\), since \(\sigma(u) > 0\). Therefore, Lemma 3 (i) applies, which guarantees convergence of the entire term (38) to \(E(f_0(E_0)) E[g(M_1, \ldots, M_k)]\) with regard to assumption \(A_0\).

(II) for \(t = k\): Since \(K(\{0\}) = 0\), there exists \(\delta > 0\) such that \(K([0, \delta]) < \epsilon\), which immediately entails \(\mu_k([0, \infty)^k \times [0, \delta]) = H_0([0, \infty)) G([0, \infty)^k \times \delta) < \epsilon\).

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

(I) for \(t + 1\): It suffices to show that for \(f_0 \in C_b([0, \infty) \times [0, \infty)^t), g \in C_b([0, \infty))\)
\[ \int_{(0,\infty) \times (0,\infty)^{t+1}} f_0(y_0, y_1, \ldots, y_t) g(y_{t+1}) \mu_{t+1}^{(u)}(dy_0, dy_1, \ldots, dy_t, dy_{t+1}) \]
\[ = \int_{(0,\infty) \times (0,\infty)^{t}} f_0(y_0, y_1, \ldots, y_t) \left[ \int_{(0,\infty)} g(y_{t+1}) \pi(b_{t+1,k}(\psi_u(y_0), y_{t+1,k}(v_u(y_0))), b_{t+1}(v_u(y_0)) dy_{t+1}) \right] \mu_t^{(u)}(dy_0, dy_1, \ldots, dy_t) \]
converges to
\[ \int_{(0,\infty) \times (0,\infty)^{t+1}} f_0(y_0, y_1, \ldots, y_t) g(y_{t+1}) \mu_{t+1}(dy_0, dy_1, \ldots, dy_t, dy_{t+1}) \]
\[ = \int_{(0,\infty) \times (0,\infty)^{t}} f_0(y_0, y_1, \ldots, y_t) \left[ \int_{(0,\infty)} g(y_{t+1}) K(dy_{t+1}/\psi_t^{(b)}(y_{t+1,k}()) \right] \mu_t(dy_0, dy_1, \ldots, dy_t). \]

From Lemma 2 and assumptions \(B_1\) and \(B_2\) (a) and (b) we know that, for any \(\delta > 0\), the (bounded) term in the brackets \([\ldots]\) of (39) converges uniformly on compact sets in the variable \(y_{t+1,k} \in \prod_{i=1}^{k}[\delta_i, \infty)\) to the continuous function
\[ \int_{(0,\infty)} g(\psi_t^{(b)}(y_{t+1,k}) y_{t+1}) K(dy_{t+1}) \]
(the term in the brackets \([\ldots]\) of (40)). This convergence holds even uniformly on compact sets in both variables \((y_0, y_{t+1,k}) \in [0, \infty) \times \prod_{i=1}^{t}[\delta_i, \infty)\) jointly, since \(\sigma(u) > 0\). Hence, the induction hypothesis (I) and Lemma 3 (i) imply that for any \(\delta > 0\) the integral in (39) converges to the integral in (40) if the integrals with respect to \(\mu_t\) and \(\mu_t^{(u)}\) were restricted to \(A_\delta := [0, \infty) \times [0, \infty)^{t-1} \times [\delta, \infty)\) (instead of integration over \([0, \infty) \times [0, \infty)^{t-1} \times [0, \infty)\)).

Therefore (and since \(f_0\) and \(g\) are bounded) it suffices to control the mass of \(\mu_t\) 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 \(u\), such that \(\mu_t(A_\delta^c) < \varepsilon\) and \(\mu_t^{(u)}(A_\delta^c) < 2\varepsilon\). Because of the induction hypothesis (II), we have indeed \(\mu_t(A_\delta^c) < \varepsilon\) for some \(\delta \geq \delta/2\) and note that the sets of the form \(A_\delta\) are nested. Let \(C_\delta\) be a continuity set of \(\mu_t\) with \(A_\delta^c \subset C_\delta \subset A_{\delta/2}^c\). Then the value of \(\mu_t\) on all three sets \(A_\delta^c, C_\delta, A_{\delta/2}^c\) is smaller than \(\varepsilon\) and because of 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
\( \mu_t^{(w)}(A_\delta) < \mu_t^{(w)}(C_\delta) < \mu_t(C_\delta) + \varepsilon < 2\varepsilon \), as desired.

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

\[
\mu_{t+1}([0, \infty) \times [0, \infty)^t \times [0, \delta]) = \int_{[0, \infty)^t \times [0, \delta]} K([0, \delta/\psi_{t}(y_{t+1,k})]) \mu_t(dy_0, \ldots, dy_t).
\]

Splitting the integral according to \{\psi_{t}(y_{t+1,k}) > c\} or \{\psi_{t}(y_{t+1,k}) \leq c\} yields

\[
\mu_{t+1}([0, \infty) \times [0, \infty)^t \times [0, \delta]) \leq K([0, \delta/c]) + \mu_t([0, \infty) \times [0, \infty)^{t-1} \times (\psi_{t})^{-1}([0, c])).
\]

By assumption \( B_3(c) \) and the induction hypothesis (II) we may choose \( c > 0 \) sufficiently small, such that the second summand \( \mu_t([0, \infty) \times [0, \infty)^{t-1} \times (\psi_{t})^{-1}([0, c])) \) is smaller than \( \varepsilon/2 \). Secondly, since \( K([0]) = 0 \), it is possible to choose \( \delta_{t+1} = \delta > 0 \) accordingly small, such that the first summand \( K([0, \delta]) \) is smaller than \( \varepsilon/2 \), which shows (II) for \( t + 1 \).

6.4 Proof of Proposition 1

Proof of Proposition 1. The proof of Proposition 1 follows at once after noticing that the recurrence relation

\[
a_t(x) = c \left[ \sum_{i=1}^{k} \gamma_i (\gamma_i)^{a_{t-k+i-1}(x)} \right]^{1/\delta},
\]

can be converted to the order-\( k \) homogeneous linear recurrence relation, given by \( y_t = \sum_{i=1}^{k} c_i y_{t-k+i-1} \), for \( \{y_t\} = \{a_t(1)^{\delta}\} \) and \( c_i = c^{\delta} \gamma_i^{1+\delta} \). Solving this latter recurrence relation and transforming the solution back to the original sequence \( \{a_t(x) = x a_t(1)\} \) leads to the claim.

References

Billingsley, P. (1999), *Convergence of Probability Measures*, Wiley Series in Probability and Statistics: Probability and Statistics, second edn, John Wiley & Sons, Inc., New York.

Coles, S. G. and Tawn, J. A. (1991), ‘Modelling extreme multivariate events’, *J. Roy. Statist. Soc., B* 53, 377–392.

Heffernan, J. E. and Resnick, S. I. (2007), ‘Limit laws for random vectors with an extreme component’, *Ann. Appl. Prob.* 17, 537–571.

Heffernan, J. E. and Tawn, J. A. (2004), ‘A conditional approach for multivariate extreme values (with discussions and reply by the authors)’, *J. Roy. Statist. Soc., B* 66(3), 1–34.

Janßen, A. and Segers, J. (2014), ‘Markov tail chains’, *J. Appl. Probab.* 51(4), 1133–1153.

Joe, H. (1997), *Multivariate Models and Dependence Concepts*, Chapman & Hall, London.

Joe, H. (2015), *Dependence Modeling with Copulas*, Vol. 134 of *Monographs on Statistics and Applied Probability*, CRC Press, Boca Raton, FL.

Kifer, Y. (2012), *Ergodic theory of random transformations*, Vol. 10, Springer Science & Business Media.
Kulik, R. and Soulier, P. (2015), ‘Heavy tailed time series with extremal independence.’, Extremes pp. 1–27.

Laherrère, J. and Sornette, D. (1998), ‘Stretched exponential distributions in nature and economy: “fat tails” with characteristic scales’, The European Physical Journal B - Condensed Matter and Complex Systems 2(4), 525–539.

Lamperti, J. (1962), ‘Semi-stable stochastic processes’, Trans. Amer. Math. Soc. 104, 62–78.

Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983), Extremes and Related Properties of Random Sequences and Series, Springer–Verlag, New York.

Ledford, A. W. and Tawn, J. A. (1997), ‘Modelling dependence within joint tail regions’, J. Roy. Statist. Soc., B 59, 475–499.

Ledford, A. W. and Tawn, J. A. (2003), ‘Diagnostics for dependence within time series extremes’, J. R. Statist. Soc., B 65, 521–543.

Papastathopoulos, I., Strokorb, K., Tawn, J. A. and Butler, A. (2017), ‘Extreme events of Markov chains’, Advances in Applied Probability 49, 134–161.

Perfekt, R. (1994), ‘Extremal behaviour of stationary Markov chains with applications’, Ann. Appl. Probab. 4(2), 529–548.

Perfekt, R. (1997), ‘Extreme value theory for a class of Markov chains with values in R^d’, Adv. Appl. Probab. 29(1), 138–164.

Resnick, S. I. (1987), Extreme Values, Regular Variation and Point Processes, Springer–Verlag, New York.

Resnick, S. I. and Zeber, D. (2013), ‘Asymptotics of Markov kernels and the tail chain’, Adv. in Appl. Probab. 45(1), 186–213.

Resnick, S. I. and Zeber, D. (2014), ‘Transition kernels and the conditional extreme value model’, Extremes 17, 263–287.

Rootzén, H. (1988), ‘Maxima and exceedances of stationary Markov chains’, Adv. in Appl. Probab. 20, 371–390.

Smith, R. L. (1992), ‘The extremal index for a Markov chain’, J. Appl. Probab. 29(1), 37–45.

Tawn, J. A. (1990), ‘Modelling multivariate extreme value distributions’, Biometrika 77, 245–53.

Walker, G. (1931), ‘On periodicity in series of related terms’, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 131(818), 518–532.

Yule, G. U. (1927), ‘On a method of investigating periodicities in disturbed series, with special reference to Wolfer’s sunspot numbers’, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 226(636-646), 267–298.

Yun, S. (1998), ‘The extremal index of a higher-order stationary Markov chain’, Ann. Appl. Probab. 8(2), 408–437.