

Tail behaviour and extremes of two-state Markov-switching autoregressive models

Péter Elek^{a,b,*}, András Zempléni^b

^a *Economic Policy Department, Ministry of Finance, Hungary*

^b *Department of Probability Theory and Statistics, Eötvös Loránd University, H-1117 Budapest, Pázmány Péter sétány 1/c, Hungary*

Received 13 July 2007; received in revised form 16 August 2007; accepted 27 September 2007

Abstract

We examine the tail behaviour and extremal cluster characteristics of two-state Markov-switching autoregressive models where the first regime behaves like a random walk, the second regime is a stationary autoregression, and the generating noise is light-tailed. Under additional technical conditions we prove that the stationary solution has asymptotically exponential tail and the extremal index is smaller than one. The extremal index and the limiting cluster size distribution of the process are calculated explicitly for some noise distributions, and simulated for others. The practical relevance of the results is illustrated by examining extremal properties of a regime-switching autoregressive process with Gamma-distributed noise, already applied successfully in river flow modeling. The limiting aggregate excess distribution is shown to possess Weibull-like tail in this special case.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Extremal clusters; Extremal index; Markov-switching autoregressive model; River flow modeling; Tail behaviour

1. Introduction

There has been an extensive development in the theory and applications of extreme value models for time series in the last decades. From theoretical and practical points of view, the tail behaviour of the stationary distribution of the time series and the clustering tendencies of its high values are equally important. For instance, in a hydrological setting, the tail behaviour describes the height of a flood, while the measures of extremal clustering (which show the extent to which high values occur together) give additional information on flood lengths, flood volumes, and hence on the severity of extreme events.

In this paper we examine the tail behaviour and extremal clusters of a particular class of nonlinear time series models, namely two-state Markov-switching first order autoregressions. These processes are governed by a latent Markov chain with two states (regimes), and they behave as an AR(1) model in each regime. Their extremal properties depend substantially on the stability of the dynamics in the particular regimes. If in both regimes the parameters lie within the open interval $(-1, 1)$ and the generating noise is light-tailed, the stationary solution is also light-tailed

* Corresponding author at: Department of Probability Theory and Statistics, Eötvös Loránd University, H-1117 Budapest, Pázmány Péter sétány 1/c, Hungary.

E-mail addresses: elekpeti@cs.elte.hu (P. Elek), zempleni@ludens.elte.hu (A. Zempléni).

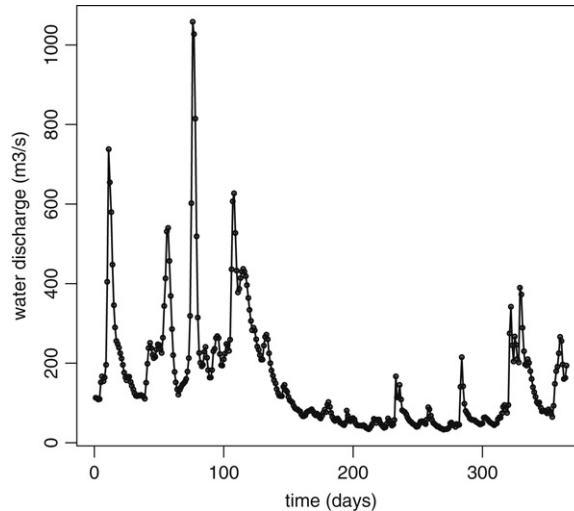


Fig. 1. A one year portion of the water discharge series at Tivadar (river Tisza, Hungary).

and there is no extremal clustering (meaning that, heuristically, the process behaves as an i.i.d. sequence above high thresholds). On the other hand, if in one of the regimes the AR(1) parameter is greater than one, we obtain an extremely clustered heavy-tailed output from a light-tailed input (see [1]). There remains the case when one of the parameters is exactly one, i.e. when the process behaves as a random walk in one regime.

This parameter choice – not yet studied previously in the literature – is qualitatively different from the above mentioned ones. We prove in this article that for a wide class of light-tailed noise distributions the tail of the stationary distribution will be asymptotically exponential, its scale parameter being determined by the moment-generating function of the noise. We also show that extremal clustering is present in the model, and, using Wiener–Hopf equations, give explicit results on the extremal index and also on the limiting cluster size distribution in some special cases. For other types of noise distributions, we proceed by simulation to determine these quantities.

Our research is motivated by the wide range of applications of such models e.g. in finance (see the sequence of articles originating from [2]), engineering or hydrology ([3–5]). Regarding the latter field, Fig. 1 shows a portion of a daily water discharge series measured at river Tisza in Hungary, and it is clear that short and steep rising periods are followed apparently randomly by longer, gradually falling ones on the graph. Taking into account this fact, [5] modelled the river flow series by a two-state regime-switching autoregressive model where the noise in the random walk regime is Gamma-distributed. In this paper we apply our theoretical findings to this model as well, and, furthermore, prove that the limiting aggregate excess distribution has Weibull-like tail in this Gamma-distributed case.

The paper is organised as follows. The model is defined and the assumptions are introduced in Section 2. Section 3 investigates the tail behaviour, while Section 4 analyses extremal clustering in the model. We draw the hydrological implications of our theoretical findings in Section 5, and conclude in Section 6. The Appendix contains the proof of Theorem 4.

2. The model and the assumptions

We define the X_t process as

$$X_t = a_1 X_{t-1} + \varepsilon_{1,t} \quad \text{if } I_t = 1, \quad (1)$$

$$X_t = a_0 X_{t-1} + \varepsilon_{0,t} \quad \text{if } I_t = 0, \quad (2)$$

where I_t is a two-state Markov chain with transition probabilities

$$p_1 = P(I_t = 0 \mid I_{t-1} = 1),$$

$$p_0 = P(I_t = 1 \mid I_{t-1} = 0).$$

We assume that $\{\varepsilon_{1,t}\}$ and $\{\varepsilon_{0,t}\}$ are both independent, identically distributed noise sequences (but the two distributions need not be the same), independent from each other and from the $\{I_t\}$ sequence as well.

It follows from [6] that the model has a unique stationary solution if

$$p_1 \log |a_0| + p_0 \log |a_1| < 0. \tag{3}$$

Hence, local stationarity (i.e. $|a_0| < 1$ and $|a_1| < 1$) is a sufficient but not necessary condition for the “global” stationarity of the model. Further probabilistic properties are given e.g. in [7].

Let us assume that $|a_1| \geq |a_0|$. If $|a_1| > 1$ but (3) still holds, it is easy to show that not all moments of X_t exist even when $\varepsilon_{1,t}$ is sufficiently light-tailed, e.g. normally distributed. In fact, under some additional conditions, it follows from [1] that for $|a_1| > 1$, there exist a $K > 0$ and $\lambda > 0$ (which certainly depend on a_1, a_0 and the distribution of the noises) such that $P(X_t > u) \sim K u^{-\lambda}$. Hence, the tail is typically regularly varying in this case. On the other hand, if $|a_0| \leq |a_1| < 1$, the precise form of the stationary solution depends very much on the noises, but it is certainly light-tailed without extremal clustering if the noise is light-tailed. (For more discussion, see Section 5.)

Our main focus is the following choice of parameters:

Assumption 1. $a_1 = 1$ and $0 \leq a_0 < 1$.

This assumption implies that the process behaves like a random walk in the first regime, while it is a stationary autoregression in the second one. (The $0 \leq a_0 < 1$ assumption can be weakened, see Remark 4.) The stationary solution for X_t always exists by (3), and unless otherwise indicated, all probability statements in the following correspond to this unique stationary distribution. The parameter choice is on the border of the two previously mentioned cases because – as it will turn out in Sections 3 and 4 – if the noise is light-tailed the stationary solution is light-tailed (as when $|a_1| < 1$), but there is asymptotic clustering of high values (as when $a_1 > 1$).

Throughout the paper, the following assumptions on the noise sequences will be applied. Assumption 2 essentially states that the upper tail of $\varepsilon_{1,t}$ is light (excluding e.g. subexponential noises), while Assumption 3 implies that $\varepsilon_{0,t}$ is not much heavier-tailed than the positive part of $\varepsilon_{1,t}$. In what follows, $L_X(s) = E(\exp(sX))$ denotes the moment-generating function, $F_X(\cdot)$ the distribution function, $\bar{F}_X(\cdot) = 1 - F_X(\cdot)$ the survival function and $f_X(\cdot)$ the density function of a random variable X . For two functions $g(\cdot)$ and $h(\cdot)$, $g(u) \sim h(u)$ means that $g(u)/h(u) \rightarrow 1$ as $u \rightarrow \infty$. We will omit to indicate $u \rightarrow \infty$ if no ambiguity arises. Constants are denoted by K , but they are not intended to mean the same constant all the time. $X_n \rightarrow_d X$ indicates convergence in distribution. Finally, the notations $y^+ = \max\{y, 0\}$ and $y^- = -\min\{y, 0\}$ are used.

Assumption 2. The distribution of $\varepsilon_{1,t}$ is absolutely continuous with respect to the Lebesgue-measure and $E|\varepsilon_{1,t}| < \infty$. Moreover, there exists a $\kappa > 0$ such that

$$(1 - p_1) L_{\varepsilon_{1,t}}(\kappa) = 1, \tag{4}$$

and $L'_{\varepsilon_{1,t}}(\kappa) < \infty$.

Assumption 3. The distribution of $\varepsilon_{0,t}$ is absolutely continuous with respect to the Lebesgue-measure, and its probability density is positive on the whole real line. With κ defined in (4), there exists an $s_0 > \kappa$ such that $L_{|\varepsilon_{0,t}|}(s_0) < \infty$.

The importance of Assumption 2 comes from the following observation. Let

$$S_0 = 0, \quad S_n = S_{n-1} + \varepsilon_n \quad (n = 1, 2, \dots) \tag{5}$$

be a random walk with step distribution the same as the distribution of $\varepsilon_{1,t}$, and let T be a $\text{Geom}(p_1)$ -distributed random variable, i.e. $P(T \geq k) = (1 - p_1)^{k-1}$, independent of S_n . Then, the asymptotic distribution of

$$M = \max\{S_i : i \leq T - 1\} \tag{6}$$

is given under Assumption 2 by

Proposition 1. Under Assumption 2 there exists a $K > 0$ such that

$$P(M > u) \sim K \exp(-\kappa u). \tag{7}$$

This is essentially the Cramér–Lundberg approximation (see [8, Chapter XII., Thms 5.2. and 5.3.]), but here the random walk has a defective step distribution. Let $S'_0 = 0, S'_n = S'_{n-1} + \varepsilon'_n$ ($n = 1, 2, \dots$) where ε'_n takes $-\infty$ with probability p_1 and is equal to ε_n with probability $1 - p_1$. Then $M = \max\{S'_i : 0 \leq i\}$, and the Cramér–Lundberg approximation states that the exponent κ of the tail of M comes from the equation $L_{\varepsilon'_n}(\kappa) = 1$. However, $L_{\varepsilon'_n}(s) = (1 - p_1) L_{\varepsilon_n}(s)$, hence (7) follows. The argument can be made precise by considering step distributions which take $-m$ with probability p_1 and the value of ε_n with probability $1 - p_1$, and allowing $m \rightarrow \infty$.

Assumption 2 is satisfied for a wide range of distributions, Examples 1–3 state a few practically important ones. Example 3 is particularly interesting because then the model is very similar to the one developed by [5] for river discharge series. The derivative condition of Assumption 2 is satisfied in all the cases below.

Example 1. If $\varepsilon_{1,t}$ is normally distributed with mean μ and variance σ^2 ,

$$\kappa = \frac{(\mu^2 - 2\sigma^2 \log(1 - p_1))^{1/2} - \mu}{\sigma^2}$$

is the positive solution of equation (4).

Example 2. Let $\varepsilon_{1,t}$ be distributed as skewed double exponential with parameters c, λ_L and λ_U , i.e. have probability density

$$f(x) = \frac{\lambda_L \lambda_U}{\lambda_L + \lambda_U} \exp(\lambda_L(x - c)) \quad \text{if } x < c,$$

$$f(x) = \frac{\lambda_L \lambda_U}{\lambda_L + \lambda_U} \exp(-\lambda_U(x - c)) \quad \text{if } x \geq c.$$

Then, κ is the solution of the equation

$$(1 - p_1) \exp(c\kappa) = \frac{(\kappa + \lambda_L)(\lambda_U - \kappa)}{\lambda_L \lambda_U}. \tag{8}$$

If $c = 0$ and $\lambda_L = \lambda_U = \lambda$, then $\kappa = p_1^{1/2} \lambda$.

Example 3. Let $\varepsilon_{1,t}$ be distributed as $\Gamma(\alpha, \lambda)$, then $\kappa = \lambda(1 - (1 - p_1)^{1/\alpha})$.

The following simple observations will be applied throughout the article.

Lemma 1. Let Q_1 be a random variable with $P(Q_1 > u) \sim K_1 \exp(-\kappa u)$, and let Q_2 be an independent variable with $L_{Q_2}(s) < \infty$ for an $s > \kappa$. Then,

$$P(Q_1 + Q_2 > u) \sim K_1 L_{Q_2}(\kappa) \exp(-\kappa u).$$

Proof. According to Breiman’s theorem (see [9]), if X and Y are two independent nonnegative random variables such that the tail of X is regularly varying with index $-\delta$ ($\delta > 0$) and $E(Y^{\delta+\eta}) < \infty$ for some $\eta > 0$, then $P(XY > v) \sim E(Y^\delta) P(X > v)$ as $v \rightarrow \infty$. Thus the statement follows with the choice $X = \exp(Q_1), Y = \exp(Q_2), \delta = \kappa$ and $\eta = s - \kappa > 0$. \square

Lemma 2. Let Q_1 and Q_2 be two independent random variables with tails $P(Q_i > u) \sim K_i \exp(-\kappa u)$ ($i = 1, 2$) and let $|a| < 1$. Then for every fixed $v \geq 0$, as $u \rightarrow \infty$,

$$P(aQ_1 + Q_2 > u + v \mid Q_1 > u) \sim K_2(1 - a)^{-1} \exp(-(1 - a)\kappa u) \exp(-\kappa v).$$

Proof. For $0 < a < 1$, let $\eta = (2 - a + 1/a)/2$ and for $-1 < a \leq 0$, let $\eta > 2 - a$ be arbitrary. Then $\eta > 1$ and $a\eta < 1$, hence for $u \rightarrow \infty$

$$P(aQ_1 + Q_2 > u + v, u < Q_1 \leq \eta u) = - \int_u^{\eta u} \bar{F}_{Q_2}(u + v - ax) d\bar{F}_{Q_1}(x)$$

$$\sim - \int_u^{\eta u} K_2 \exp(-\kappa(u + v - ax)) d\bar{F}_{Q_1}(x).$$

By partial integration, this is equal to

$$\begin{aligned}
 &= - \left[K_2 \exp(-\kappa(u + v - ax)) \bar{F}_{Q_1}(x) \right]_{x=u}^{x=\eta u} + \int_u^{\eta u} K_2 \kappa a \exp(-\kappa(u + v - ax)) \bar{F}_{Q_1}(x) dx \\
 &\sim - \left[K_2 \exp(-\kappa(u + v - ax)) K_1 \exp(-\kappa x) \right]_{x=u}^{x=\eta u} \\
 &\quad + \int_u^{\eta u} K_2 \kappa a \exp(-\kappa(u + v - ax)) K_1 \exp(-\kappa x) dx \\
 &= - \left[K_2 K_1 (1 - a)^{-1} \exp(-\kappa(u + v + (1 - a)x)) \right]_{x=u}^{x=\eta u} \\
 &\sim K_2 K_1 (1 - a)^{-1} \exp(-\kappa(u + v + (1 - a)u)) \\
 &\sim P(Q_1 > u) K_2 (1 - a)^{-1} \exp(-\kappa(1 - a)u) \exp(-\kappa v).
 \end{aligned}$$

Moreover, as $u \rightarrow \infty$, $P(Q_1 > \eta u) \sim P(Q_1 > u) \exp(-(\eta - 1)\kappa u)$ is negligible compared to the above term for every fixed $v \geq 0$ because $\eta - 1 > 1 - a$. Hence the statement follows. \square

3. Tail behaviour

As already noted, we examine the $a_1 = 1$ case. In what follows we will call the $I_t = 1$ regime the “random walk” or nonstationary regime, while the $I_t = 0$ regime as the stationary one. Since I_t is a Markov chain, regime durations are independent and geometrically distributed. To examine the extremal behaviour of our model, let us introduce a few auxiliary processes. Let ξ_m and ζ_m , respectively, denote the series of time points when the $I_t = 1$ and $I_t = 0$ regimes end. (The indexing is chosen to ensure that $\xi_{m-1} < \zeta_m < \xi_m < \zeta_{m+1}$.) For later reference, let $\gamma_m(u) = \min\{\zeta_m + 1 \leq t \leq \xi_m : X_t > u\}$ be the time of first reaching a threshold u in a nonstationary regime (and $\gamma_m(u)$ remains undefined if there is no such t). We use the notations $N_{1m} = \xi_m - \zeta_m$ and $N_{0m} = \zeta_m - \xi_{m-1}$. Finally, let $B(t) = \max(\xi_m \mid \xi_m \leq t)$ be the time of the end of the last nonstationary regime up to t , and similarly we use $D(t) = \max(\zeta_m \mid \zeta_m \leq t)$ for the time of the end of the last stationary regime.

Then $Y_m = X_{\xi_m}$ (the sequence of last values in the nonstationary regimes) is a Markov chain, and we expect that its tail behaviour characterises the tail of X_t . Similarly, $Z_m = X_{\zeta_m}$ (the series of last values in the stationary regimes) is a Markov chain as well. For later reference, let $M^{(m)} = \max\{X_t : \zeta_m + 1 \leq t \leq \xi_m\}$ be the maximum in a nonstationary regime.

The sequence Y_m can be written as

$$Y_m = Z_m + V_m = A_{0m} Y_{m-1} + U_m + V_m. \tag{9}$$

Here, $A_{0m} = a_0^{N_{0m}}$, $V_m = \sum_{t=\zeta_m+1}^{\xi_m} \varepsilon_{1,t}$ is a geometric random sum of i.i.d. variables (because the time spent in a regime is geometrically distributed), and $U_m = \sum_{t=\xi_{m-1}+1}^{\zeta_m} a_0^{\zeta_m-t} \varepsilon_{0,t}$ is a geometric random weighted sum of i.i.d. variables. If $m \neq k$, (A_{0m}, U_m, V_m) is independent of (A_{0k}, U_k, V_k) , but A_{0m} is not independent of U_m . Hence, standard results on the solutions of stochastic difference equations are not directly applicable.

The lemma below gives the tail behaviour of V_m . Based on this lemma, Theorem 1 states that X_t has asymptotically exponential upper tail.

Lemma 3. *Under Assumption 2, there exists a $K > 0$ constant such that*

$$P(V_m > u) \sim K \exp(-\kappa u).$$

Proof. The distribution of V_m is the same as that of S_T defined in (5) where T is $\text{Geom}(p_1)$ -distributed. Theorem 1 in [10] states that

$$S_{T-1} = \max\{S_n : n \leq T - 1\} + \min\{S_n : n \leq T - 1\}$$

if the terms on the right are added *independently*. Here the first term is just M defined in (6), which has $\text{Exp}(\kappa)$ -tail by Proposition 1. As the second term in the sum is nonpositive, Lemma 1 yields that S_{T-1} has the same tail, too. But $S_T = S_{T-1} + \varepsilon_1$, and $L_{\varepsilon_1}(s) < \infty$ for an $s > \kappa$ by Assumption 2. Hence Lemma 1 can be applied again to obtain the tail of S_T and thus of V_m . \square

Theorem 1. *If Assumptions 1–3 hold there is a constant $K > 0$ such that*

$$P(X_t > u) \sim K \exp(-\kappa u).$$

Proof. Let $L_0(s) = L_{|\varepsilon_{0,t}|}(s)$. By Jensen’s inequality, $L_0(as) \leq (L_0(s))^a$ for all $0 \leq a < 1$. According to Assumption 3, for all $s \leq s_0$ given there,

$$\log L_{|U_m|}(s) \leq \sum_{k=0}^{\infty} \log L_0(|a_0|^k s) \leq \sum_{k=0}^{\infty} |a_0|^k \log L_0(s) < \infty. \tag{10}$$

To determine the tail behaviour of X_t , we first use the drift condition for the stability of Markov chains (cf. [11]) to prove that for all $0 < s < \kappa$

$$L_{Y_m}(s) < \infty. \tag{11}$$

Clearly, Y_m is a ψ -irreducible and aperiodic Feller chain, with ψ being the Lebesgue-measure. Thus, by [11, Thms 5.5.7 and 6.0.1], every compact set is small and smallness is equivalent to petiteness. (For the definition and properties of small and petite sets, see [11, Chapter 5].) Following [11, Thms 14.0.1 and 15.0.1], for every $0 < s < \kappa$ it is enough to find a suitable test function $h \geq 1$ which satisfies $h(y) \geq \exp(sy)$ for $y \geq 0$, a compact set C and constants d and $0 < \beta < 1$ such that

$$E(h(Y_m) | Y_{m-1} = y) \leq (1 - \beta)h(y) + d\chi_C(y),$$

where χ_C denotes the indicator function of the set C . Then $E(h(Y_m)) < \infty$, and thus (11) also holds.

In our case we can choose $h(y) = y^- + \exp(sy^+)$. By (9), $Y_m^+ \leq a_0y^+ + U_m^+ + V_m^+$. Therefore, Lemma 3 and (10) yield

$$E(\exp(sY_m^+) | Y_{m-1} = y) \leq \exp(sa_0y^+) L_{U_m^+}(s) L_{V_m^+}(s) \leq K(s) \exp(sa_0y^+)$$

for all $0 < s < \kappa$. Furthermore, $\exp(sa_0y) / \exp(sy) \rightarrow 0$ as $y \rightarrow \infty$, and $E(Y_m^- | Y_{m-1} = y) \leq a_0y^- + E(U_m^-) + E(V_m^-)$. Thus, there exist a $C = \{y : |y| \leq N\}$ compact set and $\beta < 1 - a_0$ for which

$$E(Y_m^- + \exp(sY_m^+) | Y_{m-1} = y) \leq (1 - \beta)(y^- + \exp(sy^+)) + d\chi_C(y),$$

hence (11) is proven.

It follows that

$$E(\exp(rZ_m^+)) \leq E(\exp(ra_0Y_{m-1}^+)) E(\exp(rU_m^+)) < \infty$$

if $0 < r < \min\{\kappa/a_0, s_0\} > \kappa$. Since Y_m is the independent sum of Z_m and V_m , Lemma 1 with the choice $Q_1 = V_m$ and $Q_2 = Z_m$ immediately implies that Y_m has $\text{Exp}(\kappa)$ upper tail.

Finally, it is easy to show by the constant hazard property of the geometric distribution that the same asymptotic results hold for the tail of X_t in the whole nonstationary period (i.e. $X_t | (I_t = 1)$), not just of Y_m . On the other hand, Assumption 3 ensures that $E(\exp(sX_t) | I_t = 0) < \infty$ for all $0 < s < \min\{s_0, \kappa/a_0\} > \kappa$, hence the tail of X_t is completely determined by the $I_t = 1$ regime. Thus the theorem is proven. \square

Remark 1. It follows from the proof that $P(X_t > u | I_t = 0) / P(X_t > u | I_t = 1) \rightarrow 0$ as $u \rightarrow \infty$, hence $P(I_t = 1 | X_t > u) \rightarrow 1$ as $u \rightarrow \infty$.

Remark 2. Since there exists an $s > \kappa$ with $L_{\varepsilon_{1,t}}(s) < \infty$, a related consequence is that $(\gamma_m(u) - \zeta_m) | (M^{(m)} > u) \rightarrow \infty$ as $u \rightarrow \infty$, i.e. the time necessary to reach a high threshold u in a nonstationary regime goes to infinity.

Remark 3. By Proposition 1 (see also the proof of Lemma 3), $M^{(m)} - Z_m$ and hence also $M^{(m)}$ have asymptotically $\text{Exp}(\kappa)$ upper tail.

Remark 4. It is clear from the proof that the $0 \leq a_0$ assumption can be substituted with a weaker one which ensures that even when $-1 < a_0 < 0$, the lower tail of $\varepsilon_{1,t}$ does not influence the upper tail of Y_m . In particular, if $a_0 < 0$ and there exists a $\kappa_- > 0$ such that

$$(1 - p_1) L_{-\varepsilon_{1,t}}(\kappa_-) = 1,$$

then V_m has asymptotically $\text{Exp}(\kappa_-)$ lower tail and $a_0 V_m$ has $\text{Exp}(-\kappa_-/a_0)$ upper tail. Hence the upper tail of Y_m is not affected and the theorem is valid if $\kappa_-/\kappa > -a_0$. Or, in the case of $L_{-\varepsilon_{1,t}}(s) < \infty$ for all $s > 0$ (e.g. in Examples 1 and 3), $|a_0| < 1$ is sufficient for the statement of the theorem to hold.

Corollary 1. If $\varepsilon_{1,t}$ is normally distributed with mean μ and variance σ^2 , and Assumptions 1 and 3 hold,

$$P(X_t > u) \sim K \exp\left(-\frac{(\mu^2 - 2\sigma^2 \log(1 - p_1))^{1/2} - \mu}{\sigma^2} u\right).$$

Corollary 2. If $\varepsilon_{1,t}$ is distributed as skewed double exponential with parameters c, λ_L and λ_U , and Assumptions 1 and 3 hold,

$$P(X_t > u) \sim K \exp(-\kappa u),$$

where κ is the positive root of (8).

Corollary 3. If $\varepsilon_{1,t}$ is $\Gamma(\alpha, \lambda)$ -distributed and Assumptions 1 and 3 hold, then

$$P(X_t > u) \sim K \exp\left(-\lambda \left(1 - (1 - p_1)^{1/\alpha}\right) u\right).$$

4. Extremal clustering behaviour

Let X_1, X_2, \dots, X_n be a strictly stationary time series with marginal distribution function F and let u_n be a real-valued sequence such that

$$\lim_{n \rightarrow \infty} n(1 - F(u_n)) \rightarrow \tau > 0. \tag{12}$$

Here, as $F(u_n) \rightarrow 1$, u_n plays the role of a high threshold. Our aim is to examine the asymptotic distribution of an extremal functional

$$C_n(u) = \sum_{t=1}^n g(X_t - u)$$

as $n \rightarrow \infty$ and $u = u_n$ where g is a $\mathbf{R} \rightarrow \mathbf{R}_+$ function with $g(x) = 0$ for $x < 0$. Two examples include the total number of exceedances of u , which arises by choosing $g(x) = \chi_{(x>0)}$, and the aggregate excess above u , which is obtained by $g(x) = x^+$.

The following theorem, due to [12], essentially states that under some technical conditions, which most series of practical interest satisfy, the distribution of $C_n(u_n)$ converges as $n \rightarrow \infty$ to the distribution of a Poisson sum of i.i.d. variables. The essence of the theorem is that high-level exceedances of a stationary time series occur in clusters, and each cluster contributes independently to the determination of $C_n(u_n)$. The most basic measure of extremal clustering, the extremal index comes as the reciprocal of the average length of an extremal cluster. (Obviously, the extremal index is equal to one for an i.i.d. sequence, but may remain the same for dependent sequences, too. Examples include all stationary Gaussian AR(1) processes, see [13].)

Theorem 2 ([12]). Let us assume that the X_t process is strongly mixing, i.e. there exists a $\phi(l)$ function such that for all $A \in \sigma\{X_j, -\infty < j \leq t\}$ and $B \in \sigma\{X_j, t + l \leq j < \infty\}$, $|P(A \cap B) - P(A)P(B)| \leq \phi(l)$, and $\phi(l) \rightarrow 0$ as $l \rightarrow \infty$. Given such a function, we can define a p_n sequence satisfying

$$p_n \rightarrow \infty, \quad \frac{p_n}{n} \rightarrow 0, \quad \frac{n\phi(p_n)}{p_n} \rightarrow 0. \tag{13}$$

We also assume that, with p_n chosen this way and a u_n sequence defined in (12), the following conditions hold:

$$\lim_{p \rightarrow \infty} \lim_{n \rightarrow \infty} \sum_{k=p}^{p_n} P(X_k > u_n \mid X_0 > u_n) = 0 \tag{14}$$

and

$$E(C_{p_n}(u_n) \mid M_{1,p_n} > u_n) \leq K < \infty, \tag{15}$$

where $M_{1,p} = \max(X_1, X_2, \dots, X_p)$. Finally, we assume that there exists a C^* random variable such that

$$P(C^* \leq y) = \lim_{p \rightarrow \infty} \lim_{u \rightarrow \infty} P(C_p(u) \leq y \mid M_{1,p} > u). \tag{16}$$

Then the distribution of $C_n(u_n)$ converges as $n \rightarrow \infty$ to the distribution of $C_1^* + C_2^* + \dots + C_L^*$ where L is a Poisson random variable and C_1^*, C_2^*, \dots are independent, of each other and of L , random variables with the same distribution as C^* . The mean of L is $\theta\tau$ where τ is defined by (12) and θ is the extremal index. The latter can be calculated, for instance, as

$$\theta = \lim_{p \rightarrow \infty} \lim_{u \rightarrow \infty} \theta(u, p) = \lim_{p \rightarrow \infty} \lim_{u \rightarrow \infty} P(M_{1,p} \leq u \mid X_0 > u).$$

Not surprisingly, the key to analysing extremal functionals in our model is to examine their behaviour in a typical nonstationary regime exceeding a high threshold u . Thus let

$$C'(u) = \left(\sum_{t=\xi_m+1}^{\xi_m} g(X_t - u) \right) \mid (M^{(m)} > u) = \sum_{t=\gamma_m(u)}^{\xi_m} g(X_t - u).$$

To obtain the distributional limit of $C'(u)$ as $u \rightarrow \infty$, we first recall some facts about the behaviour of the random walk introduced in (5) above a high threshold u . Let $\tau_u = \min\{n : S_n > u\}$ be the entrance time to (u, ∞) and $B_u = (S_{\tau_u} - u) \mid (\tau_u < \infty)$ be the overshoot of u (when it exists). Then, if $E\varepsilon_n = E\varepsilon_{1,t} \geq 0$, it follows from [8, Chapter VII., Thm. 2.4.] that $P(\tau_u < \infty) = 1$ for all $u \geq 0$. Moreover, Assumption 2 automatically implies in this case that $E\varepsilon_{1,t}^2 < \infty$, hence [14, Chapter XVIII., Theorem 2.] and [8, Chapter VII., Theorem 2.1.] yield that $E(B_0) < \infty$ and $B_u \rightarrow_d B_\infty$ as $u \rightarrow \infty$. B_∞ has probability density function

$$f_{B_\infty}(y) = \bar{F}_{B_0}(y) / E(B_0) \tag{17}$$

for $y > 0$. On the other hand, if $E\varepsilon_{1,t} < 0$, then $L'_{\varepsilon_{1,t}}(0) < 0$ and so Assumption 2 ensures that there is a κ' such that $L_{\varepsilon_{1,t}}(\kappa') = 1$ and $L'_{\varepsilon_{1,t}}(\kappa') < \infty$. Hence, although $P(\tau_u < \infty) \rightarrow 0$ as $u \rightarrow \infty$ in this case, it is still true by [15] that the distributional limit of B_u exists (and we also denote this by B_∞).

Now, define the S_n^* random walk as

$$S_0^* = B_\infty, \quad S_n^* = S_{n-1}^* + \varepsilon_n \quad (n = 1, 2, \dots)$$

where – as earlier – ε_n is distributed as $\varepsilon_{1,t}$. Let T be a Geom(p_1)-distributed random variable, independent of S_n^* . Define

$$C^* = \sum_{k=0}^{T-1} g(S_k^*). \tag{18}$$

Then

Proposition 2. Under Assumptions 1–3, $C'(u) \rightarrow_d C^*$ as $u \rightarrow \infty$.

Proof. If $M^{(m)} > u$, let $\varepsilon^*(u) = X_{\gamma_m(u)} - u$. By conditioning on the value of the end of the last $I_t = 0$ regime (Z_m),

$$\bar{F}_{\varepsilon^*(u)}(y) = \int_{-\infty}^u \bar{F}_{B_{u-z}}(y) f_{Z_m \mid (M^{(m)} > u)}(z) dz + \bar{F}_{Z_m \mid (M^{(m)} > u)}(u + y). \tag{19}$$

By Remark 3, as $u \rightarrow \infty$,

$$f_{Z_m|M^{(m)}>u}(z) = P\left(M^{(m)} > u \mid Z_m = z\right) f_{Z_m}(z) / P\left(M^{(m)} > u\right) \\ \sim K \exp(-\kappa(u - z)) f_{Z_m}(z) / \exp(-\kappa u) = K \exp(\kappa z) f_{Z_m}(z).$$

It follows from the proof of Theorem 1 that $L_{Z_m}(s) < \infty$ for every $0 < s < \min\{\kappa/a_0, s_0\} > \kappa$, hence for every $\delta > 0$ there exists a z_0 such that $\lim_{u \rightarrow \infty} \bar{F}_{Z_m|M^{(m)}>u}(z_0) < \delta$. Thus the second term on the right-hand side of (19) and also the integral on (z_0, u) is negligible. Therefore, since $\lim_{u \rightarrow \infty} \bar{F}_{B_{u-z}}(y) = \bar{F}_{B_\infty}(y)$ for every fixed z and $y > 0$, we obtain that $\lim_{u \rightarrow \infty} \bar{F}_{g^*(u)}(y) = \bar{F}_{B_\infty}(y)$. Hence, as $u \rightarrow \infty$, $\{X_t : \gamma_m(u) \leq t \leq \xi_m\}$ behaves like $\{S_k^* : 0 \leq k \leq T - 1\}$, and the statement of the proposition holds. \square

Proposition 3 below states that the asymptotic distribution of an extremal functional in the case of our Markov-switching autoregressive model depends only on its behaviour in a typical $I_t = 1$ regime. The proof relies on Lemma 4 which states that there is no extremal clustering among values in different $I_t = 1$ regimes because they are asymptotically independent in the extreme value sense. That is, e.g. for the end points of two subsequent such regimes, $P(Y_m > u \mid Y_{m-1} > u) \rightarrow 0$ as $u \rightarrow \infty$.

Lemma 4. *Let $g(x) = 0$ for $x < 0$, and $g(x) = o(\exp(\kappa x))$ as $x \rightarrow \infty$. Then there exists a $K > 0$ such that for all j integers*

$$E(g(Y_m - u) \mid Y_{m-j} > u) \leq K \exp\left[-\kappa\left(1 - a_0^{|j|}\right)u\right]. \tag{20}$$

The same bound holds for $E(g(X_t - u) \mid X_{t-l} > u)$ provided that there are j stationary regimes in $(t - l, t)$, and also for $E(g(M^{(m)} - u) \mid M^{(m-j)} > u)$ where $M^{(k)}$ is the maximum of the k th nonstationary regime.

It follows with the choice $g(x) = \chi_{(x>0)}$ that as $u \rightarrow \infty$,

$$P(Y_m > u \mid Y_{m-1} > u) \rightarrow 0.$$

Proof. Let us first assume that $j = 1$. We know from the proof of Theorem 1 that $U_m + V_m$ and Y_{m-1} are independent and both have asymptotically $\text{Exp}(\kappa)$ tail. If $Y_{m-1} > 0$, $Y_m \leq a_0 Y_{m-1} + U_m + V_m$. Since $g(x) = o(\exp(\kappa x))$, the bound (20) for $j = 1$ follows from Lemma 2 with the choice $Q_1 = Y_{m-1}$, $Q_2 = U_m + V_m$ and $a = a_0$.

To prove the bound for $E(g(X_t - u) \mid X_{t-l} > u)$, Remark 1 implies that we only have to deal with the case when X_t and X_{t-l} are both in nonstationary regimes. Let $m = D(t)$ be the end of the last stationary regime before t , $Q_1 = X_{t-l}$ and $Q_2 = a_0(Y_{m-1} - X_{t-l}) + U_m + (X_t - Z_m)$. Then, for $Y_{m-1} > 0$, $X_t \leq a_0 Q_1 + Q_2$. By the constant hazard property of the geometric distribution, $X_t - Z_m$ has asymptotically $\text{Exp}(\kappa)$ -tail. $a_0(Y_{m-1} - X_{t-l})$ is lighter-tailed than $X_t - Z_m$, and independent of X_{t-l} , U_m and $(X_t - Z_m)$. Hence, after using Lemma 1 to obtain the tail of Q_2 , we can apply Lemma 2 with $a = a_0$ to get the required upper bound.

Finally, in the case of regime maxima,

$$E\left(g\left(M^{(m)} - u\right) \mid M^{(m-1)} > u\right) \leq E\left(g\left(M^{(m)} - u\right) \mid Y_{m-1} > u\right),$$

and for $Y_{m-1} > 0$, $M^{(m)} \leq a_0 Y_{m-1} + U_m + (M^{(m)} - Z_m)$. By Remark 3, $M^{(m)} - Z_m$ has $\text{Exp}(\kappa)$ -tail, and it is independent of $a_0 Y_{m-1}$ and U_m , hence the statement holds by Lemma 2.

The $j \neq 1$ cases can be treated similarly. \square

Proposition 3. *If $g(x) = 0$ for $x < 0$ and $g(x) = o(\exp(\kappa x))$ as $x \rightarrow \infty$, the conditions of Theorem 2 are satisfied with C^* defined by (18).*

Proof. Strong mixing of X_t can be proven either directly using the fact that Y_m is strongly mixing because of the existence of the test function constructed in the proof of Theorem 1, or along the lines of e.g. [7]. (Details are omitted.) It follows that $\phi(l) = K\rho^l$ for some $0 < \rho < 1$, hence p_n can be chosen as $K \log n$ with an appropriate $K > 0$. Meanwhile, Theorem 1 implies that $u_n \sim \log n / \kappa$ for u_n defined in (12).

Moreover, Lemma 4 and Theorem 1 yield that for all $m > 0$ and $g(x) = o(\exp(\kappa x))$,

$$E(g(X_t - u_n) \mid X_{t-l} > u_n) \leq K n^{a_0^j - 1},$$

provided that there are at least j stationary regimes in $(t - l, t)$. Hence, for each $0 < \eta < 1$,

$$E(g(X_t - u_n) \mid X_{t-l} > u_n) \leq Kn^{a_0^{\lfloor \eta l \rfloor - 1}} + K'P \text{ (less than } \lfloor \eta l \rfloor \text{ stationary regimes in } (t - l, t)).$$

Here the last term can be bounded from above by $K'P(N_0 > l/2) + K'P(N_1 > l/2)$ where N_0 is the sum of the durations of $\lfloor \eta l \rfloor$ independent stationary regimes and N_1 is the sum of the lengths of $\lfloor \eta l \rfloor$ nonstationary ones. Since regime durations are independent and geometrically distributed, $N_i - (\lfloor \eta l \rfloor - 1)$ ($i = 0, 1$) are negative binomially distributed with parameter $(\lfloor \eta l \rfloor, p_i)$. It can then be shown that if $\eta < p_i/2$, $P(N_i > l/2) \leq K_i \rho_i^l$ ($i = 0, 1$) with $0 < \rho_i < 1$. Therefore,

$$E(g(X_t - u_n) \mid X_{t-l} > u_n) \leq Kn^{a_0^{\eta l} - 1} + K_1 K' \rho_1^l + K_2 K' \rho_2^l, \tag{21}$$

hence

$$\sum_{k=p}^{P_n} E(g(X_k - u_n) \mid X_0 > u_n) \leq K p_n n^{a_0^{\eta p} - 1} + K'_1 \rho_1^p + K'_2 \rho_2^p,$$

which tends to 0 as $n \rightarrow \infty$ for all p . Thus, putting $g(x) = \chi_{(x>0)}$, the X_t process satisfies (14).

This argument also yields that the number of nonstationary regimes that exceed u_n in time interval $[1, p_n]$, provided that $M_{1,p_n} > u_n$, converges in probability to one as $n \rightarrow \infty$. Thus, using (21) again, (15) follows. Similarly, if $M_{1,p} > u$ for a fixed p , the distribution of $C_p(u)$ deviates from the distribution of $C'(u)$ only for two reasons. First, there may be at least two nonstationary regimes exceeding u in $[1, p]$ (the probability of this event vanishes as $u \rightarrow \infty$) and second, the single such regime (say, the m th one) may be too long to belong entirely to $[1, p]$. If the latter event occurs, $\xi_m - \gamma_m(u) \geq p^{1/2}$ or $\xi_m \in [1, p^{1/2}]$ or $\xi_m \in [p, p + p^{1/2}]$. Since $(\xi_m - \gamma_m(u)) \mid (M^{(m)} > u)$ is geometrically distributed,

$$\limsup_{u \rightarrow \infty} |P(C_p(u) \leq y \mid M_{1,p} > u) - P(C'(u) \leq y)| \leq 2p^{1/2} / (p + p^{1/2}) + (1 - p_1)p^{1/2}$$

for all p integer and $y \geq 0$. Letting $p \rightarrow \infty$ and using Proposition 2, (16) holds with C^* defined by (18). \square

In short, as long as extremes are concerned, our Markov-switching autoregressive model behaves like a Markov chain, moreover, like a random walk with defective step distribution function $H_F(x)$ putting p_1 mass to $-\infty$: $H_F(x) = p_1 + (1 - p_1) F_{\varepsilon_{1,t}}(x)$. It is also true that the joint distribution of (X_{t-1}, X_t) belongs to the domain of attraction of a bivariate extreme value distribution. The spectral measure H of this extreme value law (see e.g. [16]) is given by the relationship

$$H_F(x) = \int_{a(\kappa x)}^1 w dH(w),$$

where $a(x) = \exp(-x) / (1 + \exp(-x))$. (Hence $dH(1) = p_1$.)

The asymptotic Markovity, together with this domain of attraction result, means that the methodology developed by [17] and [12] to analyse the extremal clusters of Markov chains is applicable. In particular, the extremal index of the process – which is asymptotically the reciprocal of the length of an average cluster of high-level exceedances – can be calculated as

$$\theta = \int_{-\infty}^0 \kappa \exp(\kappa x) Q(x) dx, \tag{22}$$

where $Q(x)$ is the solution of the Wiener–Hopf equation (see [17])

$$Q(x) = \int_0^\infty Q(y) dH_F(x - y) = p_1 + (1 - p_1) \int_0^\infty Q(y) f_{\varepsilon_{1,t}}(x - y) dy. \tag{23}$$

In fact, $Q(x) = P(M < x)$ for $x > 0$ with M defined by (6). It follows from Assumption 2 that $P(\varepsilon_{1,t} > 0) > 0$, thus $P(M > 0) > 0$ which yields that θ , the extremal index is smaller than one. However, θ can be rarely obtained analytically because the Wiener–Hopf equation rarely has an explicit solution. A trivial exception is when $\varepsilon_{1,t} \geq 0$ a.s., then the extremal index is obviously p_1 . This is the case for Example 3. Another distribution for which an

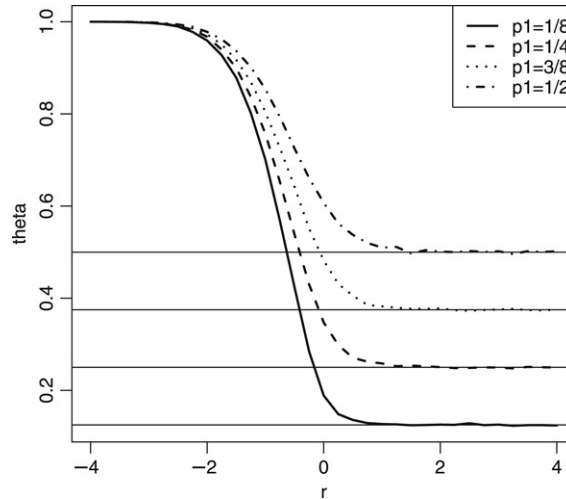


Fig. 2. θ as a function of $r = \mu/\sigma$ for $p_1 = 1/8, 1/4, 3/8$ and $1/2$ for the Gaussian model.

explicit expression is available is the double exponential one, introduced in Example 2. To illustrate this, the following Proposition gives the result for $c = 0$.

Proposition 4. *If $\varepsilon_{1,t}$ has skewed double exponential distribution with parameters $c = 0, \lambda_L$ and λ_U , the extremal index of X_t is*

$$\theta = p_1 + (1 - p_1) \left(\frac{\kappa}{\kappa + \lambda_L} \right)^2,$$

where κ is defined by (8). In particular, for $\lambda_U = \lambda_L$,

$$\theta = \frac{2p_1}{1 + p_1^{1/2}}. \tag{24}$$

Proof. The Wiener–Hopf equation can be solved explicitly because the right tail of $\varepsilon_{1,t}$ is now exponential. By [8, Chapter IX., Thm. 1.2.] (or by direct calculations), the maximum M of the stopped random walk has distribution function

$$Q(x) = P(M < x) = 1 - \left(1 - \frac{\kappa}{\lambda_U} \right) \exp(-\kappa x)$$

for $x > 0$. Hence, for $x < 0$, (23) yields

$$Q(x) = p_1 + (1 - p_1) \exp(\lambda_L x) \frac{\kappa}{\kappa + \lambda_L},$$

and finally integration in (22) shows the statement. \square

There is no closed form for $Q(x)$ in the Gaussian case (Example 1). Instead, we use simulations to approximate the extremal index in this setting. It is easy to show that θ is determined by p_1 and μ/σ , so Fig. 2 displays θ as a function of $r = \mu/\sigma$ for $p_1 = 1/8, 1/4, 3/8$ and $1/2$. Obviously, $\theta > p_1$ and $\theta \rightarrow p_1$ as $r \rightarrow \infty$. However, according to Fig. 3, θ is only moderately higher than p_1 even for $\mu = r = 0$. (The difference is highest when p_1 is neither close to 0 nor to 1.) Fig. 3 also displays $\theta - p_1$ as a function of p_1 for the symmetric double exponential case with $c = 0$ (see (24)). Clustering is higher (i.e. θ is lower) in the process driven by double exponential noise than in the one generated by Gaussian noise for the same p_1 and $\mu = c = 0$.

More elaborate extremal characteristics such as the limiting cluster size distribution or the limiting aggregate excess distribution can be calculated explicitly in even fewer cases. For instance, when $\varepsilon_{1,t} > 0$ a.s. (as in Example 3), $S_k^* > 0$ for all k , hence the limiting cluster size distribution is geometric with parameter p_1 . However, this is no longer the case in the Gaussian or double exponential setting.

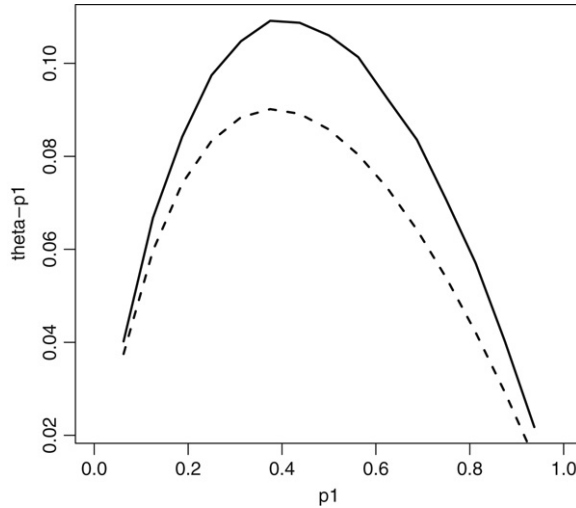


Fig. 3. $\theta-p_1$ for the Gaussian model (continuous) and for the double exponential model (dashed) as a function of p_1 if $E(\varepsilon_{1,t}) = 0$.

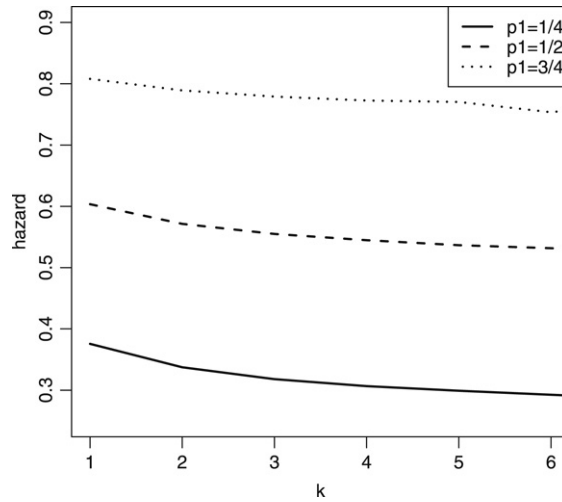


Fig. 4. Hazard function of the limiting cluster size distributions of the model with symmetric double exponential noise with zero mean for $p_1 = 1/4, 1/2$ and $3/4$.

To illustrate this, we simulated limiting cluster size distributions N for $p_1 = 0.25, 0.5$ and 0.75 when the generating noise is the symmetric double exponential distribution with $c = 0$. The calculated empirical hazard functions $P(N = k | N \geq k)$ – plotted in Fig. 4 – show that hazards are decreasing with k , therefore the cluster size distributions belong to the DFR (decreasing failure rate) family for these parameters.

5. Application

Vasas et al. [5] fitted a two-state regime-switching AR(1) model to water discharge series. In their setting, the following assumptions are made on the noises of the generating equations (1) and (2) and on the latent I_t process.

Assumption 4. $0 < a_1 \leq 1, 0 \leq a_0 < 1, \varepsilon_{1,t}$ is $\Gamma(\alpha, \lambda)$ -distributed and $\varepsilon_{0,t}$ is Gaussian with zero mean and σ^2 variance. The lengths of the $I_t = 1$ regimes are negative binomial with parameter (b, p_1) , while the $I_t = 0$ regime durations are geometrically distributed with parameter p_0 (and the durations are independent).

Since regime durations are geometrically distributed when I_t is Markovian, and the negative binomial distribution $\{q_k\}$ ($k \geq 1$) with parameter (b, p_1) is defined as

$$q_k = \frac{\Gamma(k + b - 1)}{\Gamma(k)\Gamma(b)} p_1^b (1 - p_1)^{k-1}$$

where $\Gamma(k)$ is the Gamma-function, we obtain the Markov-switching autoregressive case for $b = 1$. However, discrete IFR (increasing failure rate) and DFR duration distributions can also be modelled within this more general parametrisation, depending upon the sign of $b - 1$.

Also note that the distribution of $\varepsilon_{1,t}$ is as in Example 3, and $\varepsilon_{0,t}$ satisfies Assumption 3, hence the model fits directly into the previous setting if $b = a_1 = 1$ and $0 \leq a_0 < 1$. But even when $b \neq 1$ or $0 < a_1 < 1$, tail behaviour and extremes of the model can be examined by similar techniques as in Sections 3 and 4. Here, $V_m = \sum_{t=\xi_m+1}^{\xi_m} a_1^{\xi_m-t} \varepsilon_{1,t}$, which is now a weighted negative binomial random sum of i.i.d. Gamma-distributed variables, determines the tail of X_T . It can be shown (details are available from the authors upon request) that if $0 < a_1 < 1$, $P(V_m > u) \sim K \bar{F}_{\Gamma(\alpha, \lambda)}(u)$, and if $a_1 = 1$, $P(V_m > u) \sim Ku^{b-1} \exp(-\kappa u)$ with κ defined in Example 3. It follows that the model has Gamma-like tail irrespective of the value of $a_1 \in (0, 1]$ and b , but the shape and scale parameters are different in the various cases:

Theorem 3. *If Assumption 4 holds, there exists a $K > 0$ such that*

$$\begin{aligned} P(X_t > u) &\sim Ku^{\alpha-1} \exp(-\lambda u) \quad \text{if } 0 < a_1 < 1, \\ P(X_t > u) &\sim Ku^{b-1} \exp\left(-\lambda \left(1 - (1 - p_1)^{1/\alpha}\right) u\right) \quad \text{if } a_1 = 1. \end{aligned}$$

If $a_1 = 1$, Remarks 1–3 are still valid. Turning to the extremal clustering behaviour, if $a_1 = 1$, Remark 2 and basic properties of the negative binomial distribution imply that the distribution of $(\xi_m - \gamma_m(u) + 1) \mid (M^{(m)} > u)$, i.e. the distribution of the time spent in a typical $I_t = 1$ regime after first reaching u , tends to a geometric distribution as $u \rightarrow \infty$. Hence if $a_1 = 1$, Propositions 2 and 3 are still valid, and the essence of extremal clustering does not change compared to the Markov-switching case. However, if $0 < a_1 < 1$, it is easy to see that $P(X_t > u \mid X_{t-1} > u) \rightarrow 0$ as $u \rightarrow \infty$, so high observations do not make clusters asymptotically. Since, by Theorem 3, $P(X_t > u + z \mid X_t > u) \rightarrow \exp(-\lambda z)$ for all fixed $z > 0$, it follows that $C'(u) \rightarrow_d C^* = g(E)$ in this case where E is an $\text{Exp}(\lambda)$ -distributed random variable. The conditions of Theorem 2 hold with this C^* if $a_1 < 1$, thus applying the above for $g(x) = \chi_{(x>0)}$ we obtain

Proposition 5. *Let Assumption 4 hold. If $0 < a_1 < 1$, the limiting cluster size distribution of X_t takes value 1 with probability one, and the extremal index is 1. If $a_1 = 1$, the limiting cluster size distribution is geometric with parameter p_1 , and the extremal index is p_1 .*

One particular feature of the Gamma-model is that the tail of the limiting aggregate excess distribution can also be examined. Let

$$W_n(u_n) = \sum_{t=1}^n |X_t - u_n|^+,$$

then $W_n(u_n)$ converges in distribution to a Poisson sum of i.i.d. random variables distributed as W^* . If $0 < a_1 < 1$, W^* is exponential with parameter λ , but if $a_1 = 1$, the following Theorem states that it is heavier-tailed than the exponential distribution, having Weibull-like tail with exponent parameter $1/2$. The proof – given in the Appendix – relies on renewal theoretical arguments and on Laplace’s method of sums.

Theorem 4. *If Assumption 4 holds and $a_1 = 1$, there exist $K_i > 0$ ($i = 1, 2$) constants such that*

$$K_1 \exp\left(-2^{3/2} \left(\lambda_0^{-1} - \alpha \lambda_0\right) (\lambda y)^{1/2}\right) \leq \bar{F}_{W^*}(y) \leq K_2 \exp\left(-2 \left(\lambda_0^{-1} - \alpha \lambda_0\right) (\lambda y)^{1/2}\right), \tag{25}$$

where λ_0 is the unique real number satisfying

$$\lambda_0^{-2} - 2\alpha \log \lambda_0 + \log(1 - p_1) - \alpha(1 + \log \alpha) = 0. \tag{26}$$

According to [5], the dynamics of the daily river flow series measured at Tivadar (river Tisza in Hungary) can be well described by a model where $a_1 = 1$, the durations of the $I_t = 1$ regimes follow a negative binomial distribution with parameters $b = 4.8$ and $p_1 = 0.75$ (hence I_t is far from Markovian) and the rising Gamma-increments have $\alpha = 0.97$ (hence the increments are close to exponential). The $I_t = 0$ regimes are characterised by $a_0 = 0.815$ and $p_0 = 0.07$. Thus, our findings indicate that the river discharge series has Gamma-like tail with shape parameter $b = 4.8$ and there is extremal clustering because $a_1 = 1$. Nevertheless, the clustering is moderate at extreme levels, the extremal index is $p_1 = 0.75$, so the average size of a cluster of high-level exceedances is just $1/p_1 = 1.33$. (However, we stress that the subasymptotic behaviour – which can also be simulated from the fitted model – substantially differs from this because the a_0 parameter is close to 1 and thus a very high threshold is needed to ensure that the $I_t = 0$ regime does not influence the behaviour of the threshold exceedances.)

Moreover, Theorem 4 allows the calculation of the tail of the aggregate excess (flood volume) distribution, based on the parameters of the fitted model. The theorem is particularly interesting because it gives a theoretical background of the method advocated by [18]. There, based purely on empirical analyses, aggregate excesses of hydrological time series were proposed to be modelled by a Weibull-distribution. Our theoretical result suggests that the aggregate excess distribution of river flow series is not exactly Weibull but it indeed has a Weibull-like tail.

6. Conclusions

We have investigated the tail behaviour and extremal clustering of a general class of regime-switching autoregressive models. Our results are not only theoretically interesting but prove to be practically useful, too. Provided that our physically motivated regime-switching time series model describes adequately the behaviour of hydrological processes (the fit is discussed in [5]), the tail (i.e. high quantiles) and the extremal cluster functionals of water discharge series may be estimated on the basis of the whole time series, not just on high-level exceedances. This procedure substantially reduces the estimation errors of important characteristics of, for instance, the flood length or flood volume distributions.

Acknowledgement

The authors are grateful to Tamás Móri for his helpful suggestions.

Appendix

Proof of Theorem 4. We may assume that $\lambda = 1$. By Propositions 2 and 3, and since $S_k^* > 0$ for all k in the Gamma-case,

$$W^* = \sum_{k=0}^{T-1} (S_k^*)^+ = TB_\infty + \sum_{k=1}^{T-1} (T - k)\varepsilon_k. \tag{27}$$

T is a Geom(p_1)-distributed random variable, ε_k ($k = 1, 2, \dots$) are distributed as $\Gamma(\alpha, 1)$ and the density of B_∞ is given by (17). Since $\varepsilon_i \geq 0$ a.s., B_0 in (17) is also distributed as $\Gamma(\alpha, 1)$.

Let us first examine the $\alpha \geq 1$ case. Then the $\Gamma(\alpha, 1)$ distribution belongs to the class of NBUE-distributions (“new better than used in expectation”), thus for all $y \in \mathbf{R}^+$

$$\bar{F}_{B_\infty}(y) \leq \bar{F}_{\Gamma(\alpha,1)}(y). \tag{28}$$

Since $\bar{F}_{\Gamma(\alpha,1)}(v) \geq \exp(-v) E(\Gamma(\alpha, 1))$ for all $v \geq 0$, a lower bound can also be given straightforwardly for $\bar{F}_{B_\infty}(y)$:

$$\bar{F}_{B_\infty}(y) = \frac{\int_y^\infty \bar{F}_{\Gamma(\alpha,1)}(v) dv}{E(\Gamma(\alpha, 1))} \geq \bar{F}_{\text{Exp}(1)}(y).$$

To obtain the upper bound for the tail of W^* we first observe that $W^* \leq R_1 = T \left(\sum_{k=0}^{T-1} \varepsilon_k \right)$ by (27) and (28) (here ε_0 is also chosen as $\Gamma(\alpha, 1)$), hence it is enough to examine the tail of R_1 . Since $R_1 \mid (T = k)$ is distributed as $\Gamma(k\alpha, 1/k)$, we obtain for $y \rightarrow \infty$ that

$$f_{R_1}(y) = \sum_{k=1}^{\infty} p_1 (1 - p_1)^{k-1} \frac{1}{\Gamma(k\alpha)} \left(\frac{1}{k}\right)^{k\alpha} y^{k\alpha-1} \exp(-y/k) = \sum_{k=1}^{\infty} \exp(h(y, k)), \tag{29}$$

where

$$h(y, k) = \log p_1 + (k - 1) \log(1 - p_1) - \log \Gamma(k\alpha) + (k\alpha - 1) \log y - k\alpha \log k - y/k.$$

The asymptotic behaviour of this sum as $y \rightarrow \infty$ can be examined by Laplace’s method for sums (see e.g. [19]). After finding the location of the maximum $k_{\max}(y)$ of the function $k \rightarrow h(y, k)$, we use Taylor series expansion around $k_{\max}(y)$ to obtain the following result (the proof is given later):

Lemma 5. *There exists a $K > 0$ constant such that with λ_0 defined by (26)*

$$f_{R_1}(y) \sim Ky^{-1/2} \exp\left(-2\left(\lambda_0^{-1} - \alpha\lambda_0\right)y^{1/2}\right). \tag{30}$$

Integrating this directly gives the upper bound in (25). To give a lower bound for $P(W^* > y)$, let us introduce $R_2 = (T - 1) \left(\sum_{k=1}^{T-1} \varepsilon_k \right)$. (The notation implies that R_2 takes 0 with probability p_1 .) Clearly, for $y > 0$, $f_{R_2}(y) = (1 - p_1) f_{R_1}(y)$, hence the approximation in Lemma 5 – though with a different constant – applies for $f_{R_2}(y)$. Moreover, the variables $W'_1 = \sum_{k=1}^{T-1} (k - 1/2) \varepsilon_k$ and $W''_1 = \sum_{k=1}^{T-1} (T - k - 1/2) \varepsilon_k$ are identically distributed, take only positive values and $R_2 = W'_1 + W''_1$, hence $\bar{F}_{R_2}(2y) \leq 2\bar{F}_{W''_1}(y)$ for all y . Additionally, by (27), $W^* > W''_1$ a.s., hence $\bar{F}_{R_2}(2y) \leq 2\bar{F}_{W^*}(y)$, which yields the lower bound in (25). This concludes the proof when $\alpha \geq 1$.

When $\alpha < 1$, similar calculations give $\bar{F}_{\Gamma(\alpha,1)}(y) \leq \bar{F}_{B_\infty}(y) \leq \bar{F}_{\text{Exp}(1)}(y)$ for all $y > 0$ hence the lower bound for $P(W^* > y)$ can be obtained by observing that the variables $W'_2 = \sum_{k=0}^{T-1} (k + 1/2) \varepsilon_k$ and $W''_2 = \sum_{k=0}^{T-1} (T - k - 1/2) \varepsilon_k$ are identically distributed, their sum is R_1 and they are stochastically smaller than W^* . On the other hand, $R_3 = \sum_{k=T}^{T+\lceil 1/\alpha \rceil} \varepsilon_k$ is distributed as $\Gamma(\alpha\lceil 1/\alpha \rceil, 1)$, hence is stochastically larger than an $\text{Exp}(1)$ -distributed variable. Therefore, W^* is stochastically smaller than $R_4 = T \left(\sum_{k=1}^{T+\lceil 1/\alpha \rceil} \varepsilon_k \right)$ and this can be easily shown to have the same tail as R_1 , though with a different constant. This also concludes the proof for $\alpha < 1$.

Proof of Lemma 5. Denoting the logarithmic derivative of the Gamma-function by $\Psi(\cdot)$ and using $\Psi(x) = \log x + O(x^{-1})$, $\Psi'(x) = x^{-1} + O(x^{-2})$ and $\Psi''(x) = O(x^{-2})$ (see [20]) we obtain that

$$\begin{aligned} h'_k(y, k) &= \log(1 - p_1) - \alpha \Psi(k\alpha) + \alpha \log y - \alpha \log k - \alpha + yk^{-2} \\ &= yk^{-2} + \alpha \log(yk^{-2}) + \log(1 - p_1) - \alpha(1 + \log \alpha) + O(k^{-1}) \\ h''_k(y, k) &= -\alpha^2 \Psi'(k\alpha) - \alpha k^{-1} - 2yk^{-3} = -2\alpha k^{-1} - 2yk^{-3} + O(k^{-2}) \\ h'''_k(y, k) &= O(k^{-2}) + O(yk^{-4}). \end{aligned}$$

Solving the equation $h'_k(y, k) = 0$ for k yields $k_{\max}(y) = \lambda_0 y^{1/2} + O(1)$ with λ_0 defined by (26). (It is easy to check that λ_0 satisfies $0 < \lambda_0 < \alpha^{-1/2}$.) If we use the notation $k_y = \lambda_0 y^{1/2}$ and apply Stirling’s formula, we obtain from above and from the definition of λ_0 that

$$\begin{aligned} h(y, k_y) &= \lambda_0 y^{1/2} \left(\log(1 - p_1) - \alpha \log \alpha + \alpha - 2\alpha \log \lambda_0 - \lambda_0^{-2} \right) - 3/4 \log y + K + O(y^{-1/2}) \\ &= -2\lambda_0 y^{1/2} \left(\lambda_0^{-2} - \alpha \right) - 3/4 \log y + K + O(y^{-1/2}) \\ h'_k(y, k_y) &= O(y^{-1/2}) \\ h''_k(y, k_y) &= -2\lambda_2 y^{-1/2} + O(y^{-1}) \end{aligned}$$

$$h_k'''(y, k_y) = O(y^{-1}),$$

where $\lambda_2 = \alpha\lambda_0^{-1} + \lambda_0^{-3}$.

To examine the sum in (29) we distinguish between different values of k . If $|k - k_y| < y^{3/10}$, a Taylor-series expansion around k_y gives

$$\begin{aligned} h(y, k) - h(y, k_y) &= (k - k_y) O(y^{-1/2}) - (k - k_y)^2 (\lambda_2 y^{-1/2} + O(y^{-1})) \\ &\quad + (k - k_y)^3 O(y^{-1}) = - (k - k_y)^2 \lambda_2 y^{-1/2} + O(y^{-1/10}). \end{aligned} \quad (31)$$

Therefore, as $y \rightarrow \infty$,

$$\begin{aligned} \sum_{|k - k_y| < y^{3/10}} \exp(h(y, k) - h(y, k_y)) &\sim \sum_{|j| < y^{3/10}} \exp\left(-\frac{1}{2} \left(\frac{2^{1/2} \lambda_2^{1/2}}{y^{1/4}} j\right)^2\right) \\ &\sim \frac{y^{1/4}}{(2\lambda_2)^{1/2}} \int_{-(2\lambda_2)^{1/2} y^{1/20}}^{(2\lambda_2)^{1/2} y^{1/20}} \exp(-t^2/2) dt \sim Ky^{1/4}. \end{aligned}$$

On the other hand, using the fact that $k \rightarrow h'_k(y, k)$ is a decreasing function for all y and $h'_k(y, k_y) = 0$, we obtain from (31) that for all $|k - k_y| \geq y^{3/10}$

$$h(y, k) - h(y, k_y) \leq -y^{6/10} \lambda_2 y^{-1/2} + O(y^{-1/10}).$$

Moreover, if $k > y$ then $h(y, k) < -\log \Gamma(k\alpha)$. Hence, as $y \rightarrow \infty$,

$$\sum_{\{k > 0; |k - k_y| \geq y^{3/10}\}} \exp(h(y, k) - h(y, k_y)) \leq Ky \exp(-\lambda_2 y^{1/10}) + \exp(-h(y, k_y)) \sum_{k > y} 1/\Gamma(k) = o(1).$$

Putting together the above estimates yields

$$\begin{aligned} f_{R_1}(y) &\sim \exp(h(y, k_y)) \sum_{k=1}^{\infty} \exp(h(y, k) - h(y, k_y)) \\ &\sim Ky^{1/4} \exp(h(y, k_y)) \sim Ky^{-1/2} \exp\left(-2(\lambda_0^{-1} - \alpha\lambda_0) y^{1/2}\right). \end{aligned}$$

References

- [1] B. Saporta, Tail of the stationary solution of the stochastic equation $Y_{n+1} = a_n Y_n + b_n$ with Markovian coefficients, *Stochastic Process. Appl.* 115 (2005) 1954–1978.
- [2] J.D. Hamilton, Analysis of time series subject to changes in regime, *J. Econometrics* 45 (1990) 39–70.
- [3] Z.Q. Lu, L.M. Berliner, Markov switching time series models with application to a daily runoff series, *Water Resour. Res.* 35 (2) (1999) 523–534.
- [4] J. Szilágyi, G. Bálint, A. Csík, Hybrid, Markov chain-based model for daily streamflow generation at multiple catchment sites, *J. Hydrol. Eng.* 11 (2006) 245–256.
- [5] K. Vasas, P. Elek, L. Márkus, A two-state regime switching autoregressive model with an application to river flows, *J. Statist. Plann. Inference* 137 (2007) 3113–3126.
- [6] A. Brandt, The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients, *Adv. in Appl. Probab.* 18 (1986) 211–220.
- [7] J.F. Yao, J.G. Attali, On stability of nonlinear AR-processes with Markov switching, *Adv. in Appl. Probab.* 32 (2000) 394–407.
- [8] S. Asmussen, *Applied Probability and Queues*, Wiley, Chichester, 1987.
- [9] L. Breiman, On some limit theorems similar to the arc-sin law, *Theory Probab. Appl.* 10 (1965) 323–331.
- [10] P. Greenwood, Wiener–Hopf decomposition of random walks and Lévy-processes, *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 34 (1971) 193–198.
- [11] S.P. Meyn, R.L. Tweedie, *Markov Chains and Stochastic Stability*, Springer, London, 1993.
- [12] R.L. Smith, J.A. Tawn, S.G. Coles, Markov chain models for threshold exceedances, *Biometrika* 84 (1997) 249–268.
- [13] P. Embrechts, C. Klüppelberg, T. Mikosch, *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin, 1997.
- [14] W. Feller, *An Introduction to Probability Theory and Its Applications*, Wiley, New York, 1971.

- [15] S. Asmussen, Conditioned limit theorems relating a random walk to its associate, with applications to risk reserve processes and the GI/G/1 queue, *Adv. in Appl. Probab.* 14 (1) (1982) 143–170.
- [16] S. Coles, J. Tawn, Modelling extreme multivariate events, *J. R. Statist. Soc. B* 53 (1991) 377–392.
- [17] R.L. Smith, The extremal index for a Markov chain, *J. Appl. Probab.* 29 (1992) 37–45.
- [18] C.W. Anderson, G.P. Dancy, The severity of extreme events, Research Report 409/92, University of Sheffield, 1992.
- [19] C.M. Bender, S.A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers*, Springer, Berlin, 1999.
- [20] M. Abramowitz, I.A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, in: National Bureau of Standards Applied Mathematics Series, vol. 55, 1965.