Stochastic Processes and their Applications 85 (2000) 189–207

stochastic processes and their applications

www.elsevier.com/locate/spa

Extremal behavior of the autoregressive process with ARCH(1) errors

Milan Borkovec [1]

Center of Mathematical Sciences, Munich University of Technology, D-80290 Munich, Germany

Received 2 September 1998; received in revised form 18 March 1999

Abstract

We investigate the extremal behavior of a special class of autoregressive processes with ARCH(1) errors given by the stochastic difference equation

$$X_n = \alpha X_{n-1} + \sqrt{\beta + \lambda X_{n-1}^2}\varepsilon_n, \quad n \in \mathbb{N},$$

where $(\varepsilon_n)_{n\in\mathbb{N}}$ are i.i.d. random variables. The extremes of such processes occur typically in clusters. We give an explicit formula for the extremal index and the probabilities for the length of a cluster. © 2000 Elsevier Science B.V. All rights reserved.

MSC: Primary: 60G70; 60J05; Secondary: 60F05; 60G55

Keywords: ARCH model; Autoregressive process; Compound Poisson process; Coupling; Extremal behavior; Extremal index; Fréchet distribution; Heavy tail; Heteroscedastic homogeneous Markov process; Recurrent Harris chain; Separating sequence; Strong mixing

1. Introduction

Random recurrence equations have been used in numerous fields of applied probability. We refer, for instance, to Kesten (1973), Vervaat (1979) and Embrechts and Goldie (1994). Stochastic models in finance are an important field of application for random recurrence equations. Over the last years a variety of these models have been suggested as appropriate models for financial time series (see e.g. Priestley, 1988; Tong, 1990; Taylor, 1986). Due to the random recurrence structure, many of these models possess the property that their conditional variance depends on the past information (conditional heteroscedasticity). Empirical work has confirmed that such models fit quite many types of financial data. The most known examples of volatility models in finance with random recurrence structure are autoregressive conditionally heteroscedastic (ARCH) processes and generalized ARCH (GARCH) processes. These models were introduced by Engle (1982) and Bollerslev (1986), respectively. They serve as special

[1] Present address: School of Operations Research and Industrial Efficiency, 232 Rhodes Hall, Cornell University, Ithaca, NY 14853-3801, USA.

E-mail address: borkovec@orie.cornell.edu (M. Borkovec)

0304-4149/00/$ - see front matter © 2000 Elsevier Science B.V. All rights reserved.
PII: S0304-4149(99)00073-3

exchange rate or asset price models and are very popular in econometrics. In a series of papers, the ARCH and GARCH models have been analyzed and generalized, see for instance the survey article by Bollerslev et al. (1992) and the statistical review paper by Shephard (1996).

The class of autoregressive (AR) models with (G)ARCH errors proposed first by Weiss (1984) are a natural extension of ARCH and GARCH processes. They are defined by the random recurrence equation

$$X_n = f(X_{n-1},\dots,X_{n-k}) + \sigma_n \varepsilon_n, \quad n \geqslant k, \tag{1.1}$$

where f is a linear function in its arguments, the innovations $(\varepsilon_n)_{n\in\mathbb{N}}$ are i.i.d. symmetric random variables with mean zero and σ_n is given by

$$\sigma_n^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j X_{n-j}^2 + \sum_{l=1}^{q} \beta_l \sigma_{n-l}^2, \tag{1.2}$$

where $\alpha_0 > 0$, $\alpha_1,\dots,\alpha_p \geqslant 0$, $\alpha_p > 0$, $\beta_1,\dots,\beta_q \geqslant 0$, $\beta_q > 0$ for some $p \geqslant 1$ and $q \geqslant 0$ with the convention that $\sum_{l=1}^{0} \beta_l \sigma_{n-l}^2 = 0$. These models combine the advantages of AR models which target more on the conditional mean of X_n given the past and ARCH and GARCH models which concentrate on the conditional variance of X_n (given the past). Autoregressive models with (G)ARCH errors capture the structure of financial data quite well, i.e. the tendency of volatility clustering and the fact that unconditional price and return distributions tend to have fatter tails than the normal distribution. Statistical and/or probabilistic properties of such models have been investigated, for instance, by Weiss (1984), Diebolt and Guégan (1990), Maercker (1997) and Borkovec and Klüppelberg (1998).

In the present paper we study the extremal behavior of autoregressive processes of order 1 with ARCH(1) errors, i.e. $f(X_{n-1},\dots,X_{n-k}) = \alpha X_{n-1}$ for some $\alpha \in \mathbb{R}$ and σ_n is given in (1.2) with $p = 1$ and $q = 0$. This Markovian model is analytically tractable and serves as a prototype for the larger class of models (1.1). In particular, the results in this paper can be seen as a step towards the theoretical description of the extremal behavior of GARCH-type models which are more successful in practice. Note that in the special case $\alpha = 0$ we get just the ARCH(1) model of Engle (1982) and hence our results for the extremes are an extension of the results in de Haan et al. (1989).

Extremal behavior of a Markov process $(X_n)_{n\in\mathbb{N}}$ is, for instance, manifested in the asymptotic behavior of the maxima

$$M_n = \max_{1 \leqslant k \leqslant n} X_k, \quad n \geqslant 1.$$

The limit behavior of M_n is a well-studied problem in extreme value theory. Two review paper on this and related problems are Rootzén (1988) and Perfekt (1994). For a general overview of extremes of Markov processes, see also Leadbetter et al. (1983) and the references therein. Loosely speaking, under quite general mixing conditions, one can show that for n and x large

$$P(M_n \leqslant x) \approx F^{n\theta}(x), \tag{1.3}$$

where F is the stationary distribution function of $(X_n)_{n\in\mathbb{N}}$ and $\theta \in [0,1]$ is a constant called *extremal index*. A natural interpretation of θ is that of the reciprocal of mean cluster size (see e.g. Embrechts et al. (1997, Chapter 6) and the references therein). The practical implication of (1.3) is that dependence in data does often not invalidate the application of classical extreme value theory. There are many methods for determing the extremal index. However, most are very technical and often useless in practice. An alternative is then to estimate θ from the data.

For the AR(1) process with ARCH(1) errors, we derive an explicit formula for the extremal index. Moreover, we investigate the point process of exceedances of a high threshold u of $(X_n)_{n\in\mathbb{N}}$ which characterizes the extremal behavior of the process in detail. This point process converges in distribution to a compound Poisson process with a well-specified intensity and a well-specified distribution of the size of the jumps.

The paper is organized as follows: in Section 2 we present the model and introduce the required assumptions on the innovations $(\varepsilon_n)_{n\in\mathbb{N}}$. The conditions are the same as in Borkovec and Klüppelberg (1998), namely the so-called general conditions and the technical conditions (D.1)–(D.3). The general conditions guarantee the existence of a stationary version of $(X_n)_{n\in\mathbb{N}}$ whereas (D.1)–(D.3) allow us to describe the tail behavior of the stationary distribution. Furthermore, we present in Theorem 2.1 some results on the AR(1) process with ARCH(1) errors $(X_n)_{n\in\mathbb{N}}$ which were basically proved in Borkovec and Klüppelberg (1998). In particular, the stationary distribution of $(X_n)_{n\in\mathbb{N}}$ has a Pareto-like tail. Later, we construct an auxiliary process $(Z_n)_{n\in\mathbb{N}}$ which has the same law in distribution as the process $(\ln(X_n^2))_{n\in\mathbb{N}}$. It turns out that the process $(Z_n)_{n\in\mathbb{N}}$ is the key to the description of the extremal behavior of $(X_n)_{n\in\mathbb{N}}$. The main reason for the process is Lemma 2.3 where we show that $(Z_n)_{n\in\mathbb{N}}$ behaves above a high threshold asymptotically as a random walk with negative drift. Section 3 contains the main results (Theorem 3.1) concerning the extremal behavior of $(X_n)_{n\in\mathbb{N}}$. We interprete these results and present some simulations. We conclude the paper in Section 4 with the proof of Theorem 3.1.

2. Preliminaries

We consider an autoregressive model of order 1 with autoregressive conditional heteroscedastic errors of order 1 (AR(1) model with ARCH(1) errors) which is defined by the stochastic difference equation

$$X_n = \alpha X_{n-1} + \sqrt{\beta + \lambda X_{n-1}^2}\,\varepsilon_n, \quad n \in \mathbb{N}, \tag{2.1}$$

where $(\varepsilon_n)_{n\in\mathbb{N}}$ are i.i.d. random variables, $\alpha \in \mathbb{R}$, $\beta, \lambda > 0$ and the parameters α and λ satisfy in addition the inequality

$$E(\ln|\alpha + \sqrt{\lambda}\varepsilon|) < 0. \tag{2.2}$$

This condition is required to guarantee the existence and uniqueness of a stationary distribution. Here ε is a generic random variable with the same distribution as ε_n. Throughout this paper, we assume the same conditions for ε as in Borkovec and

Klüppelberg (1998). These are the so-called *general conditions*:

$$\begin{aligned}&\varepsilon \text{ is symmetric with continuous Lebesgue density } p(x),\\ &\varepsilon \text{ has full support } \mathbb{R},\\ &\text{the second moment of } \varepsilon \text{ exists,}\end{aligned} \tag{2.3}$$

and the *technical conditions* (D.1)–(D.3):

(D.1) $p(x) \geqslant p(x')$ for any $0 \leqslant x < x'$.

(D.2) For any $c \geqslant 0$ there exists a constant $q = q(c) \in (0,1)$ and functions $f_+(c,\cdot)$, $f_-(c,\cdot)$ with $f_+(c,x), f_-(c,x) \to 1$ as $x \to \infty$ such that for any $x > 0$ and $t > x^q$

$$p\left(\frac{x+c+\alpha t}{\sqrt{\beta+\lambda t^2}}\right) \geqslant p\left(\frac{x+\alpha t}{\sqrt{\beta+\lambda t^2}}\right) f_+(c,x),$$

$$p\left(\frac{x+c-\alpha t}{\sqrt{\beta+\lambda t^2}}\right) \geqslant p\left(\frac{x-\alpha t}{\sqrt{\beta+\lambda t^2}}\right) f_-(c,x).$$

(D.3) There exists a constant $\eta > 0$ such that

$$p(x) = o(x^{-(N+1+\eta+3q)/(1-q)}) \quad \text{as } x \to \infty,$$

where $N := \inf\{u \geqslant 0;\ E(|\sqrt{\lambda}\varepsilon|^u) > 2\}$ and q is the constant in (D.2).

There exists a wide class of distributions which satisfy these assumptions. Examples are the normal distribution, the Laplace distribution or the Students distribution. Conditions (D.1)–(D.3) are necessary for determing the tail of the stationary distribution. For further details concerning the conditions and examples we refer to Borkovec and Klüppelberg (1998). Note that the process $(X_n)_{n\in\mathbb{N}}$ is evidently a homogeneous Markov chain with state space $\mathbb{R}$ equipped with the Borel σ-algebra. The transition kernel density is given by

$$P(X_1 \in \mathrm{d}y \mid X_0 = x) = \frac{1}{\sqrt{\beta+\lambda x^2}}\, p\left(\frac{y-\alpha x}{\sqrt{\beta+\lambda x^2}}\right) \mathrm{d}y. \tag{2.4}$$

The next theorem collects some results on $(X_n)_{n\in\mathbb{N}}$ from Borkovec and Klüppelberg (1998).

Theorem 2.1. *Consider the process* $(X_n)_{n\in\mathbb{N}}$ *in* (2.1) *with* $(\varepsilon_n)_{n\in\mathbb{N}}$ *satisfying the general conditions* (2.3) *and with parameters* α *and* λ *satisfying* (2.2). *Then the following assertions hold*:

(a) *Let* ν *be the normalized Lebesgue-measure* $\nu(\cdot) := \lambda(\cdot \cap [-M,M])/\lambda([-M,M])$. *Then* $(X_n)_{n\in\mathbb{N}}$ *is an aperiodic positive* ν*-recurrent Harris chain with regeneration set* $[-M,M]$ *for* M *large enough. In particular, there exists a constant* $C \in (0,1)$ *such that for any Borel-measurable set* B *and* $x \in [-M,M]$

$$P(X_1 \in B \mid X_0 = x) \geqslant C\nu(B). \tag{2.5}$$

(b) $(X_n)_{n\in\mathbb{N}}$ *is geometric ergodic. In particular,* $(X_n)_{n\in\mathbb{N}}$ *has a unique stationary distribution and satisfies the strong mixing condition with geometric rate of convergence. The stationary* d*f is continuous and symmetric.*

(c) *Let $\bar{F}(x)=P(X>x)$, $x\geqslant 0$, be the right tail of the stationary* df *and conditions* (D.1)–(D.3) *are in addition fulfilled. Then*

$$\bar{F}(x)\sim cx^{-\kappa},\quad x\to\infty, \tag{2.6}$$

where

$$c=\frac{1}{2\kappa}\frac{E(|\alpha|X|+\sqrt{\beta+\lambda X^2}\varepsilon|^{\kappa}-|(\alpha+\sqrt{\lambda}\varepsilon)|X||^{\kappa})}{E(|\alpha+\sqrt{\lambda}\varepsilon|^{\kappa}\ln|\alpha+\sqrt{\lambda}\varepsilon|)} \tag{2.7}$$

and κ is given as the unique positive solution to

$$E(|\alpha+\sqrt{\lambda}\varepsilon|^{\kappa})=1. \tag{2.8}$$

Furthermore, the unique positive solution κ is less than 2 if and only if $\alpha^2+\lambda E(\varepsilon^2)>1$.

Remark 2.2. (a) Note that $E(|\alpha+\sqrt{\lambda}\varepsilon|^{\kappa})$ is a function of κ,α and λ. It can be shown that for fixed λ, the exponent κ is decreasing in $|\alpha|$. This means that the distribution of X gets heavier tails when $|\alpha|$ increases. In particular, the AR(1) process with ARCH(1) errors has for $\alpha\neq 0$ heavier tails than the ARCH(1) process (see also Table 3 in Borkovec and Klüppelberg, 1998).

(b) Theorem 2.1 is crucial for investigating the extremal behavior of $(X_n)_{n\in\mathbb{N}}$. The strong mixing property includes automatically that the sequence $(X_n)_{n\in\mathbb{N}}$ satisfies the conditions $D(u_n)$ and $\Delta(u_n)$. The condition $D(u_n)$ is a frequently used mixing condition due to Leadbetter et al. (1983) whereas the slightly stronger condition $\Delta(u_n)$ was introduced by Hsing (1984). Loosely speaking, $D(u_n)$ and $\Delta(u_n)$ give the "degree of independence" of extremes situated far apart from each other. This property together with (2.6) implies that the maximum of the process $(X_n)_{n\in\mathbb{N}}$ belongs to the domain of attraction of a Fréchet distribution. We will specify the normalizing constants of the maxima and the limit distribution in Section 3. □

In order to investigate the extremal behavior of $(X_n)_{n\in\mathbb{N}}$ and $(X_n^2)_{n\in\mathbb{N}}$ we define two auxiliary processes $(Y_n)_{n\in\mathbb{N}}$ and $(Z_n)_{n\in\mathbb{N}}$ as follows: let $(Y_n)_{n\in\mathbb{N}}$ be the process given by the random recurrence equation

$$Y_n=|\alpha Y_{n-1}+\sqrt{\beta+\lambda Y_{n-1}^2}\varepsilon_n|,\quad n\in\mathbb{N}, \tag{2.9}$$

where the notation is the same as in (2.1) and Y_0 equals $|X_0|$ a.s. Because of the symmetry of $(\varepsilon_n)_{n\in\mathbb{N}}$, the independence of ε_n and X_{n-1} in (2.1) and the homogeneous Markov structure of $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ it is readily seen that $(Y_n)_{n\in\mathbb{N}}\stackrel{\mathrm{d}}{=}(|X_n|)_{n\in\mathbb{N}}$.

Set now $(Z_n)_{n\in\mathbb{N}}=(\ln(Y_n^2))_{n\in\mathbb{N}}$. Since $(Y_n)_{n\in\mathbb{N}}$ follows (2.9) the process $(Z_n)_{n\in\mathbb{N}}$ satisfies the stochastic difference equation

$$Z_n=Z_{n-1}+\ln((\alpha+\sqrt{\beta\,\mathrm{e}^{-Z_{n-1}}+\lambda}\varepsilon_n)^2),\quad n\in\mathbb{N}, \tag{2.10}$$

where Z_0 equals $\ln(X_0^2)$ a.s. Note that $(Z_n)_{n\in\mathbb{N}}\stackrel{\mathrm{d}}{=}(\ln(X_n^2))_{n\in\mathbb{N}}$ and thus the process $(Z_n)_{n\in\mathbb{N}}$ is again regenerative and strongly mixing. Moreover, $(Z_n)_{n\in\mathbb{N}}$ does not depend on the sign of the parameter α since ε_n is symmetric. In the following, we assume therefore that $\alpha\geqslant 0$. We will see that $(Z_n)_{n\in\mathbb{N}}$ can be bounded by two random walks

$(S_n^{\mathrm{l},a})_{n\in\mathbb{N}}$ and $(S_n^{\mathrm{u},a})_{n\in\mathbb{N}}$ from below and above, respectively. This result together with $(Z_n)_{n\in\mathbb{N}} \stackrel{\mathrm{d}}{=} (\ln(X_n^2))_{n\in\mathbb{N}}$ appears to be the key to the description of the extremal behavior of $(X_n)_{n\in\mathbb{N}}$. Via results for $(Z_n)_{n\in\mathbb{N}}$, we prove for instance that the regenerative process $(X_n)_{n\in\mathbb{N}}$ has finite mean recurrence times which allow us to consider only the extremal behavior of the stationary process $(X_n)_{n\in\mathbb{N}}$. The process $(Z_n)_{n\in\mathbb{N}}$ is also important in the proof of Lemma 4.1.

For the construction of the two random walks $(S_n^{\mathrm{l},a})_{n\in\mathbb{N}}$ and $(S_n^{\mathrm{u},a})_{n\in\mathbb{N}}$ we need some more definitions. With the same notation as before, let

$$A_a(\varepsilon) := \left\{ \omega \,\middle|\, \frac{-\alpha}{\sqrt{\beta\,\mathrm{e}^{-a}+\lambda} - \sqrt{\beta}\,\mathrm{e}^{-a/2}} \leqslant \varepsilon(\omega) \leqslant \frac{-\alpha}{\sqrt{\beta\,\mathrm{e}^{-a}+\lambda} + \sqrt{\beta}\,\mathrm{e}^{-a/2}} \right\}, \tag{2.11}$$

$$p(a,\alpha,\beta,\lambda,\varepsilon) := \ln((\alpha + \sqrt{\beta\,\mathrm{e}^{-a}+\lambda}\,\varepsilon)^2),$$

$$q(a,\alpha,\beta,\lambda,\varepsilon) := \ln\left(1 - \frac{2\alpha\sqrt{\beta}\mathrm{e}^{-a/2}\varepsilon}{(\alpha+\sqrt{\beta\,\mathrm{e}^{-a}+\lambda}\,\varepsilon)^2} 1_{\{\varepsilon<0\}}\right), \tag{2.12}$$

$$r(a,\alpha,\beta,\lambda,\varepsilon) := \ln\left(1 - \frac{\beta\varepsilon^2\mathrm{e}^{-a}}{(\alpha+\sqrt{\beta\,\mathrm{e}^{-a}+\lambda}\,\varepsilon)^2} 1_{\{\varepsilon<0\}}\right).$$

Note that $q(a,\alpha,\beta,\lambda,\varepsilon), r(a,\alpha,\beta,\lambda,\varepsilon) \to 0$ a.s. for $a\to\infty$. Now define

$$S_n^{\mathrm{l},a} := \sum_{j=1}^n U_j^a \quad \text{and} \quad S_n^{\mathrm{u},a} := \sum_{j=1}^n V_j^a, \quad n\in\mathbb{N}, \tag{2.13}$$

where

$$U_j^a := -\infty\cdot 1_{A_a(\varepsilon)} + (p(a,\alpha,\beta,\lambda,\varepsilon_j) + r(a,\alpha,\beta,\lambda,\varepsilon_j))\cdot 1_{A_a(\varepsilon)^c\cap\{\varepsilon_j<0\}} + \ln(\alpha+\sqrt{\lambda}\varepsilon_j)^2\cdot 1_{\{\varepsilon_j\geqslant 0\}} \tag{2.14}$$

and

$$V_j^a := p(a,\alpha,\beta,\lambda,\varepsilon_j) + q(a,\alpha,\beta,\lambda,\varepsilon_j) \tag{2.15}$$

for some $a\geqslant 0$. The following lemma shows that the random walks defined in (2.13)–(2.15) are really upper and lower bounds for $(Z_n)_{n\in\mathbb{N}}$ above a high level.

Lemma 2.3. *Let a be large enough, $N_a := \inf\{j\geqslant 1 \mid Z_j\leqslant a\}$ and $Z_0 > a$. Then*

$$Z_0 + S_k^{\mathrm{l},a} \leqslant Z_k \leqslant Z_0 + S_k^{\mathrm{u},a} \quad \textit{for any } k\leqslant N_a \textit{ a.s.} \tag{2.16}$$

Proof. We prove only the lower bound. The proof of the upper bound is similar but easier. Let $x\geqslant a$ be arbitrary. If $\varepsilon\geqslant 0$ it is obvious that

$$(\alpha+\sqrt{\beta\mathrm{e}^{-x}+\lambda}\varepsilon)^2 \geqslant (\alpha+\sqrt{\lambda}\varepsilon)^2. \tag{2.17}$$

Consider now $\varepsilon < 0$, then

$$\begin{aligned}
&(\alpha+\sqrt{\beta e^{-x}+\lambda}\varepsilon)^2-(\alpha+\sqrt{\beta e^{-a}+\lambda}\varepsilon)^2\\
&\quad=2\alpha(-\varepsilon)(\sqrt{\beta e^{-a}+\lambda}-\sqrt{\beta e^{-x}+\lambda})-\beta(e^{-a}-e^{-x})\varepsilon^2\\
&\quad\geqslant -\beta e^{-a}\varepsilon^2. \qquad (2.18)
\end{aligned}$$

Note that we have a non-trivial lower bound for $(\alpha+\sqrt{\beta e^{-x}+\lambda}\varepsilon)^2$ if and only if

$$(\alpha+\sqrt{\beta e^{-a}+\lambda}\varepsilon)^2-\beta e^{-a}\varepsilon^2>0. \qquad (2.19)$$

It is straightforward that (2.19) is equivalent to

$$\varepsilon>\frac{-\alpha}{\sqrt{\beta e^{-a}+\lambda}+\sqrt{\beta}e^{-a/2}} \quad \text{or} \quad \varepsilon<\frac{-\alpha}{\sqrt{\beta e^{-a}+\lambda}-\sqrt{\beta}e^{-a/2}}. \qquad (2.20)$$

From (2.17), (2.18) and (2.20), we obtain

$$\begin{aligned}
&(\alpha+\sqrt{\beta e^{-x}+\lambda}\varepsilon(\omega))^2\\
&\quad\geqslant\begin{cases}(\alpha+\sqrt{\lambda}\varepsilon(\omega))^2, & \omega\in\{\varepsilon\geqslant 0\},\\ (\alpha+\sqrt{\beta e^{-a}+\lambda}\varepsilon(\omega))^2-\beta e^{-a}\varepsilon(\omega)^2, & \omega\in A_a(\varepsilon)^c\cap\{\varepsilon<0\},\\ 0, & \omega\in A_a(\varepsilon).\end{cases} \qquad (2.21)
\end{aligned}$$

Now take logarithms and use the additive structure (2.10) of $(Z_n)_{n\in\mathbb{N}}$. □

Remark 2.4. (a) If a is large enough then $S_n^{u,a}$ and $S_n^{l,a}$ are random walks with negative drift.

Proof. Note that

$$\begin{aligned}
E(V_1^a)&=E(p(a,\alpha,\beta,\lambda,\varepsilon_1)+q(a,\alpha,\beta,\lambda,\varepsilon_1))\\
&=E(\ln((\alpha+\sqrt{\beta e^{-a}+\lambda}\varepsilon_1)^2+2\alpha\sqrt{\beta}e^{-a/2}(-\varepsilon_1)1_{\{\varepsilon_1<0\}}))\\
&\to E(\ln(\alpha+\sqrt{\lambda}\varepsilon_1)^2)<0 \quad \text{as } a\to\infty,
\end{aligned}$$

where we used the dominated convergence theorem and (2.2) in the last step. Hence for a large enough the statement follows. □

(b) Let $(S_n)_{n\in\mathbb{N}} := (\sum_{j=1}^n \ln((\alpha+\sqrt{\lambda}\varepsilon_j)^2))_{n\in\mathbb{N}}$. For $a\uparrow\infty$ we have

$$S_k^{l,a}\xrightarrow{P}S_k \quad \text{and} \quad S_k^{u,a}\xrightarrow{a.s.}S_k \qquad (2.22)$$

for any $k\in\mathbb{N}$, i.e. both random walks converge at least in probability to the same random walk. Furthermore,

$$\sup_{k\geqslant 1}S_k^{l,a}\xrightarrow{d}\sup_{k\geqslant 1}S_k \quad \text{and} \quad \sup_{k\geqslant 1}S_k^{u,a}\xrightarrow{a.s.}\sup_{k\geqslant 1}S_k. \qquad (2.23)$$

Proof. The a.s. convergence of $(S_n^{u,a})_{n\in\mathbb{N}}$ and $\sup_{k\geqslant 1}S_k^{u,a}$ is straightforward since p,q and r converge a.s. Consider therefore the lower random walk $(S_n^{l,a})_{n\in\mathbb{N}}$. Note that for $a\uparrow\infty$

$$P(A_a(\varepsilon))\to 0$$

and hence

$$1_{A_a(\varepsilon)^c\cap\{\varepsilon<0\}} \xrightarrow{\mathrm{P}} 1_{\{\varepsilon<0\}} \quad \text{and} \quad 1_{A_a(\varepsilon)\cap\{\varepsilon<0\}} \xrightarrow{\mathrm{P}} 0. \tag{2.24}$$

Moreover,

$$p(a,\alpha,\beta,\lambda,\varepsilon_1)+r(a,\alpha,\beta,\lambda,\varepsilon_1) \xrightarrow{\text{a.s.}} \ln((\alpha+\sqrt{\lambda}\varepsilon_1)^2), \tag{2.25}$$

and therefore (2.22) holds. Finally, we note that

$$\begin{aligned} E\max(0,U_1^a) &= E\max(0,(p(a,\alpha,\beta,\lambda,\varepsilon_1)+r(a,\alpha,\beta,\lambda,\varepsilon_1))1_{A_a(\varepsilon)^c\cap\{\varepsilon_1<0\}}) \\ &\quad + E\max(0,\ln(\alpha+\sqrt{\lambda}\varepsilon_1)^2 1_{\{\varepsilon_1\geqslant 0\}}) \\ &\to E\max(0,\ln(\alpha+\sqrt{\lambda}\varepsilon_1)^2) \quad \text{as } a\to\infty, \end{aligned} \tag{2.26}$$

where we used (2.24), (2.25) and the dominated convergence theorem. By Borovkov (1976, Theorem 22, p. 53), (2.22) and (2.26) we derive that

$$\sup_{k\geqslant 1} S_k^{\mathrm{l},a} \xrightarrow{\mathrm{d}} \sup_{k\geqslant 1} S_k. \qquad \square$$

Lemma 2.3 characterizes the behavior of the process $(Z_n)_{n\in\mathbb{N}}$ above a high threshold a and hence also the behavior of $(X_n^2)_{n\in\mathbb{N}}$. This is the key to what follows: the process $(S_n)_{n\in\mathbb{N}}$ will determine completely the extremal behavior of (X_n^2). Recall from Theorem 2.1 that $(X_n)_{n\in\mathbb{N}}$ is Harris recurrent with regeneration set $[-\mathrm{e}^{a/2},\mathrm{e}^{a/2}]$ for a large enough. Thus there exists in particular a renewal point process $T_0,T_1,T_2,\ldots$ which describes the regenerative structure of $(X_n)_{n\in\mathbb{N}}$.

Corollary 2.5. *The renewal point process $(T_n)_{n\in\mathbb{N}_0}$ which describes the regenerative structure of $(X_n)_{n\in\mathbb{N}}$ is aperiodic and has finite mean recurrence times $C_0=T_0$ and $C_1=T_1-T_0$.*

Proof. Since $(Z_n)_{n\in\mathbb{N}} \stackrel{\mathrm{d}}{=} (\ln(X_n^2))_{n\in\mathbb{N}}$ it is sufficient to investigate the regenerative structure of $(Z_n)_{n\in\mathbb{N}}$. Note that $(Z_n)_{n\in\mathbb{N}}$ is Harris recurrent with regeneration set $(-\infty,a]$. The renewal process can be constructed in the following way (see e.g. Asmussen (1987), Section VI.3 for some background on regenerative Markov processes):

Define

$$\tau_1 := \inf\{k\geqslant 1 \mid Z_k\leqslant a\} = N_a$$

and $\tau_{i+1} := \inf\{k>\tau_i \mid Z_k\leqslant a\}$ for $i=1,2,3,\ldots$. Since, above level a, $(Z_n)_{n\in\mathbb{N}}$ is dominated by the random walk with negative drift $(S_n^{\mathrm{u},a})_{n\in\mathbb{N}}$ and

$$\sup_{x\in(-\infty,a]} E(\max(0,Z_1)\mid Z_0=x)<\infty, \tag{2.27}$$

it follows that $\tau_1,\tau_2,\tau_3,\ldots$ are well defined and have finite expectations. Now let $M_1 := \inf\{i\geqslant 1 \mid I_{\tau_i}=1\}$ and $M_{j+1} := \inf\{i>M_j \mid I_{\tau_i}=1\}$ for $j=1,2,3,\ldots$ with $P(I_1=1)=1-P(I_1=0)=C$ and independent of $(X_n)_{n\in\mathbb{N}}$ where C is the constant in (2.5). Note that

$$P(M_j-M_{j-1}=i)=C(1-C)^{i-1} \quad \text{for } i,j=1,2,\ldots \text{ and } M_0=0. \tag{2.28}$$

From Asmussen (1987, p. 151) and (2.5), the renewal process $(T_n)_{n\geqslant 0}$ is now given by

$$T_n := \tau_{M_{n+1}} + 1, \quad n \geqslant 0,$$

and hence, by (2.28)

$$E(C_0) = E(T_0) \leqslant E(\tau_{M_1+1}) \leqslant \text{const.}\, E(M_1 + 1) < \infty.$$

Similar calculation shows that $E(C_1) < \infty$ as well. Since the transition density of $(Z_n)_{n\in\mathbb{N}}$ is positive and continuous it follows finally that C_1 is aperiodic. $\square$

As a consequence of Corollary 2.5 we may suppose in the following that the process $(X_n)_{n\in\mathbb{N}}$ is stationary. It can be shown by a coupling argument that for any probability measure μ and any sequence $(u_n)_{n\in\mathbb{N}}$

$$\left| P^{\mu}\left(\max_{1\leqslant k\leqslant n} X_k \leqslant u_n \right) - P^{\pi}\left(\max_{1\leqslant k\leqslant n} X_k \leqslant u_n \right) \right| \to 0 \quad \text{as } n \to \infty,$$

where P^{μ} denotes the probability law for $(X_n)_{n\in\mathbb{N}}$ when X_0 starts with distribution μ and π is the stationary distribution. For the coupling argument one needs explicitly that the process $(X_n)_{n\in\mathbb{N}}$ is regenerative and that the embedded renewal process is aperiodic and has finite mean recurrence time. We refer to Lindvall (1992, Chapters II and III) for further details.

3. Extremal behavior of the AR(1) process with ARCH(1) errors

In this section we present the main results concerning the extremal behavior of the AR(1) process with ARCH(1) errors and the related squared process. Let $(\hat{X}_n)_{n\in\mathbb{N}}$ be the associated independent process of $(X_n)_{n\in\mathbb{N}}$, i.e. $\hat{X}_1, \hat{X}_2, \ldots$ are i.i.d. random variables with the stationary distribution function of $(X_n)_{n\in\mathbb{N}}$. From (2.6) and classical extreme value theory we obtain

$$\lim_{n\to\infty} P\left(n^{-1/\kappa} \max_{1\leqslant k\leqslant n} \hat{X}_k \leqslant x \right) = \exp(-cx^{-\kappa}), \quad x \geqslant 0. \tag{3.1}$$

Hence the maximum of the associated independent process $(\hat{X}_n)_{n\in\mathbb{N}}$ belongs to the domain of attraction of a Fréchet distribution. In the dependent case we prove a similar result. The limit distribution is still a Fréchet distribution but a constant θ occurs in the exponent. θ is called the *extremal index* of the process $(X_n)_{n\in\mathbb{N}}$ and is a measure of local dependence amongst the exceedances over a high threshold by the process $(X_n)_{n\in\mathbb{N}}$. It has a natural interpretation as the reciprocal of the mean cluster size. In order to describe the extremes in more detail, we also consider the point process $(N_n)_{n\in\mathbb{N}}$ of exceedances of an appropriately high chosen threshold u_n given by

$$N_n(\cdot) := \#\{k/n \in \cdot \mid X_k > u_n,\ k \in \{1, \ldots, n\}\} \tag{3.2}$$

and show that this point process converges to a compound Poisson process N. We derive the intensity and the distribution of the jumps which we denote by $(\pi_k)_{k\in\mathbb{N}}$. Note that in the extreme value theory for strong mixing processes the jumps equal the lengths of clusters of exceedances. For further background we refer to Leadbetter et al. (1983), Rootzén (1988) or Embrechts et al. (1997, Section 8.1). For the ARCH(1) process it was convenient to investigate first the squared process (see de Haan et al., 1989; Hooghiemstra and Meester, 1995). This is not possible for our model since we have a different structure due to the autoregressive part of $(X_n)_{n\in\mathbb{N}}$. Nevertheless, only for the squared process $(X_n^2)_{n\in\mathbb{N}}$ a comparison with results in the ARCH(1) case is feasible. The following theorem collects our results.

Theorem 3.1. (a) *Suppose $(X_n)_{n\in\mathbb{N}}$ is given by Eq.* (2.1) *with $(\varepsilon_n)_{n\in\mathbb{N}}$ satisfying the general conditions* (2.3) *and* (D.1)–(D.3) *with parameters α and λ satisfying* (2.2) *and $X_0 \sim \mu$. Then*

$$\lim_{n\to\infty} P^{\mu}\left(n^{-1/\kappa}\max_{1\leqslant j\leqslant n} X_j \leqslant x\right) = \exp(-c\theta x^{-\kappa}), \quad x\geqslant 0, \tag{3.3}$$

where P^{μ} denotes the law for $(X_n)_{n\in\mathbb{N}}$ when X_0 starts with the distribution μ, κ solves the equation $E(|\alpha+\sqrt{\lambda}\varepsilon|^{\kappa}) = 1$, c is defined by (2.7) *and*

$$\theta = \kappa\int_1^{\infty} P\left(\sup_{k\geqslant 1}\prod_{i=1}^{k}(\alpha+\sqrt{\lambda}\varepsilon_i)\leqslant y^{-1}\right) y^{-\kappa-1}\,\mathrm{d}y.$$

For $x\in\mathbb{R}$, let N_n be the point process of exceedances of the threshold $u_n = n^{1/\kappa}x$ by $X_1,\ldots,X_n$ given by (3.2). *Then*

$$N_n \xrightarrow{\mathrm{d}} N, \quad n\to\infty,$$

where N is a compound Poisson process with intensity $c\theta x^{-\kappa}$ and cluster probabilities

$$\pi_k = \frac{\theta_k-\theta_{k+1}}{\theta}, \quad k\in\mathbb{N}, \tag{3.4}$$

where

$$\theta_k = \kappa\int_1^{\infty} P\left(\#\left\{j\geqslant 1 \,\middle|\, \prod_{i=1}^{j}(\alpha+\sqrt{\lambda}\varepsilon_i) > y^{-1}\right\} = k-1\right) y^{-\kappa-1}\,\mathrm{d}y, \quad k\in\mathbb{N}.$$

In particular, $\theta_1=\theta$.

(b) *Let $(X_n)_{n\in\mathbb{N}}$ be the* AR(1) *process with* ARCH(1) *errors in* (a) and $(X_n^2)_{n\in\mathbb{N}}$ *the squared process. Then*

$$\lim_{n\to\infty} P^{\mu}\left(n^{-2/\kappa}\max_{1\leqslant j\leqslant n} X_j^2 \leqslant x\right) = \exp(-2c\theta^{(2)}x^{-\kappa/2}), \quad x\geqslant 0, \tag{3.5}$$

where κ, c are the same constants as in (a) *and*

$$\theta^{(2)} = \frac{\kappa}{2}\int_1^{\infty} P\left(\sup_{k\geqslant 1}\prod_{i=1}^{k}(\alpha+\sqrt{\lambda}\varepsilon_i)^2\leqslant y^{-1}\right) y^{-(\kappa/2)-1}\,\mathrm{d}y.$$

For $x \in \mathbb{R}$, let $N_n^{(2)}$ be the point process of exceedances of the threshold $u_n = n^{2/\kappa}x$ by $X_1^2, \dots, X_n^2$. Then

$$N_n^{(2)} \xrightarrow{\mathrm{d}} N^{(2)}, \quad n \to \infty,$$

where $N^{(2)}$ is a compound Poisson process with intensity $2c\theta^{(2)}x^{-\kappa/2}$ and cluster probabilities

$$\pi_k^{(2)} = \frac{\theta_k^{(2)} - \theta_{k+1}^{(2)}}{\theta^{(2)}}, \quad k \in \mathbb{N}, \tag{3.6}$$

where

$$\theta_k^{(2)} = \frac{\kappa}{2} \int_1^\infty P\left(\# \left\{ j \geqslant 1 \,\middle|\, \prod_{i=1}^{j} (\alpha + \sqrt{\lambda}\varepsilon_i)^2 > y^{-1} \right\} = k - 1 \right) y^{-(\kappa/2)-1}\, \mathrm{d}y,$$

$$k \in \mathbb{N}.$$

In particular, $\theta_1^{(2)} = \theta^{(2)}$.

Remark 3.2. (a) Theorem 3.1 is a generalization of the result of de Haan et al. (1989) in the ARCH(1) case (i.e. $\alpha = 0$). They use a different approach which does not extend to the general case because of the autoregressive part of $(X_n)_{n\in\mathbb{N}}$.

(b) Note that for the squared process one can describe the extremal index and the cluster probabilities by the random walk $(S_n)_{n\in\mathbb{N}}$, namely

$$\theta_k^{(2)} = \frac{\kappa}{2} \int_0^\infty P(\#\{j \geqslant 1 \mid S_j > -x\} = k - 1) \mathrm{e}^{-(\kappa/2)x}\, \mathrm{d}x, \quad k \in \mathbb{N}.$$

The description of the extremal behavior of $(X_n^2)_{n\in\mathbb{N}}$ by the random walk $(S_n)_{n\in\mathbb{N}}$ is to be expected since by Lemma 2.3 and Remark 2.4 the process $(Z_n)_{n\in\mathbb{N}}$ ($\stackrel{\mathrm{d}}{=} (\ln(X_n^2))_{n\in\mathbb{N}}$) behaves above a high threshold asymptotically like $(S_n)_{n\in\mathbb{N}}$. Unfortunately, this link fails for $(X_n)_{n\in\mathbb{N}}$. Another possibility for proving statement (b) is to follow the work of Hooghiemstra and Meester (1995) using the regenerative structure of $(Z_n)_{n\in\mathbb{N}}$, Lemma 2.3, Corollary 2.5 and Remark 2.4(b).

(c) Analogous to de Haan et al. (1989) we may construct "estimators" for the extremal indices $\theta^{(2)}$ and $\theta_k^{(2)}$ of $(X_n^2)_{n\in\mathbb{N}}$, respectively, by

$$\hat{\theta}^{(2)} = \frac{1}{N} \sum_{i=1}^{N} 1_{\{\sup_{1 \leqslant j \leqslant m} S_j^{(i)} \leqslant -E_{\kappa/2}^{(i)}\}}$$

and

$$\hat{\theta}_k^{(2)} = \frac{1}{N} \sum_{i=1}^{N} 1_{\{\Sigma_{j=1}^{m} 1_{\{S_j^{(i)} > -E_{\kappa/2}^{(i)}\}} = k-1\}} \quad \text{for } k \in \mathbb{N},$$

where N denotes the number of simulated sample paths of $(S_n)_{n\in\mathbb{N}}$, $E_{\kappa/2}^{(i)}$ are i.i.d. exponential random variables with intensity $\kappa/2$ and m is chosen large enough. These estimators can be studied as in the case $\alpha = 0$ and $\varepsilon \sim \mathrm{N}(0, 1)$ in de Haan et al. (1989). In particular,

$$\frac{\theta^{(2)} - \hat{\theta}^{(2)}}{(\theta^{(2)}(1 - \theta^{(2)})/N)^{1/2}}$$

Table 1
Numerical tail index κ and "estimated" extremal index θ and cluster probabilities $(\pi_k)_{1\leqslant k\leqslant 6}$ of $(X_n)_{n\in\mathbb{N}}$ dependent on α and λ in the case $\varepsilon \sim \mathrm{N}(0,1)$. We chose $N = m = 2000$. Note that the extremal index for $\alpha > 0$ is much larger than for $\alpha < 0$

α	λ	κ	θ	π_1	π_2	π_3	π_4	π_5	π_6
0	0.2	12.85	0.974	0.973	0.027	0.000	0.000	0.000	0.000
0	0.6	3.82	0.781	0.799	0.147	0.036	0.012	0.005	0.001
0	1	1.99	0.549	0.607	0.188	0.107	0.036	0.034	0.017
−0.4	0.2	8.12	0.962	0.962	0.037	0.001	0.000	0.000	0.000
0.4	0.2	8.12	0.853	0.867	0.103	0.026	0.002	0.002	0.000
−0.4	0.6	2.87	0.715	0.747	0.168	0.048	0.026	0.006	0.002
0.4	0.6	2.87	0.624	0.676	0.182	0.066	0.040	0.019	0.012
−0.4	1	1.61	0.497	0.540	0.210	0.115	0.075	0.040	0.004
0.4	1	1.61	0.445	0.533	0.185	0.080	0.109	0.032	0.017
−0.8	0.2	3.00	0.572	0.626	0.185	0.111	0.026	0.033	0.001
0.8	0.2	3.00	0.386	0.470	0.172	0.148	0.062	0.068	0.006
−0.8	0.6	1.37	0.414	0.520	0.159	0.134	0.072	0.043	0.016
0.8	0.6	1.37	0.314	0.443	0.156	0.110	0.087	0.073	0.041
−0.8	1	0.85	0.273	0.429	0.137	0.126	0.106	0.016	0.012
0.8	1	0.85	0.224	0.346	0.132	0.114	0.129	0.045	0.004

is approximately $\mathrm{N}(0,1)$ distributed. Because of Remark 3.2(b) this approach is not possible for $(X_n)_{n\in\mathbb{N}}$. We choose as "estimators" for θ and θ_k for $(X_n)_{n\in\mathbb{N}}$

$$\hat{\theta} = \frac{1}{N}\sum_{i=1}^{N} 1_{\{\sup_{1\leqslant j\leqslant m}\Pi_{l=1}^{j}(\alpha+\sqrt{\lambda}\varepsilon_l)\leqslant 1/P_\kappa^{(i)}\}} \tag{3.7}$$

and

$$\hat{\theta}_k = \frac{1}{N}\sum_{i=1}^{N} 1_{\left\{\Sigma_{j=1}^{m} 1_{\{\Pi_{l=1}^{j}(\alpha+\sqrt{\lambda}\varepsilon_l)\,>1/P_\kappa^{(i)}\}} = k-1\right\}} \quad \text{for } k\in\mathbb{N}, \tag{3.8}$$

where N denotes the number of simulated paths of $(\prod_{l=1}^{n}(\alpha+\sqrt{\lambda}\varepsilon_l))_{n\in\mathbb{N}}$, $P_\kappa^{(i)}$ are i.i.d. Pareto-distributed random variables with intensity κ, i.e. with distribution function $G(x) = 1 - x^{-\kappa}$, $x\geqslant 1$, and m is large enough. These are suggestive estimators since $\prod_{l=1}^{n}(\alpha+\sqrt{\lambda}\varepsilon_l) \to 0$ a.s. as $n\to\infty$ because of assumption (2.2).

(d) Note that the extremal index θ of $(X_n)_{n\in\mathbb{N}}$ is not symmetric in the parameter α (see also Table 1). This observation is intuitively obvious since for $\alpha > 0$ the clustering is stronger by the autoregressive part than for $\alpha < 0$.

4. The proof of Theorem 3.1

The proof of Theorem 3.1 will be an application of results in Perfekt (1994). In order to apply these results we need to check some assumptions. The next lemma provides a technical property for the squared AR(1) process with ARCH(1) errors $(X_n^2)_{n\in\mathbb{N}}$. It is the most restrictive assumption in Perfekt (1994).

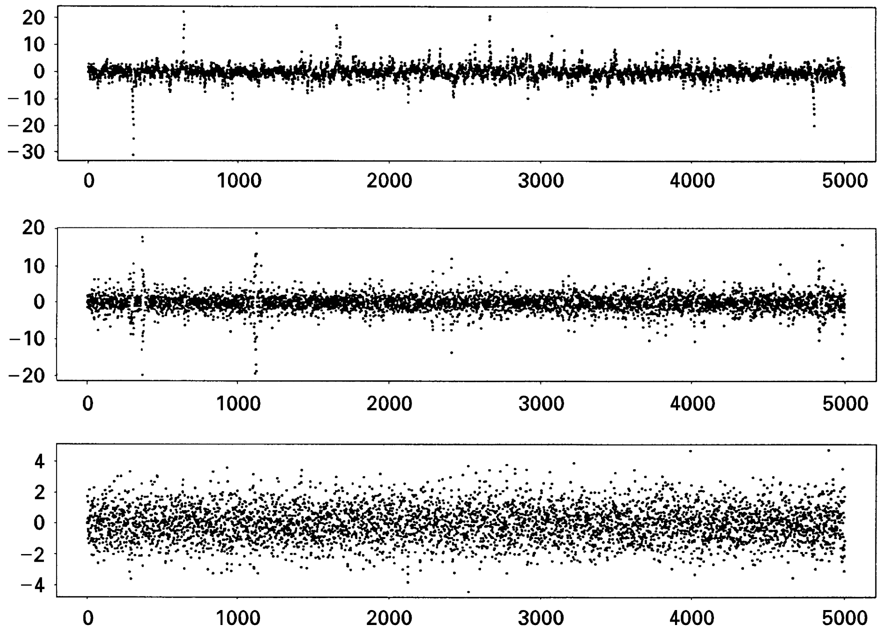

Fig. 1. Simulated sample path of $(X_n)_{n\in\mathbb{N}}$ with parameters $\alpha=0.8$, $\beta=1$, $\lambda=0.2$ (top), with $\alpha=-0.8$, $\beta=1$, $\lambda=0.2$ (middle) and with $\alpha=0$, $\beta=1$, $\lambda=0.2$ (bottom) in the case $\varepsilon\sim \mathrm{N}(0,1)$. All simulations are based on the same simulated noise sequence $(\varepsilon_n)_{n\in\mathbb{N}}$. The pictures demonstrate the tendency of clustering and confirm our comments in Remarks 2.2(a) and 3.2(d).

Lemma 4.1. *Let $(p_n)_{n\in\mathbb{N}}$ be an increasing sequence such that*

$$\frac{p_n}{n}\to 0 \quad \textit{and} \quad \frac{n\gamma(\sqrt{p_n})}{p_n}\to 0 \quad \textit{as } n\to\infty, \tag{4.1}$$

where γ is the mixing function of $(X_n)_{n\in\mathbb{N}}$, i.e. for any $m\in\mathbb{N}$

$$\gamma(m)=\sup\{|P(A\cap B)-P(A)P(B)|\colon A\in\sigma(X_j,1\leqslant j\leqslant k),$$
$$B\in\sigma(X_j,\ j\geqslant k+m),\ k\in\mathbb{N}\}.$$

Then for $u_n=n^{2/\kappa}x$

$$\lim_{p\to\infty}\limsup_{n\to\infty} P\left(\max_{p\leqslant j\leqslant p_n} X_j^2>u_n\,|\,X_0^2>u_n\right)=0. \tag{4.2}$$

Remark 4.2. (a) The strong mixing condition is a property of the underlying σ-field of a process. Hence γ is also the mixing function of $(X_n^2)_{n\in\mathbb{N}}$ and $(Z_n)_{n\in\mathbb{N}}$ and we may work in all these cases with the same sequence $(p_n)_{n\in\mathbb{N}}$. Note that because of Theorem 2.1(b) there exist constants $\eta\in(0,1)$ and $c>0$ such that $\gamma(m)\leqslant c\rho^m$ for any $m\in\mathbb{N}$ (Fig. 1).

(b) In the case of a strong mixing process, conditions (4.1) are sufficient to guarantee that $(p_n)_{n\in\mathbb{N}}$ is a $\Delta(u_n)$-separating sequence. This is a straightforward consequence of the fact that $\sigma(\{X_j\leqslant u_n\},1\leqslant j\leqslant k)\subseteq\sigma(X_j,1\leqslant j\leqslant k)$, $\sigma(\{X_j\leqslant u_n\}, j\geqslant l_n+k)\subseteq\sigma(X_j,\ j\geqslant l_n+k)$ and choosing additionally $l_n=\sqrt{p_n}$. The notion of a $\Delta(u_n)$-

separating sequence was first introduced by O'Brien (1987) and describes somehow the interval length needed to accomplish asymptotic independence of extremal events over a high level u_n in separate intervals. For a definition see also Perfekt (1994). Note that $(p_n)_{n\in\mathbb{N}}$ is in the case of a strong mixing process independent of $(u_n)_{n\in\mathbb{N}}$.

Proof. Note that

$$\begin{aligned} P&\left(\max_{p\leqslant j\leqslant p_n} X_j^2 > u_n \mid X_0^2 > u_n\right)\\ &= P\left(N_a < p,\ \max_{p\leqslant j\leqslant p_n} X_j^2 > u_n \mid X_0^2 > u_n\right)\\ &\quad + P\left(p\leqslant N_a < p_n,\ \max_{p\leqslant j\leqslant p_n} X_j^2 > u_n \mid X_0^2 > u_n\right)\\ &\quad + P\left(N_a\geqslant p_n,\ \max_{p\leqslant j\leqslant p_n} X_j^2 > u_n \mid X_0^2 > u_n\right)\\ &=: I_1 + I_2 + I_3, \end{aligned} \tag{4.3}$$

where $N_a = \inf\{j\geqslant 1 \mid Z_j\leqslant a\} = \inf\{j\geqslant 1 \mid X_j^2\leqslant \mathrm{e}^a\}$ as in Lemma 2.3. In order to get upper bounds of I_1, I_2 and I_3 we show first that there exist constants $C>0$ and $N\in\mathbb{N}$ such that for any $n>N$, $x\in[\mathrm{e}^{-n},\mathrm{e}^a]$ and $k\in\mathbb{N}$

$$nP(X_k^2 > u_n \mid X_0^2 = x)\leqslant C. \tag{4.4}$$

Assume that (4.4) does not hold. Choose C, $N>0$ arbitrary and $\eta>0$ small. Because of the continuity of the transition probability (i.e. equicontinuity on compact sets), there exist $n>N$, $x\in[\mathrm{e}^{-n},\mathrm{e}^a]$, $k\in\mathbb{N}$ and $\delta=\delta(\eta)>0$ such that for any $y\in(x-\delta,x+\delta)\cap[\mathrm{e}^{-n},\mathrm{e}^a]$

$$nP(X_k^2 > u_n \mid X_0^2 = y) > C-\eta. \tag{4.5}$$

Let F_{X^2} denote the stationary df of $(X_n^2)_{n\in\mathbb{N}}$. By Theorem 2.1 we have that

$$\lim_{n\to\infty} n\bar{F}_{X^2}(u_n) = 2cx^{-\kappa/2}, \tag{4.6}$$

where c is given by the formula in (2.7) and κ is the solution of (2.8). Furthermore, by (4.5) we have

$$\begin{aligned} n\bar{F}_{X^2}(u_n) &= \int_{(-\infty,\infty)} nP(X_k^2 > u_n \mid X_0^2 = y)\,\mathrm{d}F_{X^2}(y)\\ &\geqslant \int_{(x-\delta,x+\delta)\cap[\mathrm{e}^{-n},\mathrm{e}^a]} nP(X_k^2 > u_n \mid X_0^2 = y)\,\mathrm{d}F_{X^2}(y)\\ &> (C-\eta)P(X_0^2\in(x-\delta,x+\delta)\cap[\mathrm{e}^{-n},\mathrm{e}^a])\\ &\geqslant (C-\eta)D, \end{aligned}$$

where $D := \inf_{z\in[0,\mathrm{e}^a]}(F_{X^2}(z+\delta)-F_{X^2}(z)) > 0$ because F_{X^2} is continuous. Since $C>0$ is arbitrary this is a contradiction to (4.6).

Now we estimate (4.3).

$$
\begin{aligned}
I_1 &\leqslant \sum_{l=1}^{p-1} P\left(N_a = l, \max_{p\leqslant j\leqslant p_n} X_j^2 > u_n \,|\, X_0^2 > u_n\right) \\
&\leqslant \sum_{l=1}^{p-1}\sum_{j=l+1}^{p_n} P(N_a = l, X_j^2 > u_n \,|\, X_0^2 > u_n) \\
&= \sum_{l=1}^{p-1}\sum_{j=l+1}^{p_n} E(1_{\{N_a=l\}}\, P(X_j^2 > u_n \,|\, X_l^2)|X_0^2 > u_n) \\
&= \sum_{l=1}^{p-1}\sum_{j=l+1}^{p_n} E(1_{\{N_a=l\}}\, 1_{\{X_l^2 \geqslant \mathrm{e}^{-n}\}} P(X_j^2 > u_n \,|\, X_l^2)|X_0^2 > u_n) \\
&\quad + \sum_{l=1}^{p-1}\sum_{j=l+1}^{p_n} E(1_{\{N_a=l\}} 1_{\{X_l^2 < \mathrm{e}^{-n}\}} P(X_j^2 > u_n \,|\, X_l^2)|X_0^2 > u_n) \\
&=: J_1 + J_2. \qquad (4.7)
\end{aligned}
$$

Furthermore, by (4.4),

$$
\begin{aligned}
J_1 &\leqslant \sum_{l=1}^{p-1}\sum_{j=l+1}^{p_n} \frac{1}{n} E(1_{\{N_a=l\}}\, 1_{\{X_l^2 \geqslant \mathrm{e}^{-n}\}} n P(X_j^2 > u_n \,|\, X_l^2)\,|\, X_0^2 > u_n) \\
&\leqslant \sum_{l=1}^{p-1}\sum_{j=l+1}^{p_n} \frac{C}{n} E(1_{\{N_a=l\}}\, 1_{\{X_l^2 \geqslant \mathrm{e}^{-n}\}}\,|\, X_0^2 > u_n) \\
&\leqslant \sum_{j=1}^{p_n} \frac{C}{n} P(N_a < p \,|\, X_0^2 > u_n) \\
&\leqslant C\frac{p_n}{n} \\
&\to 0 \quad \text{as } n \to \infty, \qquad (4.8)
\end{aligned}
$$

since $p_n = \mathrm{o}(n)$. Similarly, with $B_l := \{X_1^2 > \mathrm{e}^a, \ldots, X_{l-1}^2 > \mathrm{e}^a\}$ for any $l = 2, 3, 4, \ldots$ and $B_1 = \Omega$, we obtain

$$
\begin{aligned}
J_2 &\leqslant \sum_{l=1}^{p-1}\sum_{j=l+1}^{p_n} E(1_{\{N_a=l\}}\, 1_{\{X_l^2 < \mathrm{e}^{-n}\}}\,|\, X_0^2 > u_n) \\
&= \sum_{l=1}^{p-1}\sum_{j=l+1}^{p_n} E(1_{B_l}\, P(X_l^2 < \mathrm{e}^{-n} \,|\, X_{l-1}^2)\,|\, X_0^2 > u_n) \\
&= \sum_{l=1}^{p-1}\sum_{j=l+1}^{p_n} E(1_{B_l}\, P((\alpha X_{l-1} + \sqrt{\beta + \lambda X_{l-1}^2}\,\varepsilon_l)^2 < \mathrm{e}^{-n} \,|\, X_{l-1}^2)\,|\, X_0^2 > u_n)
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{l=1}^{p-1} \sum_{j=l+1}^{p_n} E\left(1_{B_l\cap\{X_{l-1}>0\}}\, P\left(\frac{-\mathrm{e}^{-n/2}/X_{l-1}-\alpha}{\sqrt{\beta/X_{l-1}^2+\lambda}} < \varepsilon_l \right.\right.\\
&\qquad \left.\left. < \frac{\mathrm{e}^{-n/2}/X_{l-1}-\alpha}{\sqrt{\beta/X_{l-1}^2+\lambda}}\right) \middle|\, X_0^2 > u_n\right)\\
&\qquad + \sum_{l=1}^{p-1} \sum_{j=l+1}^{p_n} E\left(1_{B_l\cap\{X_{l-1}<0\}} P\left(\frac{\mathrm{e}^{-n/2}/X_{l-1}+\alpha}{\sqrt{\beta/X_{l-1}^2+\lambda}} < \varepsilon_l \right.\right.\\
&\qquad \left.\left. < \frac{-\mathrm{e}^{-n/2}/X_{l-1}+\alpha}{\sqrt{\beta/X_{l-1}^2+\lambda}}\right) \middle|\, X_0^2 > u_n\right)\\
&= \sum_{l=1}^{p-1} \sum_{j=l+1}^{p_n} E\left(1_{B_l\cap\{X_{l-1}>0\}} P\left(\frac{-\mathrm{e}^{-n/2-a/2}-\alpha}{\sqrt{\lambda}} < \varepsilon_l \right.\right.\\
&\qquad \left.\left. < \frac{\mathrm{e}^{-n/2-a/2}-\alpha}{\sqrt{\lambda}}\right) \middle|\, X_0^2 > u_n\right)\\
&\qquad + \sum_{l=1}^{p-1} \sum_{j=l+1}^{p_n} E\left(1_{B_l\cap\{X_{l-1}<0\}} P\left(\frac{-\mathrm{e}^{-n/2-a/2}+\alpha}{\sqrt{\lambda}} < \varepsilon_l \right.\right.\\
&\qquad \left.\left. < \frac{\mathrm{e}^{-n/2-a/2}+\alpha}{\sqrt{\lambda}}\right) \middle|\, X_0^2 > u_n\right)\\
&\leqslant 2\,\text{const.}\; p\, p_n \mathrm{e}^{-n/2-a/2}\\
&\to 0 \quad \text{as } n \to \infty
\end{aligned}
$$

and therefore with (4.8) $I_1 \to 0$ as $n \to \infty$.

Now we estimate $\limsup_{n\to\infty} I_3$. Note first that by the Markov inequality

$$
\begin{aligned}
&P\left(\max_{p\leqslant j\leqslant p_n} S_j^{\mathrm{u},\,a} > -z\right)\\
&\quad\leqslant \sum_{j=p}^{p_n} P\left(\mathrm{e}^{(\kappa/4)S_j^{\mathrm{u},\,a}} > \mathrm{e}^{-(\kappa/4)z}\right)\\
&\quad= \sum_{j=p}^{p_n} P\left(\prod_{m=1}^{j} ((\alpha+\sqrt{\beta e^{-a}+\lambda}\,\varepsilon_m)^2 - 2\alpha\sqrt{\beta}\mathrm{e}^{-a/2}\varepsilon_m 1_{\{\varepsilon_m<0\}})^{\kappa/4} > \mathrm{e}^{-(\kappa/4)z}\right)\\
&\quad\leqslant \mathrm{e}^{(\kappa/4)z} \sum_{j=p}^{p_n} E\left(((\alpha+\sqrt{\beta\,\mathrm{e}^{-a}+\lambda}\,\varepsilon_1)^2 - 2\alpha\sqrt{\beta}\mathrm{e}^{-a/2}\varepsilon_1 1_{\{\varepsilon_1<0\}})^{\kappa/4}\right)^j\\
&\quad\leqslant \mathrm{e}^{(\kappa/4)z} \sum_{j=p}^{p_n} \eta^j, \qquad (4.9)
\end{aligned}
$$

where $\eta < 1$ such that $E(((\alpha+\sqrt{\beta\,\mathrm{e}^{-a}+\lambda}\,\varepsilon_1)^2 - 2\alpha\sqrt{\beta}\mathrm{e}^{-a/2}\varepsilon_1 1_{\{\varepsilon_1<0\}})^{\kappa/4}) \leqslant \eta$ for a large enough. This is possible because of (2.2) which implies that $E(|\alpha+\sqrt{\lambda}\,\varepsilon_1|^u) < 1$

for all $u \in (0,\kappa)$ and the fact that

$$E\left(((\alpha+\sqrt{\beta\exp(-a)+\lambda}\,\varepsilon_1)^2-2\alpha\sqrt{\beta}\mathrm{e}^{-a/2}\varepsilon_1 1_{\{\varepsilon_1<0\}})^{\kappa/4}\right)$$
$$\to E(|\alpha+\sqrt{\lambda}\,\varepsilon_1|^{\kappa/2}),\quad a\to\infty$$

by the dominated convergence theorem. Thus from Theorem 2.1, Lemma 2.3, (4.9) and a large enough,

$$\begin{aligned}
\limsup_{n\to\infty} I_3 &\leqslant \limsup_{n\to\infty} P\left(N_a \geqslant p_n,\ \max_{p\leqslant j\leqslant p_n} Z_0+S_j^{\mathrm{u},a} > \ln u_n \,|\, Z_0>\ln u_n\right)\\
&\leqslant \limsup_{n\to\infty} P\left(\max_{p\leqslant j\leqslant p_n} Z_0+S_j^{\mathrm{u},a} > \ln u_n \,|\, Z_0>\ln u_n\right)\\
&= \limsup_{n\to\infty} \int_0^\infty P\left(\max_{p\leqslant j\leqslant p_n} S_j^{\mathrm{u},a} > -z\right)\frac{\kappa}{2}\mathrm{e}^{-(\kappa/2)z}\,\mathrm{d}z\\
&\leqslant 2\sum_{j=p}^\infty \eta^j = 2\frac{\eta^{p-1}}{1-\eta}.
\end{aligned}\tag{4.10}$$

Finally, note that

$$\begin{aligned}
I_2 &\leqslant P\left(p\leqslant N_a < p_n,\ \max_{N_a<j\leqslant p_n} X_j^2 > u_n \,|\, X_0^2>u_n\right)\\
&\quad + P\left(p\leqslant N_a < p_n,\ \max_{p\leqslant j\leqslant N_a} X_j^2 > u_n \,|\, X_0^2>u_n\right)\\
&=: K_1+K_2.
\end{aligned}$$

Similarly as for I_1 and I_3, respectively, we derive that

$$\limsup_{n\to\infty} K_1 = 0 \quad\text{and}\quad \limsup_{n\to\infty} K_2 = 2\frac{\eta^{p-1}}{1-\eta}.$$

Now plugging all together and letting $p\to\infty$ the statement follows. □

Corollary 4.3. *Let $(p_n)_{n\in\mathbb{N}}$ be the same sequence as in Lemma 4.1. Then $(p_n)_{n\in\mathbb{N}}$ is also a $\Delta(u_n)$-separating sequence for $(X_n)_{n\in\mathbb{N}}$, where $u_n = n^{1/\kappa}x$ and $x\in\mathbb{R}$ arbitrary and*

$$\lim_{p\to\infty}\limsup_{n\to\infty} P\left(\max_{p\leqslant j\leqslant p_n} X_j > u_n \,|\, X_0>u_n\right) = 0. \tag{4.11}$$

Proof. Because of Remark 4.2(a) and (b), it is straightforward that $(p_n)_{n\in\mathbb{N}}$ is a $\Delta(u_n)$-separating sequence for $(X_n)_{n\in\mathbb{N}}$. Note furthermore that

$$\begin{aligned}
P\left(\max_{p\leqslant j\leqslant p_n} X_j^2 > u_n^2 \,|\, X_0^2>u_n^2\right) &= \frac{P(\max_{p\leqslant j\leqslant p_n} X_j^2>u_n^2,\ X_0^2>u_n^2)}{P(X_0^2>u_n^2)}\\
&\geqslant \frac{P(\max_{p\leqslant j\leqslant p_n} X_j>u_n,\ X_0>u_n)}{P(X_0>u_n)+P(X_0<-u_n)}\\
&= \frac{1}{2}P\left(\max_{p\leqslant j\leqslant p_n} X_j > u_n \,|\, X_0>u_n\right)
\end{aligned}$$

and hence the statement follows using Lemma 4.1. □

Now we are finally able to prove Theorem 3.1.

Proof of Theorem 3.1. The proof is an application of a result of Perfekt (1994, p. 543) which is basically an extension of Theorem 3.2 in the same paper. We prove only statement (a), statement (b) follows along the same lines using Theorem 3.2 in Perfekt (1994). As stated already we may assume w.l.o.g. that $(X_n)_{n\in\mathbb{N}}$ is stationary. Let $x\in\mathbb{R}$ be arbitrary. Note that

$$\lim_{u\to\infty}\frac{P(X_0>u+\frac{1}{\kappa}ux)}{P(X_0>u)}=\begin{cases}\infty, & 1+\frac{1}{\kappa}x\leqslant 0\\ (1+\frac{1}{\kappa}x)^{-\kappa}, & 1+\frac{1}{\kappa}x>0\end{cases}$$

and

$$\lim_{u\to\infty}P\left(\frac{X_1}{u}\leqslant x\,|\,X_0=u\right)=P(\alpha+\sqrt{\lambda}\,\varepsilon\leqslant x).$$

By Corollary 4.3 and the strong mixing property of $(X_n)_{n\in\mathbb{N}}$ all assumptions of Perfekt's result are fulfilled and we have that the extremal index θ is given by

$$\theta=\int_1^\infty P\left(\#\left\{j\geqslant 1\,\middle|\,\left(\prod_{i=1}^{j}(\alpha+\sqrt{\lambda}\,\varepsilon_i)\right)Y_0>1\right\}=0\,\middle|\,Y_0=y\right)\kappa y^{-\kappa-1}\,\mathrm{d}y$$
$$=\int_1^\infty P\left(\max_{j\geqslant 1}\left(\prod_{i=1}^{j}(\alpha+\sqrt{\lambda}\,\varepsilon_i)\leqslant y^{-1}\right)\right)\kappa y^{-\kappa-1}\,\mathrm{d}y.$$

The cluster probabilities can be determined in the same way and hence the statement follows. □

Acknowledgements

I wish to express my gratitude to Claudia Klüppelberg for her constant support and advice. I am also grateful to Alex McNeil for many discussions and an Splus program which estimates the extremal index by the blocks method.

References

Asmussen, S., 1987. Applied Probability and Queues. Wiley, Chichester.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedastic. J. Econometrics 31, 307–327.

Bollerslev, T., Chou, R.Y., Kroner, K.F., 1992. ARCH modelling in finance: a review of the theory and empirical evidence. J. Econometrics 52, 5–59.

Borkovec, M., Klüppelberg, C., 1998. The tail of the stationary distribution of an autoregressive process with ARCH(1) errors. Preprint.

Borovkov, A.A., 1976. Stochastic Processes in Queueing Theory. Springer, New York.

Diebolt, J., Guégan, D., 1990. Probabilistic properties of the general nonlinear Markovian process of order one and applications to time series modelling. Rapp. Tech. 125, L.S.T.A., Paris 6.

Embrechts, P., Goldie, C.M., 1994. Perpetuities and random equations. In: Mandl, P., Huskova, M. (Eds.) Asymptotic Statistics. Proceedings of the 5th Prague Symposium, September 4–9, 75–86, Physica-Verlag, Heidelberg.

Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events for Insurance and Finance. Springer, Heidelberg.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. Econometrica 50, 987–1007.

Haan, L. de, Resnick, S.I., Rootzén, H., Vries, C.G. de, 1989. Extremal behavior of solutions to a stochastic difference equation with applications to ARCH processes. Stochastic Process. Appl. 32, 213–224.

Hooghiemstra, G., Meester, L.E., 1995. Computing the extremal index of special Markov chains and queues. Stochastic Process. Appl. 65, 171–185.

Hsing, T., 1984. Point processes associated with extreme value theory. Dissertation, Technical Report 83, Center for Stochastic Processes, Univ. North Carolina.

Kesten, H., 1973. Random difference equations and renewal theory for products of random matrices. Acta Math. 131, 207–248.

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

Lindvall, T., 1992. Lectures on the Coupling Method. Wiley, New York.

Maercker, G., 1997. Statistical Inference in Conditional Heteroskedastic Autoregressive Models. Shaker, Aachen (Dissertation, University of Braunschweig).

O'Brien, G.L., 1987. Extreme values for stationary and Markov sequences. Ann. Probab. 15, 281–291.

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

Priestley, M.B., 1988. Non-linear and Non-stationary Time Series. Academic Press, New York.

Rootzén, H., 1988. Maxima and exceedances of stationary Markov chains. Adv. Appl. Probab. 20, 371–390.

Shephard, N., 1996. Statistical aspects of ARCH and stochastic volatility. In: Cox, D.R., Hinkley, D.V., Barndorff-Nielsen, O.E. (Eds.), Likelihood, Time Series with Econometric and other Applications. Chapman & Hall, London.

Taylor, S.J., 1986. Modelling Financial Time Series. Wiley, Chichester.

Tong, H., 1990. Non-linear Time Series – A Dynamical System Approach. Oxford University Press, Oxford.

Vervaat, W., 1979. On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables. Adv. Appl. Probab. 11, 750–783.

Weiss, A.A., 1984. ARMA models with ARCH errors. J. Time Ser. Anal. 3, 129–143.