ISSN 0974-3235
ProbStat Forum, Volume 03, July 2010, Pages 91-107

Asymptotic and pre-asymptotic tail behavior of a power max-autoregressive model

Marta Ferreira
University of Minho, Department of Mathematics, Campus de Gualtar
4710-057 Braga - Portugal

Luísa Canto e Castro
Department of Statistics and Operation Research, Faculty of Sciences
University of Lisbon, Campo Grande, 1749-016 Lisboa, Portugal
e-mail: *msferreira@math.uminho.pt*, *luisa.loura@fc.ul.pt*

Paper received on 19 March 2010; revised, 30 May 2010; accepted, 10 June 2010.

Abstract. Max-autoregressive models for time series data are useful when we want to make inference about rare events, mainly in areas like hydrology, geophysics and finance. Here we present a power max-autoregressive (pARMAX) process, $\{X_i\}_{i\in\mathbb{Z}}$, defined in such a way that the asymptotic tail dependence coefficient of Ledford and Tawn, computed for observations lag m apart (η_m), exhibits a power decay with m for larger values of c, the main parameter of the process, namely, $\eta_m = c^m$, $c \in (1/2, 1)$. We also look at the threshold-dependent form of the extremal index, which is an important functional when extending discussions of extreme values from independent and identically distributed (i.i.d.) sequences to stationary ones. We state an approach for this functional as well as its connection with the coefficient η for the pARMAX process.

Mathematics Subject Classification. 60G70, 60J20.

Keywords. Markov Chains, Extreme value theory, Dependence conditions, Auto-asymptotic-tail-dependence function, Tail index, Extremal index.

1 Introduction

Extreme Value Theory (EVT) became widely used by many researchers in applied sciences when faced with modeling high values of certain phenomena, e.g., ocean wave modeling, wind engineering, thermodynamics of earthquakes or risk assessment on financial markets. The first results were developed considering independent observations but, more recently, models for extreme values have been constructed under the more realistic assumption of temporal dependence. Among these models, stationary Markov chains are very interesting, specially because they may have a somewhat simple treatment in what concerns extremal properties. The max-autoregressive moving average processes MARMA (Davis and Resnick 1989), and also the particular case MAR(1) or ARMAX, given by,

$$X_i = k\, X_{i-1} \vee W_i, \tag{1.1}$$

with $0 < k < 1$ and $\{W_i\}_{i\in\mathbb{Z}}$ i.i.d. (Alpuim 1989; Canto e Castro 1992; Ancona-Navarrete and Tawn 2000; Lebedev 2008) are some examples. Here we present the power max-autoregressive process (in short, pARMAX), defined as

$$X_i = X_{i-1}^c \vee Z_i\ ,\ 0 < c < 1,\ i \in \mathbb{Z},$$

with $\{Z_i\}_{i\in\mathbb{Z}}$ i.i.d.. Heavy tailed pARMAX and AR(1) have quite similar sample paths (see Figure 1), indicating that pARMAX can be thought as an alternative model, even because its finite-dimensional distributions can easily be written explicitly. Observe that the ARMAX process obtained from the logarithm of heavy tail pARMAX is Gumbel tailed, a less interesting and applicable case within these processes in what concerns extreme value analysis. The

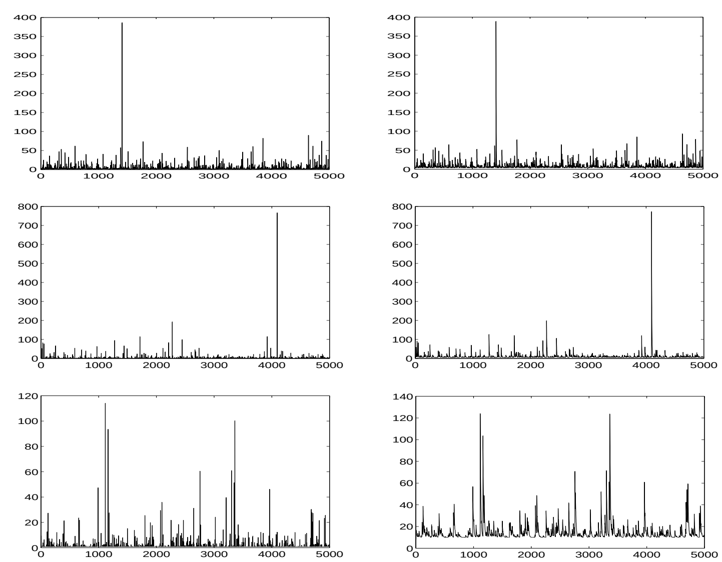

Figure 1: 5000 realizations of pARMAX, $X_i = X_{i-1}^c \vee Z_i$, on the left, and of AR(1), $X_i = c\ X_{i-1} + Z_i$, on the right, with $c = 0.8$ and with marginals Pareto(0.7)

pARMAX process has very easily derived extremal features. Actually, these are all straightforward for the more general and further applicable pRARMAX process (Ferreira and Canto e Castro, 2010), which includes random coefficients U_i (with support in $[0, 1]$),

$$X_i = U_i\ X_{i-1}^c \vee Z_i\ ,\ 0 < c < 1,\ i \in \mathbb{Z}, \tag{1.2}$$

($U_i = 1$ leads to pARMAX). We shall see that the pARMAX power parameter c is related with Ledford and Tawn coefficient η. Hence, based on the estimation of η, we will present estimators for parameter c and prove consistency and asymptotical normality using Drees conditions (Drees 2003) stated below.

1.1 Tail dependence: Ledford and Tawn approach

In the classical multivariate EVT we cannot distinguish asymptotically between exact independence of the components of a random vector and a moderate dependence which vanishes as the observations go more extreme. If we take, for instance, a bivariate normal random vector with correlation $\rho < 1$, we verify that it has the same limit distribution for the standardized maxima of the components independently of the value of ρ. In order to overcome this problem, Ledford and Tawn (1996) proposed a model, where the penultimate tail dependence is characterized by the so-called coefficient of tail dependence, $\eta \in (0, 1]$, which measures the dependence between the marginal tails.

Consider $\{(X_i, Y_i)\}_{i\geq 1}$, a sequence of i.i.d. random pairs with common d.f. F and marginal d.f.'s F_1 and F_2. Considering $U = 1 - F_1(X)$ and $V = 1 - F_2(Y)$, the basic Ledford and Tawn model assumption can be formulated as (Draisma *et al.*, 2004),

$$P\Big(\frac{U}{t} < x, \frac{V}{t} < y)|U < t, V < t\Big) = \frac{P(U < tx, V < ty)}{P(U < t, V < t)} \underset{t\downarrow 0}{\longrightarrow} h(x, y) \tag{1.3}$$

uniformly on $\{(x, y)|\max(x, y) = 1\}$ for some non-degenerate function h. It is assumed that the function $t \mapsto P(U < t, V < t)$ is regularly varying at 0 with index $1/\eta$ and that $l := \lim_{t\downarrow 0} P(U < t|V < t)$ exists. Note that we can also write

$$P(U < t|V < t) \sim t^{1/\eta - 1} L(t), \text{as } t \downarrow 0. \tag{1.4}$$

We have $l = 0$ if $\eta < 1$ and $l > 0$ if the marginals are asymptotically dependent providing that $L(t) \not\to 0$ as $t \to \infty$. Roughly speaking, four distinct cases can be considered: $\eta = 1$ in case of asymptotic dependence, $\eta \in (1/2, 1)$ and $\eta \in (1/2, 1)$ if it exists, respectively, positive and negative dependence, both vanishing asymptotically, and $\eta = 1/2$ in case of independence. The function h is homogeneous of order $1/\eta$, since $h(tx, ty) = t^{1/\eta} h(x, y)$.

However, dependence also occurs within a time series framework and η can be seen as a measure of tail dependence in time. More precisely, we derive η_m from (1.3) when applied to random pairs (X_1, X_m) separated in time by a lag m. Based on Ledford and Tawn (2003), we shall use the more readily interpreted quantity,

$$\Lambda_m = 2\eta_m - 1,$$

here denoted ATDF (*auto-asymptotic-tail-dependence function*), since $\Lambda_m = 1$ corresponds to asymptotic dependence, $0 < \Lambda_m < 1$ to positive extremal association, $\Lambda_m = 0$ to (almost) independence and $\Lambda_m < 0$ to negative extremal dependence (we will take $\eta_1 \equiv \eta$). The ATDF function Λ_m "provides a measure of serial dependence between extreme values lag m apart and may be interpreted in a manner that is broadly similar to the ACF" (Ledford and Tawn 2003).

The pARMAX process has an easily derived method for model identification through the ATDF, as we shall see.

1.2 Drees' class of tail index estimators

Consider $\{X_i\}_i$ a stationary sequence. Using a weighted approximation of the tail empirical quantile function (q.f.), $Q_n(t) := X_{n-[k_n t]:n}$, where $(k_n)_{n\geq 1}$ is an intermediate sequence, i.e., $k_n \to \infty$ and $k_n/n \to 0$, as $n \to \infty$, Drees (2003) stated its asymptotic behavior under the following conditions:

- a *β-mixing* dependence structure:

$$\beta(l) := \sup_{p\in\mathbb{N}} E\Big(\sup_{B\in\mathcal{F}(X_{p+l+1},\dots)} |P(B|\mathcal{F}(X_1,\dots,X_p)) - P(B)|\Big) \underset{l\to\infty}{\longrightarrow} 0,$$

  with $\mathcal{F}(.)$ denoting the $\sigma-$field generated by the indicated random variables. More precisely, it is assumed that there exists a sequence l_n, $n \in \mathbb{N}$, such that

$$\lim_{n\to\infty} \frac{\beta(l_n)}{l_n} n + l_n k_n^{-1/2} \log^2 k_n = 0 \tag{1.5}$$

- a regularity condition for the joint tail of (X_1, X_{1+m}):

$$\lim_{n\to\infty} \frac{n}{k_n} P\Big(X_1 > F^{-1}\big(1 - \frac{k_n}{n}x\big), X_{1+m} > F^{-1}\big(1 - \frac{k_n}{n}y\big)\Big) = c_m(x,y), \tag{1.6}$$

  for all $m \in \mathbb{N}$, $0 < x, y \leq 1 + \epsilon$ and F^{-1} denoting the inverse function of F.

- a uniform bound on the probability that both X_1 and X_{1+m} belong to an extremal interval:

$$\frac{n}{k_n} P\big(X_1 \in I_n(x,y), X_{1+m} \in I_n(x,y)\big) \leq (y-x)\Big(\tilde{\rho}(m) + D_1\Big(\frac{k_n}{n}\Big)^{\alpha}\Big), \tag{1.7}$$

  with $0 < \alpha \leq 1$, for some constant $D_1 \geq 0$, a sequence $\tilde{\rho}(m)$, $m \in \mathbb{N}$, satisfying $\sum_{m=1}^{\infty} \tilde{\rho}(m) < \infty$ for all $m \in \mathbb{N}$, $0 < x, y \leq 1 + \epsilon$ and the extremal interval, $I_n(x,y) =]F^{-1}(1 - yk_n/n), F^{-1}(1 - xk_n/n)]$.

- for the sake of simplicity,
$$F^{-1}(1-t) = dt^{-\gamma}(1+r(t))\ ,\ \text{with } |r(t)| < \Phi(t)\,. \tag{1.8}$$
for some constant $d > 0$ and a τ-varying function Φ at 0 for some $\tau > 0$, or $\tau = 0$ and Φ nondecreasing with $\lim_{t\downarrow 0}\Phi(t) = 0$

- it is assumed that $(k_n)_{n\geq 1}$ is an intermediate sequence such that
$$\lim_{n\to\infty} k_n^{1/2}\Phi(k_n/n) = 0 \tag{1.9}$$

Theorem 1.1 (Drees (2003)) *Under the conditions (1.5)-(1.9) with $l_n = o((n/k_n)^\alpha)$, there exist versions of the tail empirical q.f. Q_n and a centered Gaussian process e with covariance function $\tilde{c}$ given by*
$$\tilde{c}(x,y) := x \wedge y + \sum_{m=1}^{\infty}(c_m(x,y) + c_m(y,x)) \in \mathbb{R}\,, \tag{1.10}$$
such that
$$\sup_{t\in(0,1]} t^{\gamma+1/2}(1+|\log t|)^{-1/2}\left|k_n^{1/2}\Big(\frac{Q_n(t)}{F^{-1}(1-k_n/n)} - t^{-\gamma}\Big) - \gamma t^{-(\gamma+1)}e(t)\right| \to 0$$
in probability.

Drees (2003) observe that almost every estimator $\hat{\gamma}_n$ of the tail index parameter γ that are based only on the k_n+1 largest order statistics, can be represented as a smooth functional T (verifying some regularity conditions) applied to the tail empirical q.f.. Hill estimator, maximum likelihood estimator, moments estimator, Pickands' estimator and probability weighted moments estimator, are some examples. We shall denote these as *Drees' class of tail index estimators*. Theorem 2.2 in Drees (2002) establishes the asymptotic normality of these estimators. More precisely,
$$k_n^{1/2}(\hat{\gamma}_n - \gamma) \underset{n\to\infty}{\longrightarrow} \mathcal{N}(0, \sigma^2_{T,\gamma})$$
weakly with
$$\sigma^2_{T,\gamma} = \gamma^2 \int_{(0,1]}\int_{(0,1]} (st)^{-(\gamma+1)}\tilde{c}(s,t)\nu_{T,\gamma}(ds)\nu_{T,\gamma}(dt), \tag{1.11}$$
where $\tilde{c}$ is the function defined in (1.10). For instance, considering the Hill estimator which is defined as
$$\hat{\gamma}_n^H = \frac{1}{k_n}\sum_{i=1}^{k_n}\log\frac{X_{n-i+1:n}}{X_{n-k_n:n}}, \tag{1.12}$$
it can be proved that, in a generalized Pareto model, it has signed measure given by
$$\nu_{H,\gamma}(dt) = t^\gamma dt - \delta_1(dt), \tag{1.13}$$
where δ_1 is the Dirac measure with mass 1 at 1.

1.3 The pre-asymptotic extremal index

The extremal index, θ ($\theta \in [0,1]$), measures the tendency of extremes to occur in clusters. A cluster of high levels exceedances is defined to be a set of observations that exceed a high threshold u_n within a block of length r_n ($r_n = o(n)$), given that there is at least one exceedance in that block. For a stationary sequence $\{Y_i\}_i$ under some suitable dependence condition,

$$\theta = \lim_{n\to\infty} P(Y_i \le u_n, 2 \le i \le r_n | Y_1 > u_n)\,, \tag{1.14}$$

where $(u_n)_{n\ge 1}$ is a sequence of normalized levels, i.e., $nP(Y_1 > u_n) = O(1)$ (O'Brien 1987). ($\theta = 0$ corresponds to "pathological" cases which we do not intend to consider here). This parameter plays a very important role when estimating extremal properties of $\{Y_i\}_i$ with marginals in the domain of attraction of an extreme value distribution. More precisely, for certain levels $\{u_n\}_{n\ge 1}$, we have,

$$\lim_{n\to\infty} P(\textstyle\bigvee_{i=1}^{n} Y_i \le u_n) = \lim_{n\to\infty} \big(P(Y_1 \le u_n)\big)^{n\theta}\,. \tag{1.15}$$

Though processes with i.i.d. margins have $\theta = 1$, the converse is false. This can be evidenced through the well-known autoregressive Gaussian processes (regarded as strongly dependent by other measures but with $\theta = 1$). The pARMAX process is also an example with unit extremal index as we shall see.

An unit extremal index means that asymptotic extreme events tend to occur singly and makes the result (1.15) no informative about the dependence of the process. However, it can be possible to observe clustering of exceedances for levels of practical interest.

Based on O'Brien characterization (1.14) by taking a large finite n and $u_n \equiv u$, Bortot and Tawn (1998) stated the functional

$$\theta(u, r_{[u]}) = 1 - P(Y_2 > u | Y_1 > u) - \sum_{j=3}^{r_{[u]}} P\big(\textstyle\bigvee_{i=2}^{j-1} Y_i \le u, Y_j > u | Y_1 > u\big), \tag{1.16}$$

for which we have $\theta = \lim_{u\to\infty} \theta(u, r_{[u]})$. In order to relate this functional with the tail dependence parameter η, the authors have stated the following dependence condition:

Definition 1.1 *For a stationary process $\{X_i\}_{i\in\mathbb{Z}}$ and sequences, $(u_n)_{n\ge 1}$ and $(r_n)_{n\ge 1}$, such that, $nP(X_1 > u_n) = O(1)$ and $r_n = o(n)$, respectively, condition $\Delta^{(2)}(u_n, r_n)$ holds if*

$$\sum_{j=3}^{r_n} \frac{P(X_1 > u_n, \max(X_2, ..., X_{j-1}) \le u_n, X_j > u_n)}{P(X_1 > u_n, X_2 > u_n)} \to 0, \text{ as } n \to \infty. \tag{1.17}$$

For $r_n = 2$, which can be chosen for instance in the case of independent series, condition $\Delta^{(2)}(u_n, r_n)$ is automatically satisfied (Bortot and Tawn 1998).
Observe that, under condition $\Delta^{(2)}(u_n, r_n)$, the expression in (1.16) can be approximated as follows:

$$\theta(u) \equiv \theta(u, r_{\lfloor u \rfloor}) \sim 1 - P(Y_2 > u | Y_1 > u) \tag{1.18}$$

and a relation with η can be stated by (1.4). We will see that if we replace θ by the pre-asymptotic extremal index $\theta(u)$ in the result (1.15), we get some improvement on estimations based on this latter, like high quantiles and return periods. This will be done in the last section.

In this paper, we start by presenting the pARMAX process $\{X_i\}_{i\in\mathbb{Z}}$ and state stationarity, extremal behavior and local dependence structure. We shall focus on heavy tailed pARMAX, the most interesting case for extremal inference. We will characterize the joint limiting d.f. of the normalized first passage time of threshold u, $T = \inf\{n \in \mathbb{N} : X_n \geq u\}$, and the corresponding excess, $R_T = X_T - u$, as $u \to \infty$. In Section 4, we derive a procedure for model identification of pARMAX through the ATDF. We will see that, for $m \in \mathbb{N}$, this function presents a power decay with m if $c \in (1/2, 1)$. We also give estimators for parameter c which work for both pARMAX and pRARMAX. Consistency and asymptotic normality will be proved as well. Section 5 is devoted to the pre-asymptotic extremal index, $\theta(u)$. According to Bortot and Tawn (1998), we establish a connection between this functional and η in the pARMAX model and derive a better estimate for the return levels.

2 Stationarity and extremal properties of pARMAX

Consider $\{Z_i\}_{i\in\mathbb{Z}}$ a sequence of i.i.d. copies of a r.v., Z, having real nonnegative support and marginal d.f. F_Z. A sequence $\{X_i\}_{i\in\mathbb{Z}}$ is said to be a pARMAX process if,

$$X_i = X_{i-1}^c \vee Z_i \; , \; 0 < c < 1, \;\; i = 0, \pm 1, \pm 2, ... \tag{2.1}$$

with X_i independent of Z_j, for all integer $i < j$. The sequence $\{Z_i\}_{i\in\mathbb{Z}}$ is also known as the innovations sequence of the process.

Observe that the pARMAX process can be obtained from ARMAX given in (1.1) by taking exponentials, and hence, the stationarity of the first is straightforward from the latter. More precisely, the ARMAX process is stationary if and only if $\sum_{j=0}^{\infty} -\log F_W(c^j x)$ is finite for some $x \geq 0$, where F_W is the marginal d.f. of innovations $\{W_i\}_{i\in\mathbb{Z}}$ (Alpuim 1989). Therefore, we have,

$$\textstyle\sum_{j=0}^{\infty} -\log F_Z(x^{1/c^j}) < \infty \, , \;\; \text{for some } x \geq 0, \tag{2.2}$$

as a necessary and sufficient condition for the stationarity of the process pARMAX. Note that thus the pARMAX innovations, $\{Z_i\}_{i\in\mathbb{Z}}$, have support in $[1, \infty[$.

Denoting by $K_i(\cdot)$ the marginal d.f. of X_i for any integer i, the recurrence (2.1) and the independence assumptions lead to:

$$K_i(x) = P(X_i \leq x) = P(X_{i-1} \leq x^{1/c}, Z_i \leq x) = K_{i-1}(x^{1/c})F_Z(x)\,.$$

So, any d.f. that satisfies the stationarity equation,

$$K(x) = K(x^{1/c})F_Z(x)\,, \tag{2.3}$$

is a stationary distribution of the pARMAX process $\{X_i\}_{i\in\mathbb{Z}}$. Here is an example.

Example 2.1 *Consider* $\{Z_i\}_{i\in\mathbb{Z}}$ *with common d.f.,*

$$F_Z(x) = c\mathbf{1}_{\{x=1\}} + \frac{1-x^{-1/\gamma}}{1-x^{-1/(c\gamma)}}\mathbf{1}_{\{x>1\}}, \tag{2.4}$$

where $\mathbf{1}_{\{\cdot\}}$ *is the indicator function. Hence,* $K(x) = \big(1-x^{-1/\gamma}\big)\mathbf{1}_{\{x\geq 1\}}$*, satisfies (2.3), being, therefore, a stationary distribution for* X_i*.*

The margins of a stationary pARMAX process $\{X_i\}_{i\in\mathbb{Z}}$ are in the same max-domain of attraction of the innovations, with the same tail index value, since

$$P\big(\textstyle\bigvee_{i=1}^n X_i \leq x\big) = P\big(X_1 \vee \bigvee_{i=2}^n Z_i \leq x\big) = K(x)F_Z^{n-1}(x)\,. \tag{2.5}$$

The β-*mixing* condition is valid for ARMAX (Canto e Castro 1992) and hence for pARMAX, given the mentioned relationship between the two processes. Heavy tailed pARMAX has unit extremal index (as is the case of an ARMAX process in the Gumbel max-domain of attraction; Alpuim 1989).

3 Heavy tailed pARMAX processes

As already mentioned, we will consider stationary pARMAX processes with heavy tail innovations, a more general case than the standard Fréchet considered for the study of MARMA models (Davis and Resnick 1989; Zhang and Smith 2001; Lebedev 2008).

Indeed, using a convenient representation of a Fréchet(γ) domain of attraction (for some $\gamma > 0$), it is easy to prove that condition (2.2) is valid therein. More precisely, the assumption is equivalent to a regular variation at infinity with index $-1/\gamma$, i.e.,

$$1 - F_Z(x) = x^{-1/\gamma}L_Z(x)\,, \tag{3.1}$$

where $L_Z(\cdot)$ is a slowly varying function at ∞ (Embrechts at al. 1997, Theorem 3.3.7 on p.131 and p.152), and hence we also have,

$$1 - K(x) = x^{-1/\gamma}L_K(x) \tag{3.2}$$

for some slowly varying function $L_K(\cdot)$.

Now we will show that first passage times over high levels are also easily derived for pARMAX processes. Let $T = T_u = \inf\{n \in \mathbb{N} : X_n \geq u\}$ be the first passage of X_n over threshold u and $R_T = X_T - u$ the respective overshoot. We prove that T and R_T, properly normalized, are asymptotically independent and we calculate their joint limiting distribution as the threshold $u \to \infty$.

Theorem 3.1 *Let $\{X_i\}_{i\in\mathbb{Z}}$ be a pARMAX process. Let E and V be independent r.v.'s with distribution functions $\big(1 - e^{-x}\big)I_{(0,\infty)}$ and $\big(1-(y+1)^{-1/\gamma}\big)I_{(0,\infty)}$ respectively. Then,*

$$\big((1-F_Z(u))T, u^{-1}R_T\big) \longrightarrow (E, V) \text{ in distribution, when } u \to \infty.$$

Proof Set $d = 1 - F_Z(u)$. For positive s, v and ω, such that $s < v$,

$$\begin{aligned} P\big(s \leq dT \leq \omega, R_T > vu\big) &= \sum_j P(R_T > vu, T = j) \\ &= \sum_j K(u)F_Z^{j-2}(u)(1 - F_Z((v+1)u)\,, \end{aligned}$$

where the summation is over $[s/d] \leq j \leq [\omega/d]$ ($[\cdot]$ denotes the integer part). By (3.1), as $u \to \infty$, we have that, $(1-F_Z((v+1)u))/(1-F_Z(u)) \sim (v+1)^{-1/\gamma}$ and hence,

$$(1-F_Z(u))\textstyle\sum_j F_Z^{j-2}(u) = (1-F_Z(u))\,\frac{1-F_Z(u)^{[\frac{\omega}{d}]-[\frac{s}{d}]+1}}{1-F_Z(u)}\,F_Z(u)^{[\frac{s}{d}]-2} \sim e^{-s} - e^{-\omega}\,.$$

Since $K(u) \to 1$ as $u \to \infty$, the assertion follows. □

Remark 3.1 *A sequence of normalized levels $(u_n)_{n\geq 1}$ of the marginal d.f., K, is also a sequence of normalized levels of the innovations d.f., $F_Z(.)$, i.e., by (2.3) we have*

$$n(1-F_Z(u_n)) = \tfrac{n(1-K(u_n))-n(1-K(u_n^{1/c}))}{K(u_n^{1/c})}\,. \tag{3.3}$$

If $n(1-K(u_n)) \to \tau > 0$, as $n \to \infty$, and hence $u_n \sim n^\gamma L_K(u_n)^\gamma/\tau^\gamma$, then

$$1 - K(u_n^{1/c}) \sim (\tau/n)^{1/c}\mathcal{L}(a_{\tau/n}^{1/c})\,, \tag{3.4}$$

$$\mathcal{L}(a_t^{1/c}) = \big[L_{K^{-1}}(t)\big]^{-1/(\gamma c)}L_K\big(a_t^{1/c}\big) \sim \big[L_{K^{-1}}(t)\big]^{-1/(\gamma c)}L_K\big((t^{-\gamma}L_{K^{-1}}(t))^{1/c}\big) \tag{3.5}$$

where $a_t = K^{-1}(1-t)$ and function $\mathcal{L}$ is a slowly varying as $t \downarrow 0$. Therefore, as $n \to \infty$, we have $n(1-K(u_n^{1/c})) \sim 0$, and

$$n(1-F_Z(u_n)) \sim n(1-K(u_n)) \sim \tau\,,\; n \to \infty\,. \tag{3.6}$$

4 The function ATDF and its estimation

The function ATDF, $\Lambda_m = 2\eta_m - 1$, where η_m is the Ledford and Tawn tail dependence parameter, provides a measure of serial dependence between extreme values lag m apart, analogous to the role of the ACF (*auto-correlation function*) for linear processes (see Subsection 1.1).

A special feature of stationary heavy tailed pARMAX processes is the fact that $\eta_m = \max(c^m, 1/2)$ as can be seen in the next result. Therefore, the ATDF has a power decay as the ACF of AR(1) processes and a cut-off as the MA(p) processes (see Figure 2).

Proposition 4.1 *Let $\{X_i\}_{i\in\mathbb{Z}}$ be a pARMAX process. Then, the random pair (X_1, X_{1+m}) has a coefficient of asymptotic tail dependence $\eta_m = 1/2$ if $c^m \leq 1/2$, and c^m if $c^m > 1/2$.*

Proof Consider notation $a_{xt} = K^{-1}(1 - xt)$. By (1.3), we need to show that,

$$h(x,y) = \lim_{t\downarrow 0} P\big(X_1 > a_{xt}, X_{1+m} > a_{yt}\big)/P\big(X_1 > a_t, X_{1+m} > a_t\big). \qquad (4.1)$$

is homogeneous, where η_m corresponds to the arithmetic inverse of the degree of homogeneity of h. Developing expression in numerator we have:

$$P(X_1 > a_{xt}, X_{1+m} > a_{yt}) = 1 - K(a_{xt}) - F_Z(a_{yt})\big[K(a_{yt}^{1/c^m}) - K(a_{yt})\big] \quad (4.2)$$

By (3.4), (3.5) and (2.3), we have that,

$$P(X_1 > a_{xt}, X_{1+m} > a_{yt}) \sim xyt^2 \mathbf{1}_{\{c^m \leq 1/2\}} + (yt)^{1/c^m} \mathcal{L}\big(a_t^{1/c^m}\big)\mathbf{1}_{\{c^m \geq 1/2\}}\,. \quad (4.3)$$

For denominator in (4.1) just take $x = y = 1$ in (4.3). $\square$

We conclude that, pARMAX processes are asymptotically tail independent Markov chains ($\eta < 1$) and that, the higher the value of parameter c, the higher we must choose lag m in order to get asymptotically independent observations.

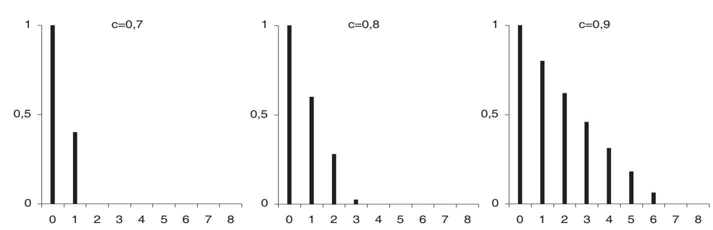

Figure 2: The ATDF (Λ_m) of pARMAX processes with parameters, $c = 0.7$, $c = 0.8$ and $c = 0.9$, respectively, for lags $m = 0, 1, ..., 8$.

For the estimation of ATDF we can use known estimators of η_m, (Ledford and Tawn 1996; Peng 1999; Draisma, Drees, Ferreira and de Haan 2004). However the properties of these latter ones were derived under the assumption of independence between random pairs, (X_i, Y_i), $i = 1, ..., n$, i.i.d. copies of (X, Y), while here we must apply to random pairs, (X_i, X_{i+m}), $i = 1, ..., n$, m fixed, which are obviously dependent. Yet this is easily overcome as we shall see in the next result.

Before going any further, we must observe that, based on (4.1), parameter η_m can be estimated as the tail index of $T_i = \min((1 - K(X_i))^{-1}, (1 - K(X_{i+m}))^{-1})$, in which replacing the unknown marginal d.f., $K(x)$, by the empirical counterpart, $K_n(x) = (1/n)\sum_{i=1}^n \mathbf{1}_{\{X_i \le x\}}$, leads to,

$$T_i^{(n)} := \min\left(\frac{n+1}{n+1-nK_n(X_i)}, \frac{n+1}{n+1-nK_n(X_{i+m})}\right),\ i = 1, ..., n, \qquad (4.4)$$

(Ledford and Tawn 1996; Draisma, Drees, Ferreira and de Haan 2004). The next result refers to tail index estimators of the class of Drees (see Subsection 1.2), which includes Hill, maximum likelihood, Pickands, moments and probability weighted moments estimators.

Theorem 4.1 *Let $\{X_i\}_{i\in\mathbb{Z}}$ be a pARMAX process and let $\left\{T_i^{(n)}\right\}_i$ be a sequence as defined in (4.4). The tail index estimators of the class of Drees based on the sequence $\left\{T_i^{(n)}\right\}_i$ are consistent and asymptotically normal.*

Proof Observe that, as $n \to \infty$

$$P\big(T_i^{(n)} > x\big) \;\sim\; P\Big(X_i > K_n^{-1}\big(1 - \tfrac{1}{x}\big), X_{i+m} > K_n^{-1}\big(1 - \tfrac{1}{x}\big)\Big). \qquad (4.5)$$

Note also that, according to the theorem in Klotz (1973), we have that, $K_n(y) = n^{-1}\sum_{j=1}^n 1_{\{X_j \le y\}} \xrightarrow{P} K(y)$, for any y in the support of X_i, since it is the sum of dependent Bernoulli trials, denoted by $Binomial(n, p, \lambda)$ in Klotz (1973), with $p = P(X_j \le y) = K(y)$ and $\lambda = P(X_j \le y | X_{j-1} \le y)$. Based on the theoretical result, usually attributed to Slutsky, in which, any given random elements, Y_n, Y and Y'_n, such that $Y_n \xrightarrow{d} Y$ and $Y'_n \xrightarrow{P} a$, then, $(Y_n, Y'_n) \xrightarrow{d} (Y, a)$, we have that, for each i, and by the continuous mapping theorem,

$$\begin{aligned} & P\big(X_i > K_n^{-1}\big(1 - \tfrac{1}{x}\big), X_{i+m} > K_n^{-1}\big(1 - \tfrac{1}{x}\big)\big) \\ \underset{n\to\infty}{\longrightarrow}\; & P\big(X_i > K^{-1}\big(1 - \tfrac{1}{x}\big), X_{i+m} > K^{-1}\big(1 - \tfrac{1}{x}\big)\big) \end{aligned} \qquad (4.6)$$

Therefore, given (4.5), we conclude that, $T_i^{(n)} \xrightarrow{d} T_i$. Similarly it is shown that,

$$P\big(T_i^{(n)} > x, T_{i+k}^{(n)} > y\big) \xrightarrow{d} P(T_i > x, T_{i+k} > y). \qquad (4.7)$$

Now we shall see that sequence $\{T_i\}_i$ (and therefore $\{T_i^{(n)}\}_i$ given (4.7)) satisfies Drees conditions in Subsection 1.2.

With respect to the *β-mixing* dependence of $\{T_i\}_i$ it is immediate from the *β-mixing* structure of *p*ARMAX.

By Proposition 4.1, the function $1 - F_T$ is regularly varying with index $\gamma^* = \max(1/2, c^m)$ and, for the sake of simplicity, it is assumed that, for some real constants, $d > 0$ and $d^* > 0$,

$$F_T^{-1}(1-t) \sim d^* t^{-\gamma_*} \quad \text{and} \quad K^{-1}(1-t) \sim dt^{-\gamma}\,, \ \text{ as } t \downarrow 0. \tag{4.8}$$

Denoting the quantile, $K^{-1}\big(1 - 1/F_T^{-1}\big(1 - \frac{k_n}{n}x\big)\big) = a_{n,x}$, we have,

$$\begin{aligned} & P\big(T_1 > F_T^{-1}\big(1 - \tfrac{k_n}{n}x\big), T_{1+m} > F_T^{-1}\big(1 - \tfrac{k_n}{n}y\big)\big) \\ = \ & P(X_1 > a_{n,x}, X_2 > a_{n,x}, X_{1+m} > a_{n,y}, X_{2+m} > a_{n,y}) \end{aligned}$$

Note that, by (4.8), $a_{n,x} \sim \big(\frac{k_n}{n}x\big)^{-\gamma^*\gamma} d(d^*)^\gamma$ and hence, $\forall j \geq 0$,

$$K\big(a_{n,x}^{1/c^j}\big) \sim 1 - \big(\tfrac{k_n}{n}x\big)^{\gamma^*} d^{-1/(\gamma c^j)} (d^*)^{-1/c^j} d^{1/\gamma} \sim 1 - \big(\tfrac{k_n}{n}x\big)^{\gamma^*} A^{1/c^j} d^{1/\gamma}, \tag{4.9}$$

where $A = d^{-1/\gamma}(d^*)^{-1}$. The same reasoning as in Proposition 4.1 leads to,

$$\begin{aligned} & P\big(T_1 > F_T^{-1}\big(1 - \tfrac{k_n}{n}x\big), T_{1+m} > F_T^{-1}\big(1 - \tfrac{k_n}{n}y\big)\big) \\ \underset{t\downarrow 0}{\sim} \ & \big(y\tfrac{k_n}{n}\big)^{\frac{\gamma^*}{c^{m+1}}} d^{1/\gamma} A^{1/c^{m+1}} + \big(xy\big)^{2\gamma^*} \big(\tfrac{k_n}{n}\big)^{4\gamma^*} d^{1/\gamma} A^4 + \big(xy\big)^{\frac{\gamma^*}{c}} \big(\tfrac{k_n}{n}\big)^{2\frac{\gamma^*}{c}} d^{1/\gamma} A^{2/c} \end{aligned}$$

Therefore, condition in (1.6) holds for $\{T_i\}_i$, and by (4.7) also for $\{T_i^{(n)}\}_i$, with $c_m(x, y) = 0$, $\forall m \geq 1$.

With respect to condition (1.7), observe that, we can state,

$$\begin{aligned} & \tfrac{n}{k_n} P\Big(T_1 \in I_n(x,y), T_{1+m} \in I_n(x,y)\Big) \\ \leq \ & \tfrac{n}{k_n}\bigg[P\Big(a_{n,y} < X_1 \leq a_{n,x}, X_1 > a_{n,y}^{1/c^m}\Big) + P\Big(a_{n,y} < X_1 \leq a_{n,x}, \bigvee_{j=0}^{m-1} Z_{m-j+1}^{c^j} > a_{n,y}\Big)\bigg] \\ + & \tfrac{n}{k_n}\bigg[P\Big(a_{n,y} < X_2 \leq a_{n,x}, X_2 > a_{n,y}^{1/c^m}\Big) + P\Big(a_{n,y} < X_2 \leq a_{n,x}, \bigvee_{j=0}^{m-1} Z_{m-j+2}^{c^j} > a_{n,y}\Big)\bigg]. \end{aligned}$$

By the independence assumptions and considering (4.9) and conditions in (1.7), the last expression can be bounded successively by,

$$\begin{aligned} & \tfrac{n}{k_n}\Big\{2\Big[K(a_{n,x}^{1/c^m}) - K(a_{n,y}^{1/c^m})\Big] + 2\Big[K(a_{n,x}) - K(a_{n,y})\Big]\Big[1 - K(a_{n,y})\Big]\Big\} \\ \leq \ & (y-x)\Big(2d^{1/\gamma} A^{1/c^m} \Big[\tfrac{k_n}{n}(1+\epsilon)\Big]^{\gamma^*/c^m - 1} \tfrac{1+\epsilon}{\delta} + \big(\tfrac{k_n}{n}\big)^{2\gamma^* - 1} 2\tfrac{(1+\epsilon)^{2\gamma^*}}{\delta} d^{2/\gamma} A^2\Big) \end{aligned}$$

Take $\delta=y-x$, $\widetilde{\rho}(m)=2d^{1/\gamma}A^{1/c^m}\left[\frac{k_n}{n}(1+\epsilon)\right]^{\gamma^*/c^m-1}\frac{1+\epsilon}{\delta}$, $D_1=2\frac{(1+\epsilon)^{2\gamma^*}}{\delta}d^{2/\gamma}A^2>0$ and $\alpha=2\gamma^*-1$. Note that, $\sum_{m=0}^{\infty}\widetilde{\rho}(m)<\infty$ since, from some order n,

$$\lim_{m\to\infty}\widetilde{\rho}(m+1)/\widetilde{\rho}(m)\sim\left\{A\left[\frac{k_n}{n}(1+\epsilon)\right]^{\gamma^*}\right\}^{\frac{1}{c^m}(1/c-1)}<1.\ \square$$

Remark 4.1 *Observe that, as* $P(\max(U_iX_{i-1},Z_i)>x)\leq P(\max(X_{i-1},Z_i)>x)$*, the survival functions of the pRARMAX process in (1.2) are upper bounded by the respective ones of the pARMAX. Hence, condition (1.6) also holds for pRARMAX with* $c_m(x,y)=0$, $\forall m$*. Condition (1.7) is also straightforward.*

Example 4.1 *Consider pARMAX series with marginals Pareto(*0.7*) and innovations with d.f. (2.4) for* $c=0.7,0.8,0.9$*, respectively (sample size* $n=5000$*). For each of these series, estimates of* η_m *are obtained, for lags* $m=1,...,6$*, by the Hill estimator based on the largest* $k+1$ *order statistics of* $T_i^{(n)}$*, defined as follows:*

$$\widehat{\eta}_m=k^{-1}\sum_{i=1}^{k}\log T^{(n)}_{n-i+1:n}-\log T^{(n)}_{n-k:n}.$$

Results in Table 1 show that the estimates are, in general, very close to their true values. The values of k *were chosen in a range of stability of Hill's estimates.*

5 Return levels in pARMAX

A very important measure of extreme events is the r-year return level, the quantile which has probability $1/r$ of being exceeded by the annual maximum in any particular year, so that it is the level expected to be exceeded on average once every r years. We will see that in pARMAX processes, return levels computation based on (1.15) works better if we replace θ by a threshold-dependent extremal index (Proposition 5.2). We follow the approach considered in Bortot and Tawn, 1998 (see Subsection 1.3).

Proposition 5.1 *Consider* $(u_n)_{n\geq1}$ *a sequence of normalized levels of* K *and* $\{r_n\}_{n\geq1}$ *a nondecreasing integers sequence as stated in (1.14). Then, condition* $\Delta^{(2)}(u_n,r_n)$ *holds for the pARMAX process.*

Proof We must prove that (1.17) holds. According to (2.5), $P\big(\bigvee_{i=2}^{j}X_i\leq u,X_1>u\big)=\big[K(u^{1/c})-K(u)\big]\big[F_Z(u)\big]^{j-1}$, so that applying (4.2) and simple calculations, expression in (1.17) becomes,

$$\frac{[K(u_n^{1/c})-K(u_n)]F_Z(u_n)[1-(F_Z(u_n))^{r_n-2}]}{(1-K(u_n)-F_Z(u_n)[K(u_n^{1/c})-K(u_n)])}.$$

Table 1: Estimates of η_m $(m = 1, ..., 6)$ obtained by Hill estimator based on the largest $k = 200, 500, 1000$ order statistics of $T_i^{(n)}$. The true values for each lag m are on the lines beginning with "η_m"

$c = 0.7$	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$
η_m	0.7	0.5	0.5	0.5	0.5	0.5
$k = 200$	0.69	0.54	0.49	0.48	0.48	0.48
$k = 500$	0.68	0.52	0.48	0.48	0.48	0.48
$k = 1000$	0.66	0.49	0.46	0.48	0.48	0.47
$c = 0.8$	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$
η_m	0.8	0.64	0.51	0.5	0.5	0.5
$k = 200$	0.81	0.67	0.56	0.51	0.49	0.49
$k = 500$	0.79	0.66	0.56	0.5	0.49	0.49
$k = 1000$	0.8	0.62	0.53	0.47	0.46	0.49
$c = 0.9$	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$
η_m	0.9	0.81	0.73	0.66	0.59	0.53
$k = 200$	0.89	0.79	0.7	0.63	0.58	0.54
$k = 500$	0.88	0.79	0.69	0.62	0.56	0.53
$k = 1000$	0.87	0.76	0.66	0.58	0.52	0.48

Considering the first order of Taylor's approach of $(F_Z(u_n))^{r_n-2}$, as $n \to \infty$, we have, $(F_Z(u_n))^{r_n-2} \sim 1-(r_n-2)(1-F_Z(u_n))$. Since $(u_n)_{n\geq 1}$ is a sequence of normalized levels of K, by (3.4), (3.6) and given (3.5), expression (1.17) can be approximated successively,

$$\frac{\left[\frac{\tau}{n}-\left(\frac{\tau}{n}\right)^{1/c}\mathcal{L}(a_{\tau/n}^{1/c})\right]\left(1-\frac{\tau}{n}\right)\frac{\tau}{n}(r_n-2)}{\frac{\tau}{n}-\left(1-\frac{\tau}{n}\right)\left[\frac{\tau}{n}-\left(\frac{\tau}{n}\right)^{1/c}\mathcal{L}(a_{\tau/n}^{1/c})\right]} \sim (r_n-2)\mathbf{1}_{\{c\leq 1/2\}} + \left(\frac{\tau}{n}\right)^{2-1/c}\frac{r_n-2}{\mathcal{L}(a_{\tau/n}^{1/c})}\mathbf{1}_{\{c>1/2\}}.$$

Hence, when $c \leq 1/2$, $\Delta^{(2)}(u_n, r_n)$ holds for $r_n = 2$ and, when $c > 1/2$, the condition holds for any r_n such that, $r_n = o\big(n^{2-1/c}\big)$. $\square$

Using the threshold-dependent extremal index stated in (1.18), the following result can now be presented.

Proposition 5.2 *Let $(u_n)_{n\geq 1}$ and $\{r_n\}_{n\geq 1}$ be real sequences defined as in Proposition 5.1. For pARMAX processs the following approximation holds*

$$P(\textstyle\bigvee_{i=1}^n X_i \leq u_n) - K^n(u_n) \sim e^{-\tau} n^{1-1/c}\mathcal{L}(a_{\tau/n}^{1/c})\tau^{1/c}, \tag{5.1}$$

which leads to,

$$P(\textstyle\bigvee_{i=1}^n X_i \leq u_n) - K^{n\theta(u_n)}(u_n) = o\big(n^{1-1/c}\mathcal{L}(a_{\tau/n}^{1/c})\big), \tag{5.2}$$

with

$$1-\theta(u_n)\sim u_n^{-1/\gamma(1/\eta-1)}\left[L_K(u_n)\mathbf{1}_{\{c\le 1/2\}}+L_K(u_n^{1/c})/L_K(u_n)\mathbf{1}_{\{c>1/2\}}\right],$$

where $\mathcal{L}$ and L_K are slowly varying functions given in (3.5) and (3.2), respectively.

Proof Consider normalized levels u_n of pARMAX such that, $\lim_{n\to\infty} n(1-K(u_n))=\tau$, with $\tau>0$ fixed. Then, using (2.5), (3.4) and (3.5), we derive (5.1) which is analogous to the approximation obtained for m-dependent Gaussian stationary sequences (Rootzén 1983). Hence, with a simple modification as stated in Bortot and Tawn (1998) we obtain (5.2). The threshold-dependent extremal index, which is defined in (1.18), is straightforward from (3.2) and (4.3). $\square$

In model pARMAX, the differences between approximation $P(\bigvee_{i=1}^n X_i\le u)\approx K(u)^{n\theta}$ obtained for $\theta=1$ and obtained for θ replaced by $\theta(u)$ become larger with higher values of parameter c, as can be seen in the following example.

Example 5.1 *Consider again the pARMAX processes of Example 4.1. We obtain somewhat significantly different return levels, if we take $\theta=\theta(u)$ (**I**) and $\theta=1$ (**II**) in the approximation (1.15). In Figure 3 we can see larger differences with larger c. Considering, for instance, the* 100 *year return level of pARMAX (taking $n=250$ observations per year), using approach (**I**) we obtain,* 1126 *with $c=0.7$,* 1184 *with $c=0.8$ and* 893 *with $c=0.9$, and for approach (**II**) we obtain* 1195.

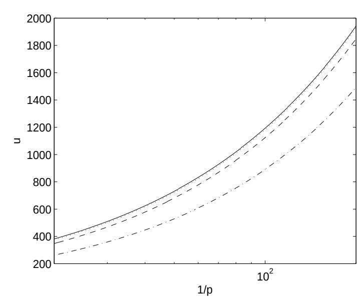

Figure 3: Return levels for pARMAX processes of Example 4.1 against $1/p$, ($p=P(\bigvee_{i=1}^n X_i>u)$), on a logarithmic scale: solid line taking $\theta=1$ (*approach* **II**); dotted line, dashed line and dotted-dashed line taking $\theta=\theta(u)$ in (1.18) (*approach* **I**), with $c=0.7$, $c=0.8$ and $c=0.9$, respectively.

6 Conclusion

In this paper, we have characterized the power max-autoregressive process pARMAX which has the interesting feature of describing an asymptotic tail independent behavior (with positive association, since $1/2 \leq \eta < 1$), i.e., the degree of dependence between exceedances of high levels gradually decreases as the levels become larger (with consecutive observations exceeding a large value occurring more frequently than under exact independence), a property that can be observed in various data series (Ledford and Tawn 1996; Bortot and Tawn 1998). Moreover, the process parameter c is related to the coefficient of asymptotic tail dependence η under the very mild assumption of heavy tailed marginals. Hence the estimation of c is easily derived from the already known estimation procedures of η. From this approach, we also derive a model identification procedure through the ATDF ($\Lambda_m = 2\eta_m - 1$) in a similar manner as ACF for linear models. We notice that the max-autoregressive MARMA models, in particular ARMAX, can only be applied in cases of asymptotic tail dependence ($\eta = 1$). Finally, the characterization presented in this paper also holds for more general and widely applicable pRARMAX in (1.2) for which a methodology of model fitting to an observed data series was derived in Ferreira and Canto e Castro (2010).

References

[1] Alpuim, M. T. (1989). An extremal markovian sequence, *J. Appl. Probab.*, **26**, 219–232.

[2] Ancona-Navarrete, M.A. and Tawn J.A. (2000). A comparison of methods for estimating the extremal index, *Extremes*, **3**, 5–38.

[3] Bortot, P. and Tawn, J.A. (1998). Models for the extremes of Markov Chains, *Biometrika*, **85**, 851–867.

[4] Canto e Castro, L. (1992). *Sobre a Teoria Assintótica de Extremos*, Ph. D. Thesis, Faculty of Sciences of University of Lisbon (FCUL), Portugal.

[5] Davis R. and Resnick, R. (1989). Basic properties and prediction of max-ARMA processes, *Adv. in Appl. Probab.*, **21**, 781–803.

[6] Draisma, G., Drees, H., Ferreira A. and de Haan L. (2004). Bivariate tail estimation: dependence in asymptotic independence, *Bernoulli* **10**, 251–280.

[7] Drees, H. (2003). Extreme quantile estimation for dependent data with applications to finance, *Bernoulli*, **9**, 617–657.

[8] Daley D. and Haslett J. (1982). A thermal energy storage with controlled input, *Adv. in Appl. Probab.*, **14(2)**, 257–271.

[9] Embrechts, P., Klppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, Springer-Verlag, Berlin.

[10] Ferreira M. and Canto e Castro L. (2010). Modeling rare events through a pRARMAX process, accepted for publication in J. Statist. Plann. Inference, doi:10.1016/j.jspi.2010.05.024.

[11] Helland I. and Nilsen T. (1976). On a general random exchange model, *J. Appl. Probab.*, **13(4)**, 781–790.

[12] Klotz, J. (1973). Statistical Inference in Bernoulli Trials with dependence, *Ann. Statist.*, **1**, 373–379.

[13] Lebedev, A. V. (2008). Statistical analysis of first-order MARMA Processes, *Math. Notes* **83**, 506–511.

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

[15] A. Ledford and J. A. Tawn (2003). Diagnostics for dependence within time series, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **65**, 521–543.

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

[17] Peng, L. (1999). Estimation of the coefficient of the tail dependence in bivariate extremes, *Statist. Probab. Lett.*, **43**, 399–409.

[18] Rootzén, H. (1983). The rate of convergence of extremes of stationary normal sequences, *Adv. in Appl. Probab.*, **15**, 54–80.

[19] Zhang, Z. and Smith, R. L. (2001). Modelling Financial Time Series Data as Moving Maxima Processes, Tech. Rep. Dept. Stat. (Univ. North Carolina, Chapel Hill, NC, 2001); http://www.stat.unc.edu/faculty/rs/ papers/RLS_Papers.html.