REVSTAT – Statistical Journal
Volume 8, Number 2, November 2010, 139–149

ESTIMATION OF THE PARAMETER OF A *pARMAX* MODEL

Author: Marta Ferreira
– Department of Mathematics, University of Minho, Portugal
msferreira@math.uminho.pt

Received: November 2009 Revised: July 2010 Accepted: September 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. In fact, they are more convenient for analysis than heavy-tailed ARMA, as their finite-dimensional distributions can easily be written explicitly. The recent power max-autoregressive model (*pARMAX*) has the interesting feature of describing an asymptotic independent tail behavior, a property that can be observed in various data series. An estimator of the model parameter c $(0 < c < 1)$ is already available in the literature, but only in the restrictive case $c > 1/2$. Here it is presented an estimator for all $c \in (0, 1)$. Consistency and asymptotic normality are also stated.

Key-Words:

- *extreme value theory; max-autoregressive processes.*

AMS Subject Classification:

- 60G70, 60J10.

1. INTRODUCTION

Extreme Value Theory (EVT) is an important tool for many applied sciences whenever we are faced with modeling high values of certain phenomena. Ocean wave modeling, wind engineering, thermodynamics of earthquakes, risk assessment on financial markets are some examples. 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 [7]), and also the particular case MAR(1) or ARMAX, given by,

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

with $0 < k < 1$ and $\{W_i\}_{i\in\mathbb{Z}}$ i.i.d. (Alpuim [2]; Canto e Castro [6]; Ancona-Navarrete and Tawn [3]; Beirlant *et al.* [4]; Lebedev [12]) are some examples. Heavy tailed MARMA and ARMA are both good choices for modeling time series data with sudden large peaks, although the former are more convenient for analysis as their finite-dimensional distributions can easily be written explicitly.

More recently, some careful attention has been given to the statistical modeling of the tail dependence between consecutive pairs from a stationary first-order Markov chain, since it is important to distinguish asymptotic dependence from asymptotic independence. More precisely, according to Bortot and Tawn ([5]), a Markov chain $\{Y_i\}$ is said to be asymptotically tail dependent or independent, whenever $b > 0$ or $b = 0$, respectively, in the limit below:

$$\lim_{y\to y^*} P\big(Y_2 > y \,|\, Y_1 > y\big) \,=\, b \ ,$$

where y^* is the right-endpoint of Y_1, i.e., $y^* = \sup\{y : P(Y_1 \le y) < 1\}$. For asymptotically tail independent Markov chains, the dependence between exceedances of y gradually decreases as $y \to y^*$, which leads to an extremal feature increasingly resembling an i.i.d. sequence at high levels. As pointed out in Bortot and Tawn ([5]), this phenomenon has been noticed in a number of data and theoretical applications. In these cases, procedures as in Smith *et al.* ([16]) assuming that the limiting behavior of the chain is exact above a fixed high threshold, and hence the dependence structure between consecutive random variables (r.v.'s) above the threshold can be modeled through a bivariate extreme value distribution, are not suitable. This problem is overcome by setting the way how $P\big(Y_2 > y \,|\, Y_1 > y\big)$ converges to zero, as $y \to y^*$, which involves the coefficient of asymptotic tail dependence η (Ledford and Tawn [13], [14]). This is a nontrivial class, including many commonly studied processes, such as Gaussian Markov chains (Sibuya [15]).

Coefficient η characterizes the asymptotic tail dependence behavior, i.e., $\eta = 1$ corresponds to tail dependence whilst $\eta < 1$ means asymptotic tail independence, with $\eta = 1/2$ occurring for the (almost) independent case. The *ARMAX* process, which has unit η, is in the group of tail dependent Markov chains (Ferreira and Canto e Castro [8]) and hence is not suitable to model data series expressing the described phenomenon.

Ferreira and Canto e Castro ([8]) introduced the power max-autoregressive process (in short, *pARMAX*), defined as,

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

with $\{Z_i\}$ i.i.d., for which η is a function of the model parameter c, under the very mild assumption of heavy tailed innovations. More precisely, we have $\eta = \max(1/2, c)$ and hence *pARMAX* is an asymptotic tail independent process, even almost independent in cases $c \leq 1/2$. Hence, it is a suitable model to describe the above mentioned phenomenon of time series exhibiting asymptotic tail independence. In Figure 1, the similarity between the sample paths of heavy tailed *pARMAX* and AR(1) processes, in this case based on marginal d.f.'s Pareto($1/\gamma$), with shape parameter $\gamma > 0$, given by

$$K(x) = 1 - x^{-1/\gamma} \ , \qquad x \geq 1 \ , \tag{1.1}$$

indicate that the former can be considered as an alternative for data modeling, particularly with respect to extreme values. The *pARMAX* process has easily derived extremal properties and also easily explicited finite-dimensional distributions (Ferreira and Canto e Castro [8], [9]). Moreover, a generalization of *pARMAX* has also been applied in modeling financial data (Ferreira and Canto e Castro [10]). Based on the estimation procedure for the Ledford and Tawn coefficient η, Ferreira and Canto e Castro ([9]) presented consistent and asymptotically normal estimators for the process parameter c, which applies only in cases where $c > 1/2$. Following a similar procedure to that of Lebedev ([12]) to estimate the parameter of unit Fréchet ARMAX, an estimator for the *pARMAX* parameter c is derived, this time covering all values of $c \in (0, 1)$. From a Klotz's result (Klotz [11], Theorem 1), consistence and asymptotic normality are easily stated.

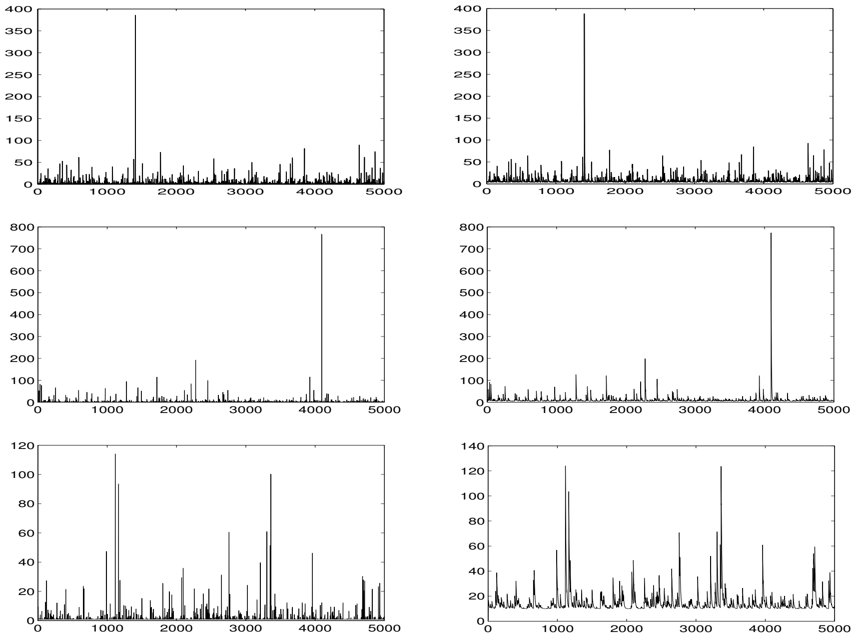

Figure 1: 5000 realizations of *pARMAX*, $X_i = X_{i-1}^c \vee Z_i$, on the left, and of AR(1), $X_i = c\,X_{i-1} + Z_i$, on the right, with, from top to bottom, $c = 0.7$, 0.8, 0.9, respectively, and with marginal Pareto(0.7).

2. THE *pARMAX* PROCESS

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

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

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

In the sequel we consider that $\{Z_i\}$ has support in $[1, \infty[$, a necessary condition for stationarity.

Let K be the marginal distribution function (d.f.) of the process. Hence K is a solution of the equation

$$K(x) = K(x^{1/c})\, F_Z(x) \ . \tag{2.2}$$

(See Ferreira and Canto e Castro [8], [9] for details). An example of a stationary *pARMAX* process is given below.

Example 2.1. Consider $\{Z_i\}$ with common d.f.,

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

where $\mathbf{1}_{\{.\}}$ is the indicator function. Hence, the Pareto($1/\gamma$) d.f. given in (1.1) satisfies (2.2), being, therefore, a stationary distribution for X_i.

The *k-step* transition probability function (t.p.f.) from x to $]-\infty, y]$, given by,

$$Q^k\big(x,]-\infty, y]\big) := P\big(X_{n+k} \le y \,|\, X_n = x\big) = \frac{K(y)}{K(y^{1/c^k})}\,\mathbf{1}_{\{x \le y^{1/c^k}\}}\ , \tag{2.3}$$

where the last step is due to (2.2), will be used in the forward results.

2.1. Parameter estimation

Now we will present an estimator for the pARMAX parameter (c) based on a similar procedure as in Lebedev ([12]) for unit Fréchet max-autoregressive, $X_i = \max\big(cX_{i-1}, (1-c)Z_i\big)$. In the pARMAX case, Pareto marginals will be considered.

Set, for each $k \ge 1$,

$$p_k = P\big(X_{k+1} \le X_1\big)\ . \tag{2.4}$$

The following result states a relation between p_k and parameter c, more precisely, c^k. For sake of simplicity, from now on consider $a_k := c^k$.

Proposition 2.1. *Let $\{X_i\}$ be a stationary pARMAX process as defined in (2.1) with marginal d.f. K satisfying (1.1). Then the equality,*

$$p_k = a_k\big(\psi(2\,a_k) - \psi(a_k)\big)\ , \tag{2.5}$$

holds where ψ is the well-known digamma function, i.e., $\psi(z) = \Gamma'(z)/\Gamma(z)$ with $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$ the Euler Gamma function.

Proof: Just observe that, using (2.3), we have,

$$\begin{aligned} p_k &= \int P\big(X_{k+1} \le x \,|\, X_1 = x\big)\, dK(x) \\ &= \int Q^k\big(x,]-\infty, x]\big)\, dK(x) \\ &= \int \frac{K(x)}{K(x^{1/c^k})}\, dK(x)\ , \qquad k \ge 1\ . \end{aligned} \tag{2.6}$$

where after some algebra (see for instance Abramowitz and Stegun [1]) and notation $a_k = c^k$, expression (2.5) can be derived. □

Note that p_k does not depend on the marginal d.f. parameter γ. There exist simple estimates for the above probabilities:

$$\widehat{p}_k = \frac{1}{n-k} \sum_{j=k+1}^{n} \mathbf{1}_{\{X_j \le X_{j-k}\}} , \qquad k \ge 1 . \tag{2.7}$$

The next result states consistency and asymptotic normality for estimators $\widehat{a}_k$ $(=\widehat{c}^k)$, obtained from equation (2.5) by plugging in the empirical estimates $\widehat{p}_k$. More precisely, we have the following result.

Proposition 2.2. *Let $\{X_i\}$ be a stationary pARMAX process as defined in Proposition 2.1. Then, for each positive integer k,*

$$n^{1/2}(\widehat{a}_k - a_k) \xrightarrow{D} N\big(0, \sigma_k^2/g'(a_k)^2\big) \tag{2.8}$$

where $g(x) = x\big(\psi(2x) - \psi(x)\big)$ and

$$\sigma_k^2 = p_k(1-p_k)\,(1 - 2p_k + \lambda_k)/(1-\lambda_k) , \tag{2.9}$$

with p_k given in (2.5) and $\lambda_k = p_k^{-1} \int_1^\infty \frac{1}{2}\Big[\frac{K(x)\,K(x^{1/a_{k-1}})}{K^2(x^{1/a_k})} + \frac{K^2(x^{a_1})}{K(x)\,K(x^{1/a_{k-1}})}\Big]\,K(dx)$.

Proof: Observe that $\widehat{p}_k$ is the mean of Bernoulli trials with Markov dependence. From Theorem 1 in Klotz ([11]), convergence $n^{1/2}(\widehat{p}_k - p_k) \xrightarrow{D} N(0, \sigma_k^2)$ holds for σ_k^2 given in (2.9), where $\lambda_k = P\big(X_j \le X_{j-k} \,|\, X_{j-1} \le X_{j-k-1}\big)$ with $\max\big(0, (2p_k - 1)/p_k\big) \le \lambda_k \le 1$. Hence, the result (2.8) is straightforward by the Delta Method.

In order to obtain the variance in (2.9) we must compute λ_k. First note that,

$$\lambda_k = \frac{P\big(X_j \le X_{j-k},\ X_{j-1} \le X_{j-k-1}\big)}{p_k} , \tag{2.10}$$

in which, using successive conditioning on the numerator lead us to,

$$\begin{aligned} &P\big(X_j \le X_{j-k},\ X_{j-1} \le X_{j-k-1}\big) = \\ &\qquad = \int_1^\infty \int_1^\infty \int_1^x Q\big(w,]-\infty, y]\big)\, Q^{k-1}(y, dw)\, Q(x, dy)\, K(dx) . \end{aligned}$$

Now considering (2.3), the following development holds:

$$\begin{aligned} &P\big(X_j \le X_{j-k},\ X_{j-1} \le X_{j-k-1}\big) = \\ &= \int_1^\infty \int_1^\infty \int_1^{\min(x, y^{1/c})} F_Z(y)\, Q^{k-1}(y, dw)\, Q(x, dy)\, K(dx) \\ &= \int_1^\infty \bigg[\int_1^{x^c} F_Z(y)\, Q^{k-1}\big(y,]-\infty, y^{1/c}]\big) + \int_{x^c}^\infty F_Z(y)\, Q^{k-1}\big(y,]-\infty, x]\big)\bigg] Q(x, dy)\, K(dx) \\ &= \int_1^\infty \bigg[\int_1^{x^c} F_Z(y)\, \frac{K(y^{1/c})}{K(y^{1/c^k})} + \int_{x^c}^{x^{1/c^{k-1}}} F_Z(y)\, \frac{K(x)}{K(x^{1/c^{k-1}})}\bigg] Q(x, dy)\, K(dx) . \end{aligned}$$

If d.f. F_Z admits density f_Z, the transition density of (2.3) is given by $q(x,y) = f_Z(y)\,\mathbf{1}_{\{x^c<y\}} + F_Z(x^c)\,\mathbf{1}_{\{x^c=y\}}$. Thus, the first term in the last integral is null and hence,

$$(2.11)\qquad \begin{aligned} &P\big(X_j \le X_{j-k},\ X_{j-1} \le X_{j-k-1}\big) = \\ &= \int_1^\infty \frac{K(x)}{K\big(x^{1/c^{k-1}}\big)}\,\frac{F_Z^2\big(x^{1/c^{k-1}}\big)-F_Z^2(x^c)}{2}\,K(dx) + \int_1^\infty \frac{K(x)}{K\big(x^{1/c^{k-1}}\big)}\,F_Z^2(x^c)\,K(dx)\ . \end{aligned}$$

Now the result follows from equation (2.2) and notation $a_k = c^k$. $\square$

Note that $p_k \in (1/2,1)$ (see Figure 2 and Table 1) and no definite results can be obtained for $\widehat{p}_k < 1/2$.

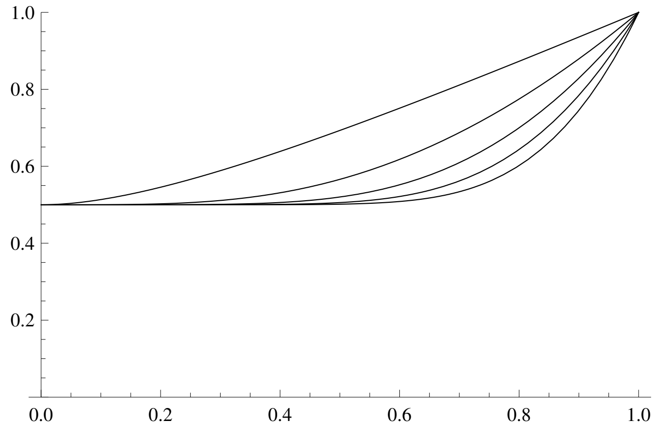

Figure 2: Plot of (from top to bottom) $p_1, ..., p_5$ given in (2.5).

Table 1: Values of p_k computed from (2.5), for Pareto marginal *pARMAX* processes with parameter values: $c = 0.1, 0.2, ..., 0.9$.

c	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
0.1	0.513472	0.500161	0.500002	0.5	0.5
0.2	0.545531	0.502419	0.500103	0.500004	0.5
0.3	0.588572	0.511114	0.501132	0.500106	0.50001
0.4	0.63855	0.531074	0.505905	0.501021	0.500169
0.5	0.693147	0.565986	0.52013	0.505648	0.501503
0.6	0.750948	0.617901	0.551814	0.521466	0.50849
0.7	0.811047	0.687525	0.609375	0.561751	0.533835
0.8	0.872845	0.77475	0.699936	0.643619	0.601832
0.9	0.935927	0.879101	0.828815	0.784424	0.74534

However, the probability of such events goes to zero as $n \to \infty$ and hence, this may be an indication of an inconsistency in our choice of the model. In what concerns the lag k, it can be chosen in order to obtain the smallest variance (σ_k^2) provided that the estimate, $\widehat{p}_k$, takes value in $(1/2, 1)$, which means as small as possible (see, for instance, Table 2).

Table 2: True values of λ_k and respective estimates, $\widehat{\lambda_k}$ in (2.12) and $\widetilde{\lambda_k}$ in (2.13), considering $n = 5000$ realizations of process *pARMAX* for cases $c = 0.3$, 0.5, 0.7, 0.9, with marginal Pareto(1).

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$c=0.3$					
λ_k	0.4805	0.5744	0.6070	0.6122	0.6128
$\widehat{\lambda_k}$	0.4778	0.5751	0.5991	0.6032	0.6071
$\widetilde{\lambda_k}$	0.4782	0.5754	0.5995	0.6035	0.6075
$c=0.5$					
λ_k	0.6393	0.6756	0.7047	0.7195	0.7250
$\widehat{\lambda_k}$	0.6461	0.6686	0.7057	0.7238	0.7236
$\widetilde{\lambda_k}$	0.6461	0.6687	0.7058	0.7239	0.7237
$c=0.7$					
λ_k	0.7930	0.8021	0.8119	0.8210	0.8283
$\widehat{\lambda_k}$	0.7973	0.8083	0.8114	0.8269	0.8277
$\widetilde{\lambda_k}$	0.7975	0.8083	0.8116	0.8270	0.8277
$c=0.9$					
λ_k	0.9341	0.9348	0.9356	0.9364	0.9373
$\widehat{\lambda_k}$	0.9334	0.9334	0.9342	0.9363	0.9362
$\widetilde{\lambda_k}$	0.9334	0.9334	0.9341	0.9362	0.9361

We remark that this procedure allows to estimate any value of $c \in (0, 1)$, and not only the case $c \in (1/2, 1)$ as in the method considered in Ferreira and Canto e Castro ([9]), which is based on the estimation of Ledford and Tawn tail dependence coefficient η. On the other hand, there is no explicit form for a_k in (2.5) and so it must be obtained numerically. Table 1 presents some computed values.

2.2. An illustrative example

An illustration is now presented. We consider 5000 realizations from *pARMAX* process in (2.1), for cases $c = 0.3, 0.5, 0.7, 0.9$, with marginal distribution Pareto(1).

In order to obtain an estimate for the variance, we can replace in (2.9), p by $\widehat{p}_k$ stated in (2.7) and λ_k by the empirical counterpart

$$\widehat{\lambda}_k = \frac{1}{n-k-1} \sum_{k+2}^{n} \mathbf{1}_{\{X_j \le X_{j-k},\, X_{j-1} \le X_{j-k-1}\}} \Big/ \widehat{p}_k \tag{2.12}$$

or alternatively, use the estimator proposed by Klotz ([11]),

$$\widetilde{\lambda}_k = \frac{r - \widehat{q}_k(2s-t) + (n-1)\,\widehat{p}_k + \Big(\big(r - \widehat{q}_k(2s-t) + (n-1)\,\widehat{p}_k\big)^2 + 4\,r\,(1-2\widehat{p}_k)\,(n-1)\,\widehat{p}_k\Big)^{1/2}}{2\,(n-1)\,\widehat{p}_k} \tag{2.13}$$

where $\widehat{q}_k = 1 - \widehat{p}_k$, $r = \sum_{i=2}^{n} x_i\, x_{i-1}$, $s = \sum_{i=1}^{n} x_i$ and $t = x_1 + x_n$, which is asymptotically equivalent to the maximum likelihood estimator. Again by Theorem 1 in Klotz ([11]), we have that $\widetilde{\lambda}_k$ is consistent, more precisely, $\sqrt{n}\,(\lambda_k - \widetilde{\lambda}_k) \overset{D}{\to} N\big(0, \lambda(1-\lambda)/p\big)$. See the very close estimates obtained for λ_k in Table 2. Results of estimation are summarized in Table 3.

Table 3: True values of a_k $(= c^k)$ and estimates obtained from (2.5), considering $n = 5000$ realizations of process *pARMAX* in (2.1), with marginal Pareto(1), for cases $c = 0.3, 0.5, 0.7, 0.9$; estimates $\widehat{c}$ were obtained by taking $\widehat{a_k}^{1/k}$; IC(λ),IC($\widehat{\lambda}$) and IC($\widetilde{\lambda}$) are 95% confidence intervals obtained, respectively, with true σ^2 and estimated σ^2 using $\widehat{\lambda}$ given in (2.12) and $\widetilde{\lambda}$ given in (2.13); non filled cells mean that a $\widehat{p}_k$ less than 0.5 was obtained.

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
a_k	0.3	0.09	0.027	0.0081	0.00243
IC(λ)	(0.2778, 0.3222)	(0.0203, 0.1598)	(−0.1840,0.2380)	(−0.6627, 0.6789)	(−2.1955, 2.2004)
$\widehat{a_k}$	0.295616	0.093871	—	—	—
IC($\widehat{\lambda}$)	(0.2734,0.3178)	(0.0262,0.1615)	—	—	—
IC($\widetilde{\lambda}$)	(0.2734, 0.3179)	(0.0262, 0.1616)	—	—	—
$\widehat{c}$	0.295616	0.306384	—	—	—
a_k	0.5	0.25	0.125	0.0625	0.03125
IC(λ)	(0.4810, 0.5190)	(0.2088, 0.2912)	(0.0526,0.1974)	(−0.0672, 0.1922)	(−0.2100, 0.2725)
$\widehat{a_k}$	0.500694	0.258363	0.137758	0.062246	0.081734
IC($\widehat{\lambda}$)	(0.4814, 0.5200)	(0.2184, 0.2983)	(0.0701,0.2054)	(−0.0678, 0.1923)	(−0.0244, 0.1880)
IC($\widetilde{\lambda}$)	(0.4814, 0.5200)	(0.2184, 0.2983)	(0.0701,0.2054)	(−0.0678, 0.1923)	(−0.0245, 0.1879)
$\widehat{c}$	0.500694	0.508294	0.516463	0.499491	0.606011
a_k	0.7	0.49	0.343	0.2401	0.16807
IC(λ)	(0.6838, 0.7162)	(0.4563, 0.5237)	(0.2947,0.3913)	(0.1760, 0.3042)	(0.0843, 0.2519)
$\widehat{a_k}$	0.682445	0.469803	0.334072	0.222017	0.149248
IC($\widehat{\lambda}$)	(0.6661, 0.6989)	(0.4355, 0.5041)	(0.2850,0.3831)	(0.1543, 0.2897)	(0.0589, 0.2396)
IC($\widetilde{\lambda}$)	(0.6660, 0.6990)	(0.4355, 0.5042)	(0.2850,0.3832)	(0.1543, 0.2897)	(0.0589, 0.2396)
$\widehat{c}$	0.682445	0.685422	0.693873	0.686430	0.683568
a_k	0.9	0.81	0.729	0.6561	0.59049
IC(λ)	(0.8896, 0.9104)	(0.7862, 0.8338)	(0.6937,0.7643)	(0.6106, 0.7016)	(0.5357, 0.6452)
$\widehat{a_k}$	0.896950	0.803367	0.721969	0.653779	0.583686
IC($\widehat{\lambda}$)	(0.8862, 0.9077)	(0.7790, 0.8277)	(0.6862,0.7577)	(0.6080, 0.6995)	(0.5286, 0.6388)
IC($\widetilde{\lambda}$)	(0.8862, 0.9077)	(0.7790, 0.8277)	(0.6863,0.7577)	(0.6081, 0.6995)	(0.5286, 0.6388)
$\widehat{c}$	0.896950	0.896307	0.897097	0.899203	0.897916

ACKNOWLEDGMENTS

This work has been supported by the grant number FCT/SFRH/BD/38867/2007 from FCT. I also acknowledge the valuable suggestions from the referee.

REFERENCES

[1] Abramowitz, M. and Stegun, I. (1972). *Handbook of Mathematical Functions: with formulas, graphs and mathematical tables*, New York.

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

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

[4] Beirlant, J.; Goegebeur, Y.; Segers, J. and Teugels, J. (2004). *Statistics of Extremes: Theory and Application*, John Wiley & Sons, England.

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

[6] Canto e Castro, L. (1992). *Sobre a Teoria Assintótica de Extremos*, PhD Thesis, FCUL.

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

[8] Ferreira, M. and Canto e Castro, L. (2008). Tail and dependence behavior of levels that persist for a fixed period of time, *Extremes*, **11**, 113–133.

[9] Ferreira, M. and Canto e Castro, L. (2010a). Asymptotic and pre-asymptotic tail behavior of a power max-autoregressive model, *ProbStat Forum*, **3**, 91–107.

[10] Ferreira, M. and Canto e Castro, L. (2010b). Modeling rare events through a pRARMAX process, *J. Statist. Plann. Inference*, **140**(11), 3552–3566.

[11] Klotz, J. (1973). Statistical inference in Bernoulli trials with dependence, *Ann. Statist.*, **1**, 373–379.

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

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

[14] Ledford, A. and Tawn, J.A. (1997). Modelling Dependence within joint tail regions, *J. Royal Statist. Soc. Ser. B*, **59**, 475–499.

[15] Sibuya, M. (1960). Bivariate extreme statistics, *Ann. Inst. Statist. Math.*, **11**, 195–210.

[16] Smith, R.L.; Tawn, J.A. and Coles, S.G. (1997). Markov chain models for threshold exceedances, *Biometrika*, **84**, 249–68.