Contributions for modeling characterization of heavy-tail time series

Marta Ferreira[1]

[1] Center of Mathematics of Minho University, Center for Computational and Stochastic Mathematics of University of Lisbon, Center of Statistics and Applications of University of Lisbon, Portugal

E-mail for correspondence: msferreira@math.uminho.pt

Abstract: The occurrence of extreme phenomena and their devastating impact have been on the agenda, especially in areas of environmental and economic-financial sciences, extending to insurance activity. The theory of extreme values allows an adequate approach in the statistical study of data associated with this type of phenomena. Heavy tail models thus play an important role and are increasingly a resource. In this work we will revisit some max/min-autoregressive and maximum-moving models and contribute to their characterization by deriving their autocorrelation structure based on the Spearman and Kendall coefficients, both useful tools in the identification of models in real data applications.

Keywords: Extreme values theory; Stationary sequences; Spearman correlation; Kendall correlation.

1 Introduction

Climate change, as well as economic and financial crisis, are affecting our daily lives, with impacts that can be tragic. Data analysis of extreme values has been assuming a growing importance between the scientific community and the theory of extreme values offers itself as an appropriate tool. The occurrence of extreme phenomena generates data from heavy tails, that is, data from random variables (r.v.'s) whose linearly normalized maximum distribution lies in the domain of attraction of a Fréchet law. ARMA linear models with Gaussian innovations, while very versatile and popular, are not suitable for heavy-tail data. MARMA or max-ARMA models, in which the sum operator (Σ) and the Gaussian law are respectively replaced by the maximum operator ($\bigvee$) and the Fréchet law, appeared as an alternative,

with a simpler theoretical treatment (Davis and Resnick, 1989). Pareto autoregressive models are based on the minimum operator ($\bigwedge$) and, as the name implies, have a Pareto marginal law. Various application properties and/or examples can be seen in, e.g., Alpuim (1989), Arnold (2001), Carcea *et al.* (2015), Ferreira and Ferreira (2013), and references therein.
In this work we will consider the first-order max-autoregressive model MARMA(1,0) or MAR(1), the moving-maximum model MARMA(0,1) or MMA(1) and the first order Pareto autoregressive model, Yeh-Arnold-Robertson Pareto(III), denoted YARP(1). Our aim is to characterize the behavior of the autocorrelation function, contributing to the identification of these processes in the context of modeling. Since this function implies the existence of finite second moments, which is not always assured in heavy-tail models, we will derive the variants of Spearman's ordinal and Kendall's concordance autocorrelation. Based on a simulation study, we will see that classical estimators of Spearman and Kendall coefficients have a good performance.

2 MAR(1), MMA(1) and YARP(1) models

Consider the innovations sequence $\{Z_n\}_{n\geq 1}$ of independent and identically distributed (i.i.d.) r.v.'s with unit Fréchet distribution function (d.f.), i.e., $F_Z(x)=\exp(-1/x)$, $x>0$. Then $\{X_n\}_{n\geq 1}$, satisfying the recursion,

$$X_n = cX_{n-1} \vee (1-c)Z_n,\ 0<c<1, \tag{1}$$

is said to be a first order max-autoregressive process, denoted MAR(1). Assume that X_0 has also a unit Fréchet d.f., and thus $F_X(x)=\exp(-1/x)=F_Z(x)\equiv F(x)$, $x>0$.
The first order moving-maximum, MMA(1), satisfies the recursion,

$$X_n = cZ_n \vee (1-c)Z_{n-1},\ 0<c<1 \tag{2}$$

Consider Z_0 with unit Fréchet d.f., and thus $F_X(x)=\exp(-1/x)=F_Z(x)\equiv F(x)$, $x>0$.
Consider the innovations sequence $\{\varepsilon_n\}_{n\geq 1}$, whose r.v.'s are i.i.d. coming from a Pareto(III)(0,σ,α), i.e., $1-F_X(x)=\left[1+(x/\sigma)^{\alpha}\right]^{-1}$, $\sigma,\alpha>0$ and sequence $\{U_n\}_{n\geq 1}$ of i.i.d. r.v.'s coming from Bernoulli(p), $0<p<1$, independent of ε_n, $n\geq 1$. The sequence $\{X_n\}_{n\geq 1}$, is said to be a first order autoregressive Yeh-Arnold-Robertson Pareto(III), in short YARP(1), if

$$X_n = p^{-1/\alpha}X_{n-1} \wedge \frac{1}{1-U_n}\varepsilon_n \tag{3}$$

holds, where $1/0$ corresponds to $+\infty$. The marginal d.f. is Pareto(III)(0,σ,α), and we consider $X_0 \frown$Pareto(III)(0,σ,α) (Arnold, 2001). In the following we assume, without loss of generalization, that $\sigma=\alpha=1$.

3 Results and conclusions

Pearson's linear correlation is perhaps the most well-known and used measure of dependence. Its popularity is essentially due to the fact that it is the main measure of dependence on the elliptical distributions, such as the multivariate normal and t-Student. However, outside this family of distributions can lead to a poor or erroneous characterization, counting that, it presupposes the existence of second finite moments. Spearman ($\rho(S)$) and Kendall (τ) correlation coefficients thus arise as alternative measures. Both coefficients can be formulated based on the copula function (Embrechts *et al.* 2002) and can be applied to vectors (X_1, X_{1+m}) of a stationary sequence $\{X_n\}_{n\geq 1}$, with common marginal fd F, now called *lag-m autocorrelation*, for $m = 0, 1, 2, \ldots$. Defining the lag-m copula function

$$C_m(u,v) = P(F(X_1) \leq u, F(X_{1+m}) \leq v),\ (u,v) \in [0,1]^2,$$

we have

$$\rho_m^{(S)} = 12 \int_0^1 \int_0^1 C_m(u,v)dudv - 3,\ \tau_m = 4 \int_0^1 \int_0^1 C_m(u,v)dC_m(u,v) - 1\,. \quad (4)$$

In this work, based on (4), we derive the expressions of Spearman and Kendall coefficients within MAR(1), MMA(1) and YARP(1) processes, stated below:

- MAR(1): $\rho_m^{(S)} = \frac{3c^m}{c^m+2}$ and $\tau_m = c^m$;
- MMA(1): $\rho_1^{(S)} = \frac{3c(1-c)}{2-c(1-c)}$, $\tau_1 = \frac{c(1-c)}{1-c(1-c)}$ and $\rho_m^{(S)} = \tau_m = 0$, if $m > 1$.
- YARP(1): $\rho_m^{(S)} = \frac{3p^m(1-p^{2m}+2p^m \log(p^m))}{(1-p^m)^3}$ and $\tau_m = \frac{2p^m(1+p^m(-1+\log p^m))}{(1-p^m)^2}$.

Observe that MAR(1) Kendal's coefficient presents a geometric decrease, similar to the first order linear auto-regressive process AR(1).
In representing the curves of each coefficient as function of the model parameter, for each lag-m, we see that Spearman's autocorrelation is greater than Kendall's autocorrelation at smaller lags in all processes. Both types of dependence increase as the model parameter also grows, in the case of MAR(1) and YARP(1) processes, which is expected according to their generation formula, respectively (1) and (3). This is not the case of the MMA(1) process in (2) where the lag-1 autocorrelation increases if $c \in]0, 1/2[$ but decreases as $c \in]1/2, 1[$ grows. Both autocorrelation functions are very similar within MAR(1) and YARP(1). However, whatever the lag-m, the YARP(1) process always presents the highest correlations. On the other hand, the MMA(1) presents the lowest ones.
A simulation study was conducted in order to evaluate the performance of the classical estimators of Spearman and Kendall autocorrelation coefficients. We have simulated 200 replicas of samples of size $n = 1000$ from

each process, considering the values 0.25, 0.5, 0.75 for the respective parameters, at the lags $m = 1, 2, 3, 4$, and derived the absolute bias (abias) and root mean squared error (rmse). The results are better for smaller lags and are quite similar between the processes and between the coefficients. In the larger lags, estimates tend to be slightly better for lower values of the model parameter. To illustrate, we present the results of the MAR(1) process in Table 1.

TABLE 1. Simulation results of MAR(1) process.

	m=1		m=2		m=3		m=4	
$\rho_m^{(S)}$	abias	rmse	abias	rmse	abias	rmse	abias	rmse
c=0.25	0.0224	0.0282	0.0255	0.0318	0.0257	0.032	0.0272	0.0338
c=0.5	0.0186	0.0244	0.0299	0.0386	0.0347	0.0439	0.0358	0.0439
c=0.75	0.0134	0.0167	0.0253	0.0311	0.0340	0.0417	0.0407	0.0501
	m=1		m=2		m=3		m=4	
τ_m	abias	rmse	abias	rmse	abias	rmse	abias	rmse
c=0.25	0.0151	0.0190	0.0170	0.0212	0.0171	0.0213	0.0181	0.0226
c=0.5	0.0147	0.0190	0.0207	0.0268	0.0232	0.0293	0.0237	0.0290
c=0.75	0.0121	0.0147	0.0201	0.0247	0.0255	0.0312	0.0294	0.0358

Acknowledgments: The author was financed by Portuguese Funds through FCT - Fundação para a Ciência e a Tecnologia within the Projects UID/MAT/00013/2013, UID/MAT/00006/2019 and by the research center CEMAT (Instituto Superior Técnico, Universidade de Lisboa) through the Project UID/Multi/04621/2013.

References

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

Arnold, B.C. (2001). Pareto Processes. In: *Handbook of Statistics (D.N. Shanbhag and C.R. Rao, eds.)*, Vol. 19, Elsevier Science B.V.

Carcea, M. and Serfling, R. (2015). A Gini Autocovariance Function for Time Series Modelling. *J. Time. Ser. Anal.*, **36**, 817 – 838

Davis R.A. and Resnick, S.I. (1989). Basic properties and prediction of max-ARMA processes. *Adv. in Appl. Probab.*, **21(4)**, 781 – 803.

Embrechts, P., McNeil, A. and Straumann, D. (2002). Correlation and Dependence in Risk Management: Properties and Pitfalls. in: *Risk Man-*

agement: Value at Risk and Beyond. 1st Edition, Cambridge University Press, Cambridge, 176 – 223.

Ferreira, M. and Ferreira, H. (2013). Extremes of multivariate ARMAX processes. *TEST*, **22(4)**, 606 – 627.