Modelling (and forecasting) extremes in time series: A naive approach

M. Manuela Neves
Instituto Superior de Agronomia, and CEAUL, Universidade de Lisboa, Lisboa, Portugal, *manela@isa.ulisboa.pt*

Clara Cordeiro
Faculdade de Ciências e Tecnologia, Universidade do Algarve, and CEAUL, Universidade de Lisboa, Lisboa, Portugal, *ccordei@ualg.pt*

Keywords: Extreme Value Theory; Extremal index estimation; Resampling procedures; Time Series.

Abstract: In *Extreme Value Theory*, we are essentially interested in the estimation of quantities related to extreme events. Whenever the focus is in large values, estimation is usually performed based on the largest k order statistics in the sample or on the excesses over a high level u. Here we are interested in modelling (and forecasting) extremes in time series. For modelling and forecasting classical time series, Boot.EXPOS is a computational procedure built in the R environment that has revealed to perform quite well in a large number of forecasting competitions. However, to deal with extreme values, a modification of that algorithm needs to be considered and is here under study.

1 Introduction and Motivation

Time series analysis deals with records that are collected over time. The records are usually dependent, and the time order of data is important. Depending on the application, data may be collected hourly, daily, weekly, monthly, yearly, etc. Time series arise in many different contexts. Its impact on scientific, economic and social applications is well recognized by the large list of fields in which important time series problems may arise. Time series can show different displays. Let us illustrate a few time series, two of them existing in the R packages `datasets` and `fma`, see Fig.1.

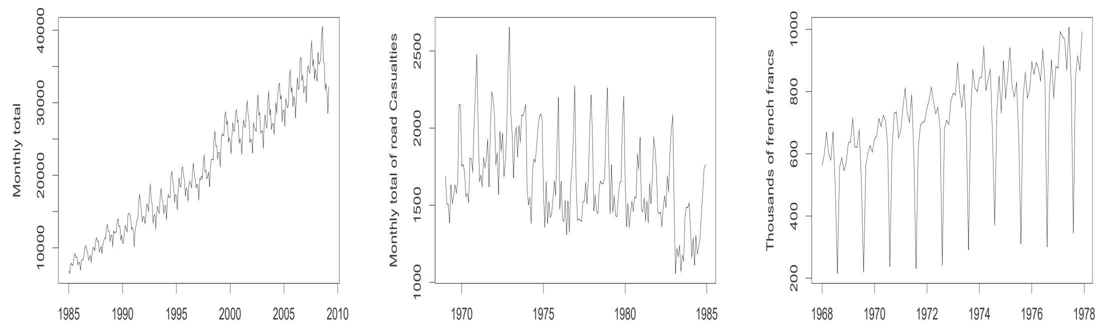


Figure 1: Number of airplanes in the FIR Lisbon(left), see [6]; Deaths and serious injuries on UK roads (center); and Sales of printing and writing paper (right).

In time series analysis, there are several challenging topics among which the treatment of extreme values has been capturing the interest of researchers. Modelling and predicting the behaviour of extreme (often maximum) values of the time series (e.g. security reasons) need special procedures.

The paper is structured as follows. In Section 2, basic results in extreme value theory both for independent and for dependent sequences are briefly reviewed. A new parameter that can appear in the limit law of the maximum of a stationary sequence, under some conditions, is described. Resampling techniques and their application together with exponential smoothing methods for modelling and prediction of a time series are reviewed in Section 3. In this section, a modification of that computational procedure, already introduced in [27] is again considered and used in extreme value theory estimation. More efficient bootstrap procedures can lead to more reliable estimates.

2 Basics in statistical analysis of univariate extremes

Statistical analysis of the extremes in time series was initially dedicated to problems in hydrology and insurance, but in the last decades the applications have spread out to a huge variety of areas, such as climatology, finance, environmental sciences (here mainly because of the direct impact in the society), etc.
The classical limiting results in Extreme Value Theory (EVT) were initially obtained through arguments that assumed an underlying process consisting of a sequence of independent and identically (i.i.d.) random variables, $(X_1,...,X_n)$, with common and unknown distribution F. Suppose we want to know the distribution of $M_n \equiv X_{n:n} := \max(X_1,...,X_n)$.
Given that $X_{n:n} \xrightarrow{\mathbb{P}} x_F =: \sup\{x \in \mathbb{R} : F(x) < 1\}$, the right endpoint of F, we are facing the situation of a degenerate distribution. First results for the existence of a non-degenerate limit for that probability date back to the beginning of the last century but were completely established by [12] and [16] that gave conditions for the existence of sequences $\{a_n\} \in \mathbb{R}^+$ and $\{b_n\} \in \mathbb{R}$ such that,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) = \lim_{n\to\infty} F^n(a_n x + b_n) = \mathrm{EV}_\xi(x), \qquad (1)$$

when $n \to \infty$ and $\forall x \in \mathbb{R}$. EV_ξ is a nondegenerate distribution function. It is called *Extreme Value* d.f., and is given by

$$\mathrm{EV}_\xi(x) = \begin{cases} \exp[-(1+\xi x)^{-1/\xi}], \ 1+\xi x > 0 & \text{if} \quad \xi \neq 0 \\ \exp[-\exp(-x)], \ x \in \mathbb{R} & \text{if} \quad \xi = 0, \end{cases}$$

where ξ, the extreme value index, is the primary parameter in extreme value theory because it measures the weight of the right tail function, $\overline{F} = 1 - F$, of the underlying model.
A function F for which the limit in (1) holds is said to be in the max-domain of attraction of EV_ξ, and we write $F \in \mathcal{D}_{\mathcal{M}}(\mathrm{EV}_\xi)$.

These models can also incorporate location (λ) and scale ($\delta > 0$) parameters, and are generally represented by

$$\mathrm{EV}_\xi(x;\lambda,\delta) \equiv \mathrm{EV}_\xi((x-\lambda)/\delta).$$

2.1 From i.i.d. to a dependent set-up

In many applications, temporal independence is unrealistic. Whenever the original scheme is no longer identically distributed, but it remains independent, those limiting results may hold true. However, when it is not possible to assume independence, we are faced with new situations. For many real problems the stationarity is the first realistic situation to be considered. In the last decades, many progresses have been made in parameter estimation of extreme values in time series, with relevance to asymptotic results. By the 1990s there was an increased interest in extremal time series, see [29, 5, 2, 4, 24], to mention a few.
Temporal dependence is common in univariate extremes of time series leading to clusters of extremes, which means that extreme values are likely to occur in temporal proximity. An excellent overview of the topic of extremal clustering is provided by [8].
As an illustration, let us consider the following sequences:

Example 2.1 *Let* $\{X_n\}$ *be a sequence of i.i.d. variables from the model* $F(x) = (1 - \exp(-x))^2$, $x \geq 0$, *and* $\{Y_n\}_{n\geq 1}$ *a two-dependent sequence defined by* $Y_n = \max(Z_{n+1}, Z_n)$, $n \geq 1$, *where* Z_n *are unit exponential i.i.d..*
We have then the underlying model for Y_n *given by* $F(y) = \mathbb{P}[Z_{n+1} \leq y, Z_n \leq y)] = (1 - \exp(-y))^2$ $y \geq 0$.

Plotting some values from $\{X_n\}$ and from $\{Y_n\}$, clusters of exceedances of high levels of size equal to 2, for the $\{Y_n\}$ sequence, can be seen, Fig.2. It can also be seen a *shrinkage* of the largest observations for the 2-dependent sequence, although we have the same model underlying both sequences.

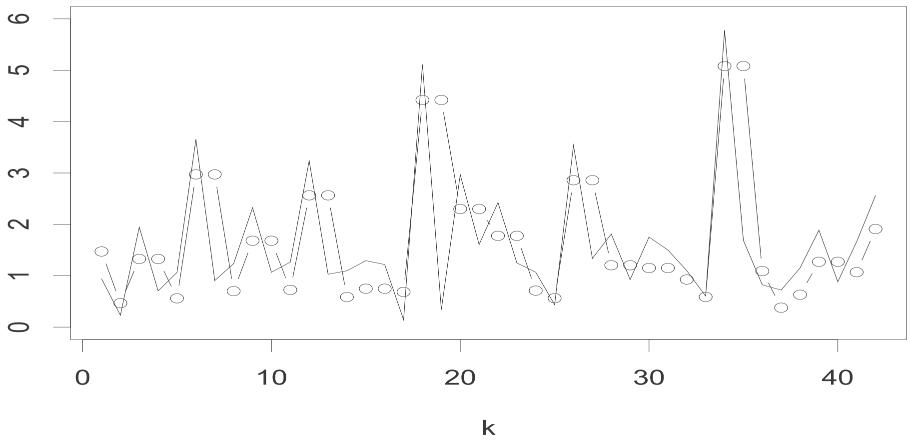


Figure 2: One realization of an i.i.d. process (solid) and a 2-dependent process (dot dash) with the same marginal d.f.

Let $\{X_n\}_{n\geq 1}$ be a stationary sequence. Under adequate conditions, the d.f. of the maximum, $X_{n:n}$, of a stationary sequence may be directly related to the maximum $Y_{n:n}$ of the associated i.i.d. sequence, through a new parameter, the so-called *extremal index.* The extremal index, θ, $0 < \theta \leq 1$, appears as

$$\mathbb{P}\left(X_{n:n} \leq x\right) \approx F^{n\theta}(x) \approx \mathrm{EV}_\xi\left(\frac{x - b'_n}{a'_n}\right) \qquad \begin{cases} a'_n = a_n \theta^\xi \\ b'_n = b_n + a_n \frac{\theta^\xi - 1}{\xi}. \end{cases}$$

In [23] conditions were established under which a stationary sequence has the same limiting EV_ξ as the associated i.i.d. sequence, but different scale and location parameters,

$$\lambda_\theta = \lambda + \delta \frac{\theta^\xi - 1}{\xi}, \qquad \delta_\theta = \delta\theta^\xi \qquad \xi_\theta = \xi,$$

where (λ, δ, ξ) are the location, scale and shape parameters of EV_ξ, respectively. A reliable estimation of θ is then required, not only by itself but because of its influence on the estimation of other parameters of interest.

2.2 The extremal index and its estimation

One common interpretation of θ is as being the reciprocal of the "mean time of duration of extreme events" which is directly related to the exceedances of high levels, see [20, 22]. Parameter θ can then be defined as $\theta = 1/(\text{limiting mean size of clusters})$.
Now, identifying clusters by the occurrence of downcrossings (or upcrossings), we can write

$$\theta = \lim_{n\to\infty} \mathbb{P}[X_2 \leq u_n | X_1 > u_n] = \lim_{n\to\infty} \mathbb{P}[X_1 \leq u_n | X_2 > u_n]$$

and the interpretation of θ has suggested the so-called Up-Crossing estimator, see [25, 10, 11], defined as:

$$\widehat{\Theta}_n^{UC} := \frac{\sum_{i=1}^{n-1} I\left(X_i \leq u_n < X_{i+1}\right)}{\sum_{i=1}^{n} I(X_i > u_n)}, \tag{2}$$

where $I(A)$ is the indicator function of A. Consistency of this estimator is obtained provided that the high level u_n is a normalized level, i.e. if with $\tau \equiv \tau_n$ fixed, the underlying distribution function F verifies

$$F(u_n) = 1 - \tau/n + o(1/n), \qquad n \to \infty \qquad \text{and} \quad \tau/n \to 0.$$

Other estimators have appeared in the literature, motivated by other forms of cluster identification, such as the blocks estimator and the runs estimator, see [18, 19, 30, 31]. Conditions for the asymptotic normality of those estimators can be seen in [19, 30, 31].
As usual in semiparametric context, the estimators considered, despite having good asymptotic properties, present high variance for high levels *vs* high bias when the level decreases, showing then a strong dependence on the high threshold u_n, for finite samples.

3 Resampling procedures

Resampling computer-intensive methodologies, like the generalised jackknife, [15], and the bootstrap, [9], have been revealing themselves as important tools for a reliable semi-parametric estimation of parameters of extreme events. Let us briefly see the application of those methodologies in the θ estimation.

3.1 The Generalized Jackknife methodology

By using generalized jackknife methodology, [13] proposed a reduced-bias Generalized Jackknife estimator of order 2, $\widehat{\Theta}^{GJ}$, based on the estimator $\widehat{\Theta}^{UC}$ computed at three levels: k, $\lfloor k/2 \rfloor+1$ and $\lfloor k/4 \rfloor+1$, ($\lfloor x \rfloor$ – integer part of x), defined as

$$\widehat{\Theta}^{GJ}(k) := 5\widehat{\Theta}^{UC}(\lfloor k/2 \rfloor + 1) - 2\big(\widehat{\Theta}^{UC}(\lfloor k/4 \rfloor + 1) + \widehat{\Theta}^{UC}(k)\big). \quad (3)$$

More generally [13] considered the levels k, $\lfloor \delta k \rfloor + 1$ and $\lfloor \delta^2 k \rfloor + 1$, depending on the *tuning parameter* δ, $0 < \delta < 1$, and got then a class of estimators. Actually $\widehat{\Theta}^{GJ}$, in (3), is obtained with $\delta = 1/2$. This estimator illustrates the simulation study performed with the "Max-Autoregressive Process (ARMAX process)", see [2].

Example 3.1 *Let $\{Z_i\}_{i\geq 1}$ be a sequence of independent, unit-Fréchet distributed random variables. For $0 < \theta \leq 1$, let*

$$X_1 = Z_1 \quad X_i = \max\{(1-\theta)X_{i-1}, \theta Z_i\} \quad i \geq 2.$$

For $u_n = nx$, $0 < x < \infty$, $\mathbb{P}\big\{M_n \leq u_n\big\} \to \exp\big(-\theta/x\big)$, as $n \to \infty$, being θ the extremal index of the sequence.

The reduced-bias estimator in (3) outperforms the associated classical estimator. However, for a given sample, the choice of the number of upper order statistics to be used is a difficulty not yet solved. See Figure 3, where three different samples were generated, considering three different values for the parameter, in an ARMAX model. The estimates paths show how difficult it is to choose k and to obtain a reliable estimate of θ.

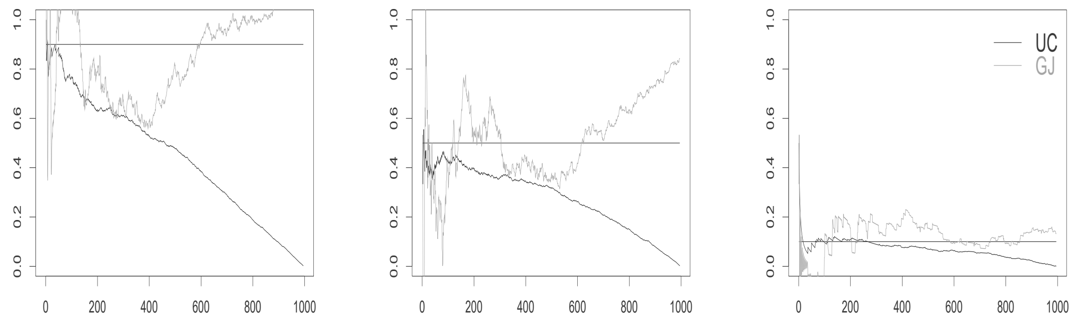


Figure 3: One sample path for UC and GJ estimates in the ARMAX model for three simulated samples with $\theta = 0.9, 0.5, 0.1$ (from the left to the right).

3.2 The bootstrap under dependence

For modelling and forecasting time series [6, 7] developed a computational procedure, built in the R environment, based on Exponential Smoothing Methods jointly with "adequate" bootstrap procedures. When applied to a large set of time series, competitive results were obtained compared with the best procedures available, see [7].

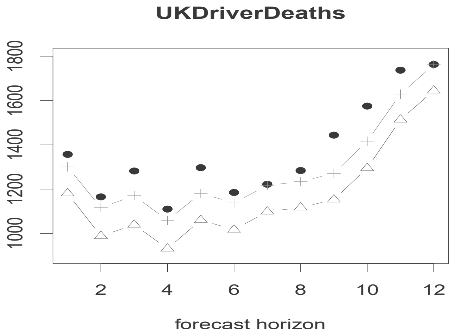
So the main motivation of this work is to explore and to modify that automatic procedure in order it can be an alternative for modelling and (forecasting) extreme values in time series. Preliminary results have been presented in [27] and are used here in the θ estimation.

The aforementioned computational procedure, for modelling and forecasting time series, chooses among a set of models, that one that best fits the data. Sieve bootstrap principle is applied to the residuals; an autoregressive model with increasing order is fitted to the residuals; stationarity is tested; transformations or differentiations are performed when necessary, and after bootstrapping the second level of residuals a bootstrap estimated series is obtained. Forecast is performed based on the bootstrap estimated values and on the model parameters estimated at the initial step. Measures of forecast errors are also included in the algorithm. A description and sketch of the algorithm is presented in [6, 7, 26] among others.

Fig.4 illustrates the result of forecasting twelve months applying Boot.EXPOS and `ets`[7] [21], using the dataset `UKDriverDeaths` available in R. The good

[7]Stands for error, trend and seasonality.

performance of the Boot.EXPOS procedure is clearly illustrated both for point forecast values and for forecasting intervals.



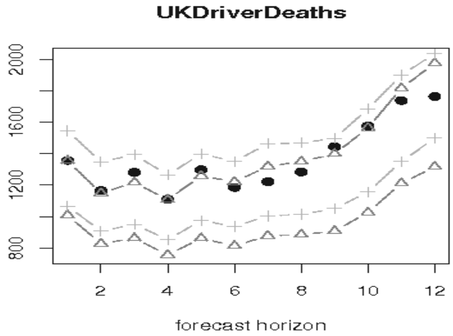


Figure 4: True values (•) compared with Boot.EXPOS values (– + –) and `ets` values (– △ -).

3.3 Modelling time series extremes

The classical bootstrap does not work in a dependent context. This was referred to [3] and later in [1], who showed that in extreme value theory the bootstrap version for the maximum (or minimum) does not converge to the extremal limit laws. Actually, [32] pointed out *"... to resample the data for approximating the distribution of the k largest observations would not work because the "pseudo-samples" would never have values greater than $X_{n:n}$".* A bootstrap method considering to resample a smaller size than the original sample was proposed in [17] for estimating mean squared error and smoothing parameter in nonparametric problems. The idea in [17] was to choose the resample size, n_1, to be less than the original sample size, n, and use knowledge of the amount by which the two samples differ to estimate mean squared error and to select the optimal smoothing parameter for deriving a bootstrap estimator of a functional of $(X_1, \ldots, X_n)$. He suggested resampling a subsample of size $n_1 = O(n^{1-\epsilon})$ with $0 < \epsilon < 1$. The procedure developed in [17] was illustrated for nonparametric density estimation, nonparametric regression and tail parameter estimation. In this latter case, the tail parameter estimators in a semi-parametric approach need an adequate choice of the number, k, of upper order statistics, that should be chosen such that the asymptotic mean squared error of the estimator is minimized. The [17] bootstrap procedure suggests the following: to draw a resample of size n_1 from de original sample of size n, to obtain the bootstrap estimate of the mean squared error of the estimator considered, let us denote it as $\widehat{\mathrm{MSE}}(n_1,k_1)$, where k_1 are the upper order statistics of the n_1-sized resample. Supposing that the asymptotically optimal k is of the form Cn^γ, where $0<\gamma<1$ is a known constant and C is unknown, what is a common result, [17] proposed, for a given class of models, if the optimal k_1 is asymptotic to Cn_1^γ, then

$$\hat{k}_0 \simeq \hat{k}_{1,0}(n/n_1)^\gamma, \tag{4}$$

is asymptotic to Cn^γ. For several models, [17] showed that $\gamma = 2/3$. This idea was exploited in a very preliminary study in [27], where the funcional under study was the *maximum*, taking advantage of the good performance of Boot.EXPOS for modelling and forecasting time series. A subsample of size $n_1 = \lfloor n^{0.995} \rfloor$ of the residuals in the algorithm was considered. Values of the resampled series were then "improved" on basis of the relation (4) – this is now called *Boot.EXPOS with subsampling*. See Fig.5 as an illustration.

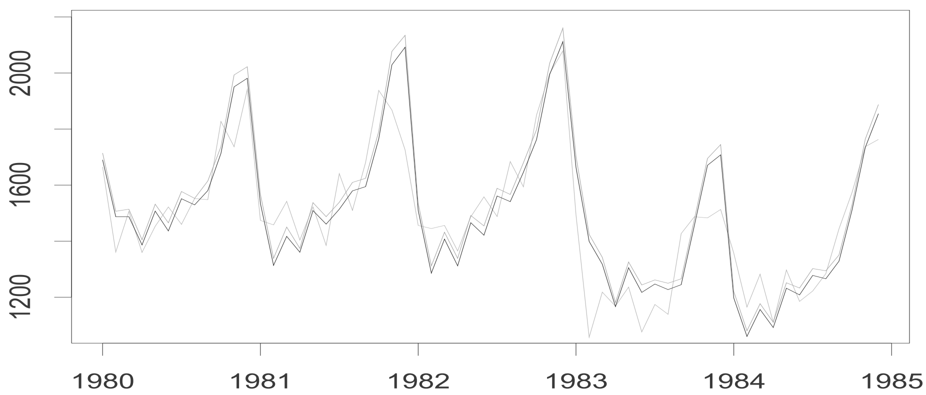


Figure 5: Subset of observed UKDriverDeaths values (solid grey) and forecasts obtained using Boot.EXPOS (dashed), Boot.EXPOS with subsampling (dotted).

The *Boot.EXPOS with subsampling*, was applied to a simulated data set and to a real data set, the `UKDriverDeaths` time series. The interest is to estimate θ. Figures 6 and 7 show sample paths for the θ-estimator, $\widehat{\Theta}_n^{UC}$, in (2), and $\widehat{\Theta}^{GJ}$, in (3), and the associated bootstrap estimates calculated using *Boot.EXPOS with subsampling*, $\widehat{\Theta}_n^{UC^*}$ and $\widehat{\Theta}^{GJ^*}$, respectively.

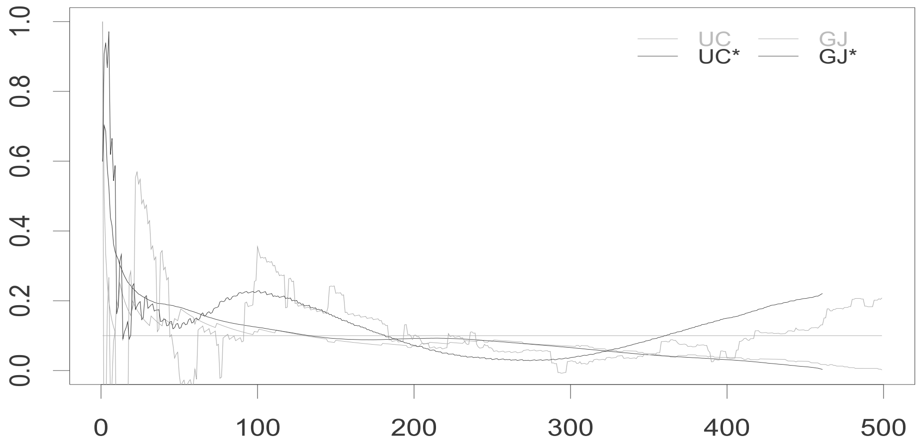


Figure 6: UC and GJ θ-estimates in an ARMAX process with $\theta = 0.1$. UC* and GJ* θ-estimates using Boot.EXPOS with subsampling.

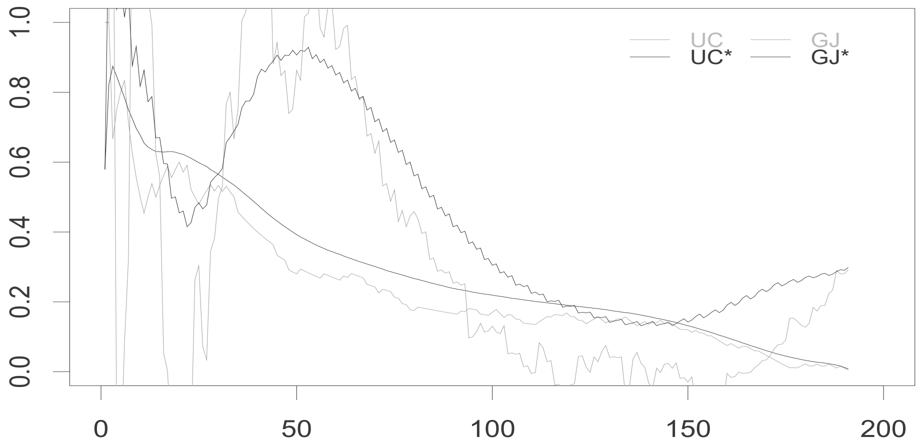


Figure 7: UC and GJ θ-estimates in the UKDriverDeaths time series and the associated UC* and GJ* θ-estimates using Boot.EXPOS with subsampling.

4 A brief discussion

The procedure here proposed and based on [17] results seems to be a promising bootstrap approach for modelling and forecasting extremes, providing more stable paths to the parameters estimates. Other values for the θ parameter in the ARMAX process have been considered, leading to similar results, not shown for reasons of space. More research needs to be performed. A large simulation study is now in progress.

Acknowledgements

We would like to thank the two anonymous referees for their valuable and helpful comments, that highly improve this version of the manuscript. Research partially supported by National Funds through Fundação para a Ciência e a Tecnologia, Portugal, throught the project FCT Portugal UID/MAT/00006/2019. The authors' thanks to **Portugal Navigation-NAV Portugal, E.P.E.** for providing the data.

References

[1] Angus, J. (1993). Asymptotic theory for bootstrapping the extremes. *Communication in Statistics Theory and Methodology*, 22(1), 15–30.

[2] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J.L. (2004). *Statistics of Extremes. Theory and Applications*. England, John Wiley & Sons.

[3] Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Annals of Statistics*, 9, 1196-1217.

[4] Chavez-Demoulin, V. and Davison, A.C. (2012). Modelling Time Series Extremes. *Revstat-Statistical Journal*, 10:1, 109–133.

[5] Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, New York.

[6] Cordeiro C. and Neves M. (2009). Forecasting time series with Boot.EXPOS procedure. *Revstat*, 7:2, 135–149.

[7] Cordeiro, C. and Neves M.(2010). Boot.EXPOS in NNGC competition): *Proceedings of the IEEE World Congress on Computational Intelligence (WCCI 2010)*, 1135–1141.

[8] Davis, R.A., Mikosch, T. and Zhao, Y. (2013). Measures of serial extremal dependence and their estimation. *Stochastic Processes and Their Applications* 123: 7, 2575–2602

[9] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics* 7:1–26.

[10] Gomes, M.I. (1990). Statistical inference in an extremal markovian model. *COMPSTAT*, 257–262.

[11] Gomes, M.I. (1992). Modelos extremais em esquemas de dependência. *I Congresso Ibero-Americano de Esdadistica e Investigacion Operativa*, 209–220.

[12] Gnedenko, B.V. (1943). Sur la distribution limite d'une série aléatoire. *Annals of Mathematics* 44, 423–453.

[13] Gomes, M.I., Hall, A. and Miranda, C. (2008). Subsampling techniques and the Jackknife methodology in the estimation of the extremal index. *Computational Statistics and Data Analysis* 52:4, 2022–2041.

[14] Gomes, M.I., Figueiredo, F. and Neves, M.M. (2011). Adaptive estimation of heavy right tails: the bootstrap methodology in action. *Extremes*, DOI: 10.1007/s10687-011-0146-6

[15] Gray, H.L., and Schucany, W.R. (1972). *The Generalized Jackknife Statistic*. Marcel Dekker.

[16] de Haan, L. (1970). *On regular variation and its application to the weak convergence of sample extremes*, Amsterdam, Mathematisch Centrum.

[17] Hall P. (1970). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of Multivariate Analysis*, 32, 177–203.

[18] Hsing, T. (1991). Estimating the parameters of rare events. *Stochastic Processes and Their Applications*, 37:117–139.

[19] Hsing, T. (1993). Extremal index estimation for a weakly dependent stationary sequence. *The Annals of Statistics*, 21:2043–2071.

[20] Hsing, J. T., Hüsler, J. and Leadbetter, M. R. (1988). On the exceedance point process for a stationary sequence. *Probability Theory and Related Fields*, 78, 1: 97–112.

[21] Hyndman, R., A. Koehler, J. Ord and R. Snyder (2008). *Forecasting with Exponential Smoothing: The State Space Approach*. Springer-Verlag.

[22] Leadbetter, M.R. and Nandagopalan, L. (1989). On exceedance point process for stationary sequences under mild oscillation restrictions. In *Extreme Value Theory*): *Proceedings, Oberwolfach 1987, J. Hüsler and R.D. Reiss (eds.)*, Lecture Notes in Statistics 52, 69–80. Springer-Verlag, Berlim.

[23] Leadbetter, M., Lindgren, G. and Rootzen, H. (1983). *Extremes and related properties of random sequences and series.* Springer-Verlag, New York.

[24] McElroy, T. (2016). On the measurement and treatment of extremes in time series. *Extremes*, 16:3, 467–490

[25] Nandagopalan, S. (1990). *Multivariate Extremes and Estimation of the Extremal Index*. PhD Thesis, University of North Carolina, Chapel Hill.

[26] Neves, M. and Cordeiro, C. (2011). Exponential smoothing and resampling techniques in time series prediction. *Discussiones Matematicae, Probability and Statistics*, 30, 87–101.

[27] Neves, M. M. and Cordeiro, C. (2014). *Statistical modelling in time series extremes: an overview and new steps.* In Gilli, M., Gonzalez-Rodriguez, G. and Nieto-Reyes, A. (eds.).): *Proceedings of COMPSTAT 2014.*

[28] R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[29] Reiss, R-Dieter, Thomas, M. (1997). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields.* Birkhauser.

[30] Smith, R. and Weissman, I. (1994). Estimating the extremal index. *Journal of the Royal Statistical Society B* 56:515–528.

[31] Weissman, I. and Novak, S. (1998). On blocks and runs estimators of the extremal index. *Journal of Statistical Planning and Inference* 66:281–288.

[32] Zelterman D.(1993, 2012). A Semiparametric Bootstrap Technique for Simulating Extreme Order Statistics, *Journal of the American Statistical Association*, 477–485.