Rao-Blackwellised estimators for parameters of extreme-value models

By Christopher A. T. Ferro

Centre for Global Atmospheric Modelling, Department of Meteorology, University of Reading, Earley Gate, P.O. Box 243, Reading, RG6 6BB, UK

c.a.t.ferro@reading.ac.uk

and Sergio Pezzulli

School of Applied Statistics, Harry Pitt Building, University of Reading, Whiteknights Road, P.O. Box 240, Reading, RG6 6FN, UK

s.pezzulli@reading.ac.uk

Summary

Fitting extreme-value models to a single set of block maxima formed by an arbitrary partition of a time series is inefficient if the complete series is available. More efficient probability-weighted moment and maximum-likelihood estimators that average over sets of admissable partitions are proposed for parameters of the generalised extreme-value distribution and for the extremal index. Simulation studies illustrate efficiency gains for typical sample sizes.

Some key words: Block maxima; Extremal index; Generalised extreme-value distribution; Probability-weighted moments.

1 Introduction

The generalised extreme-value (GEV) distribution is a fundamental model in extreme-value theory (Beirlant et al., 2004; Coles, 2001; Leadbetter et al., 1983 for example). It represents all possible, non-degenerate limit distributions for the linearly normalised maximum value M_n among independent and identically distributed random variables $X_1, \ldots, X_n$. That is, if there exist sequences of constants a_n and $b_n > 0$ such that $(M_n - a_n)/b_n$ has a non-degenerate limit distribution as n increases then the limit is the GEV distribution

$$G(x \mid \phi) = \exp\left[-\left\{1 + \gamma\left(\frac{x-\alpha}{\beta}\right)\right\}_{+}^{-1/\gamma}\right] \tag{1}$$

with location, scale and shape parameters $\phi = (\alpha, \beta, \gamma)$. The sign of the shape parameter identifies three sub-classes: Weibull ($\gamma < 0$), Gumbel ($\gamma = 0$, defined by continuity) and Fréchet ($\gamma > 0$). The limit also holds for maxima from a large class of stationary sequences, justifying its efficacy in many applications.

Traditional inference fits the three-parameter GEV model to a recorded sequence of block, say annual, maxima. This approach was developed for situations in which only block maxima, rather than complete time series, were recorded and at a time when computing power was low. The approach has remained popular even when complete time series are available, in which case block maxima are extracted from a single partition of the data. This is inefficient, however, if other partitions lead to equivalent estimators. This paper shows that averaging over admissable partitions yields estimators that have the same bias and smaller variance than the traditional estimators. The argument is similar to the Rao-Blackwell Theorem (Silvey, 1991, p. 28 for example), so the new estimators are referred to as Rao-Blackwellised (RB) estimators, and are constructed for both maximum-likelihood (ML) and probability-weighted moment (PWM) estimators for the parameters of the GEV distribution and for related quantities. Sequences of independent random variables are considered in §2; stationary sequences are considered in §3. RB estimators for the extremal index, which measures the clustering of extreme values, are also defined in §3. Simulations show that efficiency gains are practically as well as theoretically significant. The paper concludes with a discussion in §4.

2 Independent sequences

2·1 *Parameter estimators*

Let $\mathcal{X} = \{X_1, \dots, X_n\}$ be a sequence of independent and identically distributed random variables, and let $M_j = \max\{X_i : (j-1)m + 1 \leq i \leq jm\}$, $1 \leq j \leq k = \lfloor n/m \rfloor$, be the block maxima for blocks of length $m \geq 1$. Assume that the block maxima $\mathcal{M} = \{M_1, \dots, M_k\}$ follow the GEV distribution (1). Two common ways to estimate the parameters ϕ are by maximum likelihood (Prescott & Walden, 1980; Smith, 1985) and probability-weighted moments (Hosking et al., 1985).

The standard ML estimator, $\hat{\phi}_L$, for ϕ maximises the likelihood

$$L(\phi \mid \mathcal{M}) = \prod_{j=1}^{k} g(M_j \mid \phi),$$

where g is the density of the GEV distribution function (1). Now let $\mathcal{X}_i$, $1 \leq i \leq N = n!$, be the distinct permutations of $\mathcal{X}$ with associated maxima $\mathcal{M}_i$, and denote by $\hat{\phi}_{L,i} = (\hat{\alpha}_i, \hat{\beta}_i, \hat{\gamma}_i)$ the estimator that maximises $L(\phi \mid \mathcal{M}_i)$. Each $\hat{\phi}_{L,i}$ has the same distribution as $\hat{\phi}_L$, so the RB estimator

$$\bar{\phi}_L = (\bar{\alpha}_L, \bar{\beta}_L, \bar{\gamma}_L) = \frac{1}{N} \sum_{i=1}^{N} \hat{\phi}_{L,i} \tag{2}$$

satisfies, with variances and covariances taken elementwise,

$$\begin{aligned} E(\bar{\phi}_L) &= E(\hat{\phi}_L), \\ \operatorname{var}(\bar{\phi}_L) &= N^{-1}\operatorname{var}(\hat{\phi}_L) + N^{-2} \textstyle\sum_{i \neq j} \operatorname{cov}(\hat{\phi}_{L,i}, \hat{\phi}_{L,j}) \leq \operatorname{var}(\hat{\phi}_L). \end{aligned} \tag{3}$$

This remains true if $N \leq n!$ permutations are selected randomly with replacement. The connection with the Rao-Blackwell Theorem arises because $\bar{\phi}_L$ is the expectation of the estimator that equals $\hat{\phi}_{L,i}$ with probability $1/N$ conditional on the sufficient statistic $\mathcal{X}$. Segers (2001, Catholic University Leuven Ph.D. Thesis, Ch. 5) constructs U-statistics for γ with a similar averaging procedure.

The PWMs of a random variable X with GEV distribution (1) are, for non-negative integers r,

$$\mu_r = E[X\{G(X)\}^r] = (r+1)^{-1}[\alpha - \beta\{1-(r+1)^\gamma\Gamma(1-\gamma)\}/\gamma]$$

when $\gamma < 1$ (Hosking et al. 1985). Hosking et al. (1985) use the estimators

$$\hat{\mu}_r = \frac{1}{k}\sum_{j=1}^{k}\left(\frac{j-0.35}{k}\right)^r M_{(j)}, \tag{4}$$

where $M_{(1)} \leq \ldots \leq M_{(k)}$ are the ordered maxima, and solving $\mu_r = \hat{\mu}_r$ simultaneously for $r = 0$, 1 and 2 yields the PWM estimator $\hat{\phi}_M$ for ϕ. The RB version of $\hat{\mu}_r$ is

$$\bar{\mu}_r = \frac{1}{n!}\sum_{i=1}^{n!}\hat{\mu}_{r,i} = \frac{1}{n!k}\sum_{i=1}^{n}\sum_{j=1}^{k}\left(\frac{j-0.35}{k}\right)^r m_{ij}X_{(i)},$$

where $\hat{\mu}_{r,i}$ is the estimator (4) applied to permutation $\mathcal{X}_i$ and m_{ij} is the number of distinct permutations among all $n!$ permutations in which $X_{(i)}$ is the j-th smallest of the k block maxima. An estimator $\bar{\phi}^*_M = (\bar{\alpha}^*_M, \bar{\beta}^*_M, \bar{\gamma}^*_M)$ for ϕ follows by solving $\mu_r = \bar{\mu}_r$ simultaneously for $r = 0$, 1 and 2. An expression for m_{ij} is derived in a technical report available from the authors but its evaluation appears to be prohibitively expensive; a small number N of randomly selected permutations of $\mathcal{X}$ can be used instead:

$$\bar{\mu}_r = \frac{1}{N}\sum_{i=1}^{N}\hat{\mu}_{r,i}.$$

Direct Rao-Blackwellisation of the PWM estimator $\hat{\phi}_M$ yields the alternative average over N randomly selected permutations:

$$\bar{\phi}_M = \frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_{M,i}, \tag{5}$$

where $\hat{\phi}_{M,i}$ is the PWM estimator for permutation $\mathcal{X}_i$.

The standard estimator for the p-quantile of the GEV distribution (1) derived from the ML estimator $\hat{\phi}_L = (\hat{\alpha}_L, \hat{\beta}_L, \hat{\gamma}_L)$ for ϕ is

$$\hat{q}_L = \hat{\alpha}_L + \hat{\beta}_L\{(-\log p)^{-\hat{\gamma}_L} - 1\}/\hat{\gamma}_L. \tag{6}$$

Rao-Blackwellisation supports two more estimators: the RB estimator,

$$\bar{q}_L = \frac{1}{N}\sum_{i=1}^{N}\hat{q}_{L,i},$$

where $\hat{q}_{L,i} = \hat{\alpha}_i + \hat{\beta}_i\{(-\log p)^{-\hat{\gamma}_i} - 1\}/\hat{\gamma}_i$, or the quantile,

$$\tilde{q}_L = \bar{\alpha}_L + \bar{\beta}_L\{(-\log p)^{-\bar{\gamma}_L} - 1\}/\bar{\gamma}_L,$$

consistent with the RB parameter estimator $\bar{\phi}_L$. Quantile estimators, $\bar{q}_M$ and $\tilde{q}_M$, derived from PWM estimators are defined similarly.

Remark 1. Efficiency properties similar to inequality (3) hold for the RB PWM and quantile estimators.

Remark 2. These RB estimators permit contributions from all bar the smallest $m-1$ data instead of only k arbitrarily constructed block maxima. This qualifies the notion that only 'extremes' contain information relevant for tail inference. Indeed, the empirical distribution function of the block maxima from all permutations can be written as a function of the order statistics, providing a non-parametric estimator for the limiting GEV distribution.

2·2 *Simulation study*

The small-sample performances of standard and RB estimators for GEV parameters and quantiles are compared. Sequences of $n = km$ independent random variables are simulated from a marginal GEV distribution F with parameters chosen so that $F^m = G$, the GEV distribution (1) with parameters α, β and γ. These parameters and the upper 1% quantile q of G are estimated using block size m and $N = 100$ randomly selected permutations. Larger values of N confer meagre improvements. The location and scale parameters of G are set equal to $\alpha = 0$ and $\beta = 1$, and five values are considered for the shape parameter: $\gamma = -0.4$, -0.2, 0, 0.2 and 0.4. Several block and sample sizes were investigated but results are shown only for $m = 100$ and

$k = 40$; results for other choices are commented on below. Likelihoods are maximised numerically using the function `fgev` in the `evd` package (Stephenson, 2002) of `R` (Ihaka & Gentleman, 1996). Monte Carlo approximations are based on 1000 simulations.

Variances of the RB PWM estimators $\bar{\phi}_M$ for ϕ and $\bar{q}_M$ for q are compared to those of the standard PWM estimators $\hat{\phi}_M$ and $\hat{q}_M$ in Figure 1. For $k = 40$ blocks of size $m = 100$, significant efficiency gains are found in all cases: relative efficiencies of standard to RB PWM estimators increase with the shape parameter, from about 65 to 75% for the location parameter, 30 to 85% for scale, 45 to 95% for shape, and 80 to 100% for the quantile. Alternative RB PWM estimators $\bar{\phi}^*_M$ and $\tilde{q}_M$ perform similarly to $\bar{\phi}_M$ and $\bar{q}_M$. The relative efficiencies were found to decrease as m increases, but were approximately stable for $m > 10$. Similar conclusions hold for other sample sizes k. Relative efficiencies were also found to change as k increases when m is fixed, but were approximately stable for $k > 20$.

Variances of the RB ML estimators $\bar{\phi}_L$ for ϕ and $\bar{q}_L$ for q are also compared to those of the standard ML estimators $\hat{\phi}_L$ and $\hat{q}_L$ in Figure 1. The results for the quantile estimators are obtained by optimising the likelihood for the GEV distribution with parameterisation (q, β, γ). This yields more robust estimates of q than relationship (6) for the optimisation routine used. The performance of $\bar{q}_L$ under the reparameterisation is similar to the performance of $\tilde{q}_L$ defined by $\bar{\phi}_L$. Significant efficiency gains are again found in all cases, although, unlike PWM estimators, efficiencies of standard relative to RB ML estimators do not always increase with the shape parameter. For $k = 40$ blocks of size $m = 100$, efficiencies decrease from about 70 to 65% for the location parameter, increase from 30 to 70% for scale, are consistently about 70% for shape, and decrease from 95 to 75% for the quantile. Again, the relative efficiencies were found to decrease as m increases, stabilising for $m > 10$, and also changed with k when m is fixed, stabilising for $k > 20$.

3 Stationary sequences

3·1 *Parameter estimators*

Now let $\mathcal{X}$ be a stationary sequence, and let the limit distribution of block maxima from the independent sequence with the same marginal distribution as $\mathcal{X}$ be GEV with parameters $\phi = (\alpha, \beta, \gamma)$. Under weak mixing conditions, the limit distribution of block maxima from $\mathcal{X}$ is also GEV but with parameters $\phi' = (\alpha', \beta', \gamma')$ related to ϕ by

$$\alpha' = \alpha + \beta(\theta^\gamma - 1)/\gamma, \qquad \beta' = \beta\theta^\gamma, \qquad \gamma' = \gamma, \tag{7}$$

where $\theta \in (0, 1]$ is the extremal index (Leadbetter, 1983). The extremal index has several uses: relating the distribution of block maxima to the marginal distribution of the sequence; describing the strength of clustering at extreme levels of the process; and helping to identify clusters of extreme values (Beirlant et al., 2004, Ch. 10). Estimators for θ are described later; estimators for ϕ' are considered first.

When RB estimators (2) or (5) are applied to stationary sequences for N randomly selected permutations they yield estimates of ϕ. Permutations must preserve the serial order of the data within blocks if they are to yield estimators that have the same distribution as standard estimators for ϕ'. If $n = km + c$ with $1 \leq c \leq m - 1$ then there are $\binom{k+c}{c}$ distinct ways to form k blocks of m consecutive data, and averaging the estimates of ϕ' obtained from these partitions improves efficiency, but if $c = 0$ there is only one such partition and no improvement is made. A scheme that enables efficiency gains for any value of c averages over the n cycles $\mathcal{X}'_1 = \mathcal{X}$ and $\mathcal{X}'_j = \{X_j \ldots, X_n, X_1, \ldots, X_{j-1}\}$, $2 \leq j \leq n$. This alternative is adopted to define the RB estimator $\bar{\phi}' = \sum_{j=1}^{n} \hat{\phi}'_j/n$, where $\hat{\phi}'_j = (\hat{\alpha}'_j, \hat{\beta}'_j, \hat{\gamma}'_j)$ maximises $L(\phi' \mid \mathcal{M}'_j)$ and $\mathcal{M}'_j$ are the maxima associated with $\mathcal{X}'_j$. Some cycles contain a block of non-consecutive data, with a break between X_n and X_1, and the estimators $\hat{\phi}'_j$ from these cycles have different properties to $\hat{\phi}'_1$. Equalities (3) therefore hold only approximately, but the influence of the rogue blocks is negligible for large n and k. An alternative scheme not considered here averages over the $\binom{k-1+m+c}{m+c}$ ways to form $k - 1$ blocks.

The relationships (7) led Gomes (1993) to the estimator

$$\hat{\theta}_G = (\hat{\beta}_1'/\hat{\beta}_1)^{(\hat{\alpha}_1'-\hat{\alpha}_1)/(\hat{\beta}_1'-\hat{\beta}_1)}$$

for the extremal index. The exponent estimates $1/\gamma$ and makes $\hat{\theta}_G$ much less sensitive to γ, particularly when $\gamma \approx 0$, than the more obvious $1/\hat{\gamma}_1$. Ancona-Navarrete & Tawn (2000) proposed $\hat{\theta}_A = \hat{\theta}_{11}$, where $(\hat{\phi}_{ij}, \hat{\theta}_{ij})$ maximises $L(\phi \mid \mathcal{X}_i)L(\phi' \mid \mathcal{X}_j')$ subject to constraints (7) and $0 < \theta \leq 1$.

Estimators $\hat{\theta}_G$ and $\hat{\theta}_A$ can be improved by averaging over n cycles and N random permutations:

$$\begin{aligned} \bar{\theta}_G &= \frac{1}{nN}\sum_{j=1}^{n}\sum_{i=1}^{N}(\hat{\beta}_j'/\hat{\beta}_i)^{(\hat{\alpha}_j'-\hat{\alpha}_i)/(\hat{\beta}_j'-\hat{\beta}_i)}, \\ \bar{\theta}_A &= \frac{1}{nN}\sum_{j=1}^{n}\sum_{i=1}^{N}\hat{\theta}_{ij}. \end{aligned} \tag{8}$$

Several other RB estimators for θ are derived in a technical report available from the authors.

3·2 *Simulation study*

The performance of $\bar{\phi}'$ is evaluated by repeating the simulation study of §2 with data generated from a max-autoregressive process

$$X_i = \max\{(1-\theta)X_{i-1}, \theta Z_i\}, \quad i \geq 1, \tag{9}$$

where the Z_i are independent unit-Fréchet random variables. This is one of the processes used by Ancona-Navarrete & Tawn (2000). The extremal index is $\theta = 0.5$ and the data are transformed to have GEV marginal distribution (1). With $n = km$, only the first m cycles are needed to compute $\bar{\phi}'$ because $\mathcal{M}'_{j+lm}$ is then a permutation of $\mathcal{M}'_j$ for all $1 \leq j \leq m$ and $1 \leq l \leq k-1$. The amounts by which the relative efficiencies fall below 100% (not shown) are approximately halved compared to the independence case.

The performances of the estimators $\bar{\theta}_G$ and $\bar{\theta}_A$ for the extremal index are assessed in another simulation study. Sequences of $n = km$ random variables are simulated from max-autoregressive processes (9) and transformed to have GEV marginal distribution (1). Four values are considered for the extremal index: $\theta = 0.25$, 0.5, 0.75 and 1. The marginal parameters are chosen to be $\alpha = 0$, $\beta = 1$ and $\gamma = 0$. The number of blocks k and the block length m are both set equal to 20, in line with Gomes (1993) and Ancona-Navarrete & Tawn (2000). One random permutation is constructed for each of the m cycles needed when $n = km$ and $i = j$ is fixed in definition (8) so that $\bar{\theta}_G$ and $\bar{\theta}_A$ require $2m$ and m optimisations respectively. Monte Carlo approximations to the biases and variances of the estimators based on 500 simulations are plotted in Figure 2.

Estimator $\bar{\theta}_G$ has uniformly smaller bias but larger variance and mean square error than $\bar{\theta}_A$. This is unsurprising as $\bar{\theta}_G$ is formed from unconstrained estimators for the two sets of GEV parameters. If $\bar{\theta}_G$ is truncated at 1, its mean square error is smaller than that of $\bar{\theta}_A$ when $\theta = 1$. The efficiency (not shown) of $\hat{\theta}_G$ relative to $\bar{\theta}_G$ decreases from about 50 to 25% and the efficiency of $\hat{\theta}_A$ relative to $\bar{\theta}_A$ decreases from about 70 to 40% as θ increases. Similar results were found for other values of the marginal shape parameter.

4 Discussion

Rao-Blackwellised estimators for the parameters and quantiles of the GEV distribution and for the extremal index have been introduced and shown to have significant efficiency gains over standard estimators based on block maxima. The method extends to the r-largest maxima model (Tawn, 1988a) and to multivariate extreme-value models based on componentwise block maxima (Tawn, 1988b). Further work is required to determine how to estimate standard errors. One possible, computationally expensive approach uses the non-parametric bootstrap.

Point-process models for exceedances of high thresholds provide estimators for parameters of extreme-value models that can not be improved by Rao-Blackwellisation. See Davison & Smith (1990) and Ferro & Segers (2003) for example. These methods have been considered superior to methods based on block maxima in terms of mean square error (Joe, 1987; Smith, 1987, unpublished manuscript), but thresholding neglects a large portion of the data which the estimators developed in this paper do not. More work is needed to understand the relative merits of the two approaches.

References

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

Beirlant, J., Goegebeur, Y., Segers, J. & Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. London: Wiley.

Coles, S.G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. London: Springer-Verlag.

Davison, A.C. & Smith, R.L. (1990). Models for exceedances over high thresholds. *J. R. Statist. Soc.* B **52**, 393–442.

Ferro, C.A.T. & Segers, J. (2003). Inference for clusters of extreme values. *J. R. Statist. Soc.* B **65**, 545–556.

Gomes, M.I. (1993). On the estimation of parameters of rare events in environmental time series. In *Statistics for the Environment*, Ed. V. Barnett and K.F. Turkman, pp. 225–241. Chichester: Wiley.

Hosking, J.R.M., Wallis, J.R. & Wood, E.F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* **27**, 251–261.

Ihaka, R. & Gentleman, R. (1996). R: A language for data analysis and graphics. *J. Comp. Graph. Stat.* **5**, 299–314.

Joe, H. (1987). Estimation of quantiles of the maximum of N observations. *Biometrika* **74**, 347–354.

Leadbetter, M.R. (1983). Extremes and local dependence in stationary sequences. *Z. Wahr. Ver. Geb.* **65**, 291–306.

Leadbetter, M.R., Lindgren, G. & Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Series*. New York: Springer-Verlag.

Prescott, P. & Walden, A.T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika* **67**, 723–724.

Silvey, S.D. (1991). *Statistical Inference*. London: Chapman and Hall.

Smith, R.L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72**, 67–90.

Stephenson, A.G. (2002). evd: extreme value distributions. *R News* **2** (2), 31–32. URL `http://CRAN.R-project.org/doc/Rnews/`

Tawn, J.A. (1988a). An extreme value theory model for dependent observations. *J. Hydrology* **101**, 227–250.

Tawn, J.A. (1988b). Bivariate extreme value theory: models and estimation. *Biometrika* **75**, 397–415.

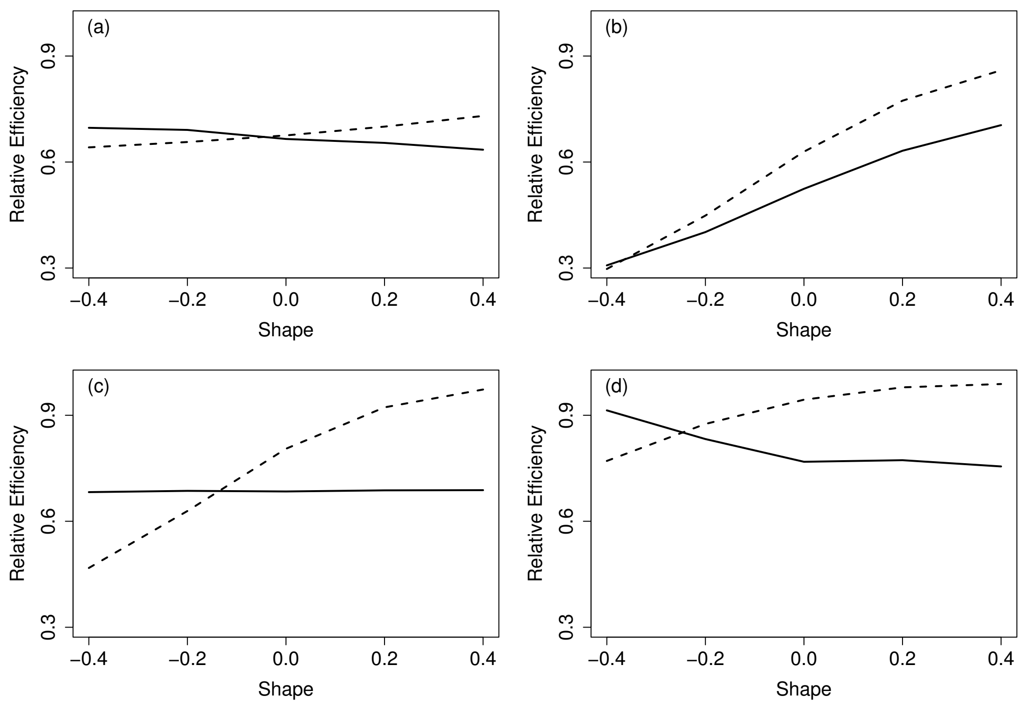

Fig. 1: Estimated relative efficiencies of standard to RB PWM (-----) and ML (——) estimators for the (a) location, (b) scale and (c) shape parameters and (d) upper 1% quantile of the GEV distribution based on $k = 40$ blocks of length $m = 100$ plotted against the true GEV shape parameter.

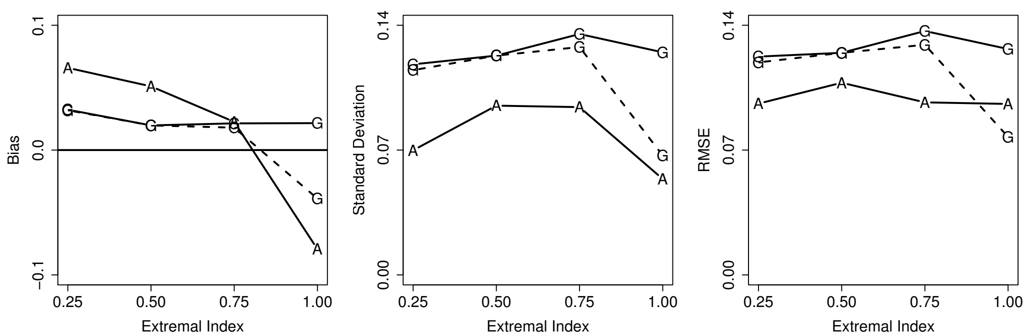

Fig. 2: Estimated biases, standard deviations and root mean square errors of three RB estimators for the extremal index: $\bar{\theta}_A$ (—A—), $\bar{\theta}_G$ (—G—) and $\bar{\theta}_G$ truncated at 1 (- - G - -).