
**Electronic Journal of Applied Statistical Analysis**
**EJASA, Electron. J. App. Stat. Anal.**
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v11n1p296

# Analysis of estimation methods for the extremal index

By Ferreira

Published: 26 April 2018

*Electronic Journal of Applied Statistical Analysis*
Vol. 11, Issue 01, April 2018, 296-306
DOI: 10.1285/i20705948v11n1p296

# Analysis of estimation methods for the extremal index

Marta Ferreira*

*Center of Mathematics of Minho University; CEMAT and CEAUL - University of Lisbon, Campus de Gualtar, 4710-057 Braga, Portugal*

Published: 26 April 2018

Many datasets present time-dependent variation and short-term clustering within extreme values. The extremal index is a primary measure to evaluate clustering of high values in a stationary sequence. Estimation procedures are based on the choice of a threshold and/or a declustering parameter or a block size. Here we revise several different methods and compare them through simulation. In particular, we will see that a recent declustering methodology may be useful for the popular runs estimator and for a new estimator that works under the validation of a local dependence condition. An application to real data is also presented.

**keywords:** declustering, extreme value theory, local dependence conditions, stationary sequences.

## 1 Introduction

Many applications require the evaluation of the impact of clusters of observations inherent to dependencies in a time series. Inference regarding clusters of exceedances over a high threshold, is important to analyze the risk for hazardous events like financial market crashes, large insurance claims, environmental catastrophes, among others.

The extremal index, $\theta$, is a dependent measure that quantifies the clustering tendency of high values. More precisely, a stationary sequence $\{X_n\}_{n\geq 1}$, with marginal distribution $F_{X_n} = F$, has extremal index $\theta \in [0,1]$ if, for each $\tau > 0$, there is a sequence of normalized levels $\{u_n \equiv u_n^{(\tau)}\}_{n\geq 1}$, i.e.,

$$n(1 - F(u_n)) \to \tau,$$

*Corresponding author: msferreira@math.uminho.pt.


ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

as $n \to \infty$, such that

$$P(M_n \leq u_n) \to e^{-\theta\tau}$$

(Leadbetter et al., 1983), where $M_{i,j} = \bigvee_{s=i+1}^{j} X_s$, with $x \vee y = \max(x, y)$, $M_{0,j} = M_j$ and $M_{i,j} = -\infty$ for $i \geq j$. We only address the non-degenerate condition $\theta \neq 0$. The case $\theta = 1$ mimics an independent context, i.e., no clustering, and $\theta < 1$ means limiting conglomerates of exceedances. This topic is addressed under a parametric framework in Ferreira (2016).

Inference within the extremal index may be divided in three main methodological groups: estimation based on a declustering scheme requiring the choice of a threshold and a declustering parameter (Nandagopalan, 1990, Hsing, 1993, Smith and Weissman, 1994, Weissman and Novak, 1998, Robert et al., 2009, Laurini and Tawn, 2003, Süveges and Davison, 2010, Fukutome et al., 2014, and references therein), estimation free-declustering parameter (Ferro and Segers, 2003, Süveges, 2007) and free-threshold maxima procedures only based on the choice of a block size (Gomes, 1993, Ancona-Navarrete and Tawn, 2000, Northrop, 2015).

Under the dependence condition $\mathrm{D}(u_n)$ of Leadbetter (1974), we have

$$P(M_n \leq u_n) \approx F^{n\theta}(u_n). \tag{1}$$

Condition $\mathrm{D}(u_n)$ limits the long-range dependence by stating $\alpha_{n,l_n} \to 0$, as $n \to \infty$, for some sequence $l_n = o(n)$, where

$$\alpha_{n,l} = \sup\{|P\left(M_{i_1,i_1+p} \leq u_n, M_{j_1,j_1+q} \leq u_n\right) - P\left(M_{i_1,i_1+p} \leq u_n\right) P\left(M_{j_1,j_1+q} \leq u_n\right)|,$$

for any integers $1 \leq i_1 < i_1 + p + l \leq j_1 < j_1 + q \leq n$.

If there exists a linear normalization of the d.f. of the maximum that converges to a non-degenerate d.f. $G$, i.e., $F^n(a_n x + b_n) \to G(x)$, as $n \to \infty$, with $a_n \in \mathbb{R}^+$, $b_n \in \mathbb{R}$, $n \geq 1$, then $G$ is a Generalized Extreme Value (GEV) distribution. If we consider $\{X_n^*\}_{n\geq 1}$ independent and with the same d.f. $F$, the associated result for $M_n^* = \bigvee_{i=1}^{n} X_i^*$ leads to the limiting d.f. $G^*(x) = G(x)^{1/\theta}$, with respective location, scale and shape parameters, $(\mu_\theta, \sigma_\theta, \xi_\theta)$ and $(\mu, \sigma, \xi)$, such that

$$\mu_\theta = \mu + \sigma(\theta^\xi - 1)/\xi \quad \text{and} \quad \sigma_\theta = \sigma\theta^\xi.$$

The maxima methods of Gomes (1993) and Ancona-Navarrete and Tawn (2000) are based on a parametric modeling of $M$ and $M_n^*$ through GEV distributions $G$ and $G^*$, the latter requiring resampling. Northrop (2015) presents a semiparametric and more efficient approach by comparing $G$ directly to $F$.

Recently, a new estimation proposal was addressed in Ferreira and Ferreira (2015) that works under the local dependence condition $\mathrm{D}^{(s)}(u_n)$ of Chernick et al. (1991). More precisely, $\mathrm{D}^{(s)}(u_n)$ holds for $\{X_n\}_{n\geq 1}$, if under $\mathrm{D}(u_n)$ and for some $\{b_n\}_{n\geq 1}$ such that,

$$b_n \to \infty,\ b_n\alpha_{n,l_n} \to 0,\ b_n l_n/n \to 0,$$

as $n \to \infty$, we have

$$nP\left(X_1 > u_n, M_{1,s} \leq u_n < M_{s,r_n}\right) \underset{n\to\infty}{\longrightarrow} 0,$$

with $\{r_n = [n/b_n]\}_{n\geq 1}$ ($[x]$ denotes the integer part of $x$).

Under $\mathrm{D}^{(s)}(u_n)$, the extremal index exists and is given by

$$\theta_X = \lim_{n\to\infty} P(M_{1,s} \leq u_n | X_1 > u_n). \tag{2}$$

See Chernick et al. (1991).

The runs estimator (see Hsing, 1993, Nandagopalan, 1990, Weissman and Novak, 1998 and references therein) can be obtained from this relation by considering the run parameter $r$ equal to $s$. The simplest case of the runs estimator $r = 2$ corresponds to the Nandagopalan (1990) proposal constructed under condition $\mathrm{D}^{(2)}(u_n)$. It is based on the ratio between the number of upcrossings (equal to the number of downcrossings) and the number of exceedances and thus avoids the choice of a length $r$ for runs of non-exceedances. The approach in Ferreira and Ferreira (2015) takes advantage of the simplest form of the Nadagopalan's estimators by estimating $\theta_X$ through the extremal index of an auxiliary sequence satisfying $\mathrm{D}^{(2)}(u_n)$.

Consider the time intervals between consecutive exceedances denoted inter-exceedance times, the ones separated by clusters of exceedances denoted inter-cluster times and the intervals between exceedances within a cluster denoted intra-cluster times. Ferro and Segers (2003) estimator is based on the theoretical result that the inter-exceedance times tend to a mixture distribution with parameter $\theta$: inter-cluster times converge to an exponential distribution with probability $\theta$ and intra-cluster times approaches zero with probability $(1-\theta)$. Süveges and Davison (2010) extends this result to truncated intervals that exceed the run parameter in length, arriving at the same limiting mixture law. In Süveges (2007) was only consider the case of a run length equal to one requiring the validity of condition $\mathrm{D}^{(2)}(u_n)$. The proposed method has the advantage of accounting the zero-length intra-cluster times in the likelihood function, avoiding a bias towards one. However, it requires the choice of an additional declustering parameter. This is similar to the runs estimator that requires the identification of the run parameter or the method in Ferreira and Ferreira (2015) where we need to validate $\mathrm{D}^{(s)}(u_n)$ for some $s$. An empirical procedure was presented in this latter reference but it may only be used as a guidance which is subjective.

The problem of a correct selection of either the threshold and/or the declustering parameter (whether the size of a block or a run), is transversal to the different inferential methods of the extremal index and has impact in their performance. Fukutome et al. (2014) introduces an automatic methodology for joint selection of threshold and run parameter based on the method of Süveges and Davison (2010). The choice of the run is also the base of the declustering scheme of the runs estimator, as well as the order $s$ of condition $\mathrm{D}^{(s)}(u_n)$ underlying the method of Ferreira and Ferreira (2015), already mentioned above.

In this paper we analyze through simulation the automatic selection procedure of Fukutome et al. (2014) within the runs estimator and estimators of Ferreira and Ferreira (2015). We also assess their performance by comparing to the estimators of Süveges and Davison (2010) and also with the recent blocks estimators of Northrop (2015). These are described in Section 2. The simulation study is presented in Section 3 and we end with an application to real data.

## 2 Estimators of the extremal index

We are going to describe the estimators of the extremal index mentioned in the introduction, that are going to be used in our analysis.

The classical runs estimator, originated in a characterization of O'Brien (1987), is strongly related to the limit (2), holding under condition $\mathrm{D}^{(s)}(u_n)$, corresponding to its empirical counterpart:

$$\widehat{\theta}^R \equiv \widehat{\theta}^R(u_n, s) = (N(u_n))^{-1} \sum_{i=1}^{n-(s-1)} \mathbf{1}_{\{X_i > u_n\}} \mathbf{1}_{\{X_{i+1} \leq u_n\}} \dots \mathbf{1}_{\{X_{i+s-1} \leq u_n\}},$$

where $N(u_n)$ denote the number of exceedances of $u_n$ and $\mathbf{1}$ is the indicator function. It can also be stated as the reciprocal of the mean cluster size, where two different groups of exceedances of the threshold $u_n$ are consider as independent clusters if there are at least $s-1$ consecutive observations below $u_n$ between them.

The Nandagopalan (1990) estimator corresponds to the particular case $s = 2$ and thus works under condition $\mathrm{D}^{(2)}(u_n)$:

$$\widehat{\theta} = (N(u_n))^{-1} \sum_{i=1}^{n-1} \mathbf{1}_{\{X_{i+1} \leq u_n < X_i\}}.$$

Observe that now the number of clusters correspond to the number of downcrossings of $u_n$.

The method introduced in Ferreira and Ferreira (2015) is intended to take advantage of the simplicity of the estimator, in particular as regards the easy identification of clusters. Under the validation of $\mathrm{D}^{(s)}(u_n)$ for some positive $s$, it is considered an auxiliary process of cycles, $Z_n = \bigvee_{t=(n-1)(s-1)+1}^{n(s-1)} X_t$, $n \geq 1$, and then estimate $\theta$ based on $\{Z_n\}_{n\geq 1}$, for which $\mathrm{D}^{(2)}(u_n)$ holds. More precisely, $\theta$ can be obtained by the ratio between the number of upcrossings of $u_n$ within $\{Z_1, \dots, Z_{[n/(s-1)]}\}$, denoted $U_n^Z(u_n)$ and the number of exceedances of $u_n$ within $\{X_1, \dots, X_n\}$ (Ferreira and Ferreira, 2015; Proposition 2.3), that is,

$$\widehat{\theta}^C = \frac{U_n^Z(u_n)}{N_n^X(u_n)}. \tag{3}$$

Moreover, if $\{X_n\}_{n\geq 1}$ is max-stable, we have

$$\theta = \theta_Z \frac{-\log F_Z(1)}{s-1},$$

where

$$\theta_Z = \frac{1}{1 - E\left(F_Z(Z_1) \vee F_Z(Z_2)\right)} - 2,$$

(see Ferreira and Ferreira, 2015; Proposition 3.2) and a new estimator is obtained by considering the respective empirical counterparts, namely, the empirical d.f. of $F_Z$ and the sample mean corresponding to $E\left(F_Z(Z_1) \vee F_Z(Z_2)\right)$. For more details, see Ferreira and Ferreira (2012). This variant of the cycles estimator given in (3) underneath max-stability will be denoted $\widehat{\theta}^{C_{ms}}$.

The maxima procedure in Northrop (2015) is based on a semiparametric and more efficient methodology than the ones in Gomes (1993) and Ancona-Navarrete and Tawn (2000). Consider blocks of size $b$, the disjoint block maximums $\{Y_i^d = M_{(i-1)b,ib}, i = 1, \ldots, n_d = \lfloor n/b \rfloor\}$, the sliding block maximums $\{Y_i^s = M_{i-1,i+b-1}, i = 1, \ldots, n_s = n - b + 1\}$, $V_i^d = -b \log F(Y_i^d)$, $i = 1, \ldots, n_d$ and $V_i^s = -b \log F(Y_i^s)$, $i = 1, \ldots, n_s$. Under (1) and if $F$ is known, then $V_i^d$ has an exponential distribution with mean $1/\theta$ and thus the respective maximum likelihood estimator is given by the sample mean. Obviously, the same holds for $V_i^s$. Whenever $F$ is unknown, empirical counterparts $\widehat{V}_i^d$ and $\widehat{V}_i^s$ are taken leading to the disjoint and sliding estimators, respectively,

$$\widehat{\theta}_b^D = \frac{n_d}{\sum_{i=1}^{n_d} \widehat{V}_i^d} \quad \text{and} \quad \widehat{\theta}_b^S = \frac{n_s}{\sum_{i=1}^{n_s} \widehat{V}_i^s}.$$

For more details, see Northrop (2015).

Ferro and Segers (2003) have proved that the limit of an inter-exceedance time, normalized by $1 - F(u_n)$, approaches in distribution an exponential-point mass mixture which is zero with probability $1 - \theta$, and the nonzero part, occurring with probability $\theta$, has an exponential distribution with rate $\theta$. Süveges and Davison (2010) modified this result by considering inter-exceedance times truncated by some fixed $K > 0$, which they called the corresponding $K$-gaps, deriving the same limiting mixture exponential law. They considered a likelihood-based estimator of $\theta$. More precisely, assuming that $N$ observations from $\{X_n\}_{n \geq 1}$ exceed the threshold $u_n$, the indices $\{j_i : X_{j_i} > u_n\}$ indicate the locations of the exceedances, the i$^{th}$ $K$-gap denoted by $s_i^{(K)} = \max(j_{i+1} - j_i - K, 0)$, for $i = 1, \ldots, N - 1$, the log-likelihood is given by

$$\ell_K(\theta; s_i^{(K)}) = (N - 1 - N_C) \log(1 - \theta) + 2N_C \log(\theta) - \theta \sum_{i=1}^{N-1} (N/n) s_i^{(K)},$$

with $N_C = \sum_{i=1}^{N-1} \mathbf{1}_{\{s_i^{(K)} \neq 0\}}$. Observe that the likelihood requires independence of the gaps which are actually dependent. However, the method is valid under condition $\mathrm{D}^{(s)}(u_n)$ with $s = K + 1$. For more details, see Süveges (2007) (Section 3.1). Inappropriate choices of the threshold $u_n$ or the run parameter $K$ compromises the model adjustment. Thus, Süveges and Davison (2010) developed misspecification tests based on the information matrix test (IMT) presented in White (1982) to the log-likelihood above. These tests ultimately provide joint graphical choice of threshold and run parameter. Fukutome et al. (2014) introduced an automation of the Süveges and Davison

(2010) method, which tests all pairs of thresholds and run parameters for misspecification of the model, selecting the pair that generates the largest number of observations after declustering, within a subset of small misspecification (IMT $< 0.05$). More precisely, let $\ell_i(\theta)$, $j_i(\theta)$, and $e_i(\theta)$ respectively denote, for each observation $i$, the log-likelihood, the score function and the expected information, and $d_i(\theta) = j_i(\theta) - e_i(\theta)$. Set the sample means of $d_i(\theta)$ and $e_i(\theta)$, respectively, $D(\theta) = (N-1)^{-1}\sum_{i=1}^{N-1} d_i(\theta)$ and $I(\theta) = (N-1)^{-1}\sum_{i=1}^{N-1} e_i(\theta)$. Then, the test statistic is $IMT(\widehat{\theta}) = nD(\widehat{\theta})^2/V(\widehat{\theta})$, where the variance of $D(\theta)$ is $V(\theta) = (N-1)^{-1}\sum_{i=1}^{N-1}\big(d_i(\theta) - D'(\theta)I(\theta)^{-1}\ell_i'(\theta)\big)$. The procedure of Fukutome et al. (2014) consists in ascertain the $(u, K)$ pairs for which IMT$<0.05$ and select the one yielding the largest $N_C$. Prior trials revealed that only pairs associated to more than 80 exceedances must be considered, otherwise there is not enough data to assure a good performance of the test. This automatic choice method of threshold-run parameter will be referred as IMT method and the ML estimator of $\theta$ will be denoted $\widehat{\theta}^{ML}$.

## 3 Simulation Study

In this section we analyze the IMT method applied to the classical runs estimator $\widehat{\theta}^{R}$, and the cycles estimator $\widehat{\theta}^{C}$ and the max-stable cycles estimator $\widehat{\theta}^{C_{ms}}$ of Ferreira and Ferreira (2015). We also include estimators $\widehat{\theta}_q^{R}$ and $\widehat{\theta}_q^{F}$ based on thresholds corresponding to quantiles $q = 0.95, 0.975, 0.99$, with the estimated IMT run parameter (at each replica). For comparison, we also consider estimator $\widehat{\theta}^{ML}$, as well as, the very recent disjoint $\widehat{\theta}_b^{D}$ and sliding $\widehat{\theta}_b^{S}$ estimators of Northrop (2015), requiring the choice of a block size $b$. We consider $b = 5, 10, 20, 40$.

Our simulation study is based on 1000 replicas of samples with size $n = 1000$, generated from the following models: a max-autoregressive process (MAR), $X_i = \phi X_{i-1} \vee \epsilon_i$, with $0 < \phi < 1$, $\{\epsilon_i\}_{i\geq 1}$ an i.i.d. sequence of r.v.'s with d.f. $F_\epsilon(x) = \exp(-(1-\phi)/x)$, $x > 0$ and $\theta = 1 - \phi$ (we consider $\phi = 1/2$ and thus $\theta = 1/2$); a moving maxima (MM), $X_i = \bigvee_{j=0,\ldots,m} \alpha_j \epsilon_{i-j}$, with $\sum_{j=0}^{m} \alpha_j = 1$, $\alpha_j \geq 0$, $\{\epsilon_i\}_{i\geq 1}$ an i.i.d. sequence of unit Féchet distributed r.v.'s and $\theta = \vee_{j=0,\ldots,m}\alpha_j$ (we consider $m = 3$, with $\alpha_0 = 1/3$, $\alpha_1 = 1/6$, $\alpha_2 = 1/2$, leading to $\theta = 1/2$); an autoregressive Cauchy (ARCauchy), $X_i = \beta X_{i-1} + \epsilon_i$, $|\beta| < 1$ and $\theta = 1 - \beta^2$ (we take $\beta = -3/5$ and thus $\theta = 0.64$); an autoregressive uniform (ARUnif), $X_i = -(1/m)X_{i-1} + \epsilon_i$, with $\{\epsilon_i\}_{i\geq 1}$ an i.i.d. sequence, $P(\epsilon_1 = j/m) = 1/m$ for $j = 1, \ldots, m$ and $\theta = 1 - 1/m^2$ (we consider $m = 2$ leading to $\theta = 3/4$); a bivariate extreme value Markov (MCBEV), $P(X_i \leq x, X_{i+1} \leq y) = \exp(-(x^{1/\gamma} + y^{1/\gamma})^\gamma)$ (we take $\gamma = 0.5$ and thus $\theta = 0.328$; see Smith, 1992); a GARCH(1,1), $X_i = \sigma_i \epsilon_i$, with $\sigma_i^2 = \alpha + \lambda X_{i-1}^2 + \beta\sigma_{i-1}^2$, $\alpha, \lambda, \beta > 0$, with $\{\epsilon_i\}_{i\geq 1}$ an i.i.d. sequence of standard Gaussian r.v.'s (we consider $\alpha = 10^{-6}$, $\lambda = 1/4$ and $\beta = 7/10$ yelding $\theta = 0.447$; see Laurini and Tawn, 2012).

Condition $\mathrm{D}^{(2)}(u_n)$ is valid for MAR (Hall, 1996) and condition $\mathrm{D}^{(3)}(u_n)$ holds for models MM (Ferreira and Ferreira, 2015), ARCauchy and ARUnif (Chernick et al., 1991). Conditions $\mathrm{D}^{(4)}(u_n)$ and $\mathrm{D}^{(5)}(u_n)$ were (empirically) validated for, respectively, models MCBEV and GARCH(1,1) in Ferreira and Ferreira (2015).

The obtained estimates of the absolute bias (abias) and the root mean squared error (rmse) are reported in Tables 1-2. The best results are in bold. The mark "+" indicates the least values. The worst performance is marked with italics.

The smallest bias lies generally in the cycles and runs estimators. Based on the rmse, the best performances are associated to the ML estimator $\widehat{\theta}^{ML}$ of Süveges and Davison (2010), followed by the runs estimator under IMT method ($\widehat{\theta}^{R}$). The IMT estimated threshold seems a nice choice when compared to the threshold quantiles 0.95, 0.975 and 0.99. The max-stable cycles estimator $\widehat{\theta}^{C_{ms}}$ presents the smallest rmse for models satisfying max-stability, but it has a bad behavior otherwise, making it very sensitive to this property. Estimators $\widehat{\theta}_b^D$ and $\widehat{\theta}_b^S$ tend to work better for block sizes within 20 and 40. Our findings are thus consistent with the rule $b \approx \sqrt{n}$, usually suggested in the literature. The GARCH model is more favorable to the maxima procedure of Northrop (2015) and the ML estimation of Süveges and Davison (2010).

As a by product, we can test the IMT method in the validation of condition $\mathrm{D}^{(s)}(u_n)$. The mean estimated $s$ values for models MM, ARUnif, ARCauchy, MAR, MCBEV and GARCH were, respectively, 3, 3, 3, 2, 3 and 6. The first four cases coincide with the known theoretical values. In the MCBEV and GARCH models, we do not know the true values of $s$ but the estimates are close to the ones derived through the empirical method applied in Ferreira and Ferreira (2015). We believe that IMT is a promising method in finding the proper order $s$ of condition $\mathrm{D}^{(s)}(u_n)$.

### 3.1 Application to real data

The data corresponds to the daily minimum temperatures (in degrees Fahrenheit) collected at Wooster, Ohio, over the period 1983-1988. We consider the negated series of winter months (thus large values means extreme cold) where the series appears stationary. The observations are plotted in Figure 1, where it is apparent a tendency of high values to occur close to one another. These data were analyzed in Coles (2001) under a parametric framework, where the choice of threshold $u = -10$ and runs 2 and 4 led to $\theta$ estimates of, respectively, 0.42 and 0.27. A longer series was also analyzed in Smith et al. (1997) and in Ferro and Segers (2003). See also references therein. Here, the application of the IMT method estimate the quantile as -2 and the run as 4. Based on these, we obtain the values, $\widehat{\theta}^{R} = 0.344$, $\widehat{\theta}^{C} = 0.313$, $\widehat{\theta}^{C_{ms}} = 0.275$ and $\widehat{\theta}^{ML} = 0.397$. In applying, respectively, the disjoint and sliding blocks estimators, we derive $\widehat{\theta}^{D} = 0.290$ and $\widehat{\theta}^{S} = 0.294$, with $b = 25$.

## Acknowledgement

The author wishes to thank the reviewer for his helpful comments that have improved this work. 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/00006/2013 and by the research center CEMAT (Instituto Superior Técnico, Universidade de Lisboa) through the Project UID/Multi/04621/2013.

Table 1: Results of the root mean squared error: the values in bold correspond to the better performances with the best one ticked with the signal +; the largest values are marked with italics.

| rmse | MM | ARUnif | ARCauchy | MAR | MCBEV | GARCH |
|---|---|---|---|---|---|---|
| $\widehat{\theta}^{C}_{0.95}$ | 0.069 | 0.063 | 0.095 | 0.079 | **0.084** | 0.158 |
| $\widehat{\theta}^{C}_{0.975}$ | **0.066** | 0.089 | 0.100 | 0.095 | 0.126 | 0.170 |
| $\widehat{\theta}^{C}_{0.99}$ | 0.096 | 0.138 | 0.148 | 0.153 | 0.205 | 0.228 |
| $\widehat{\theta}^{C}$ | 0.089 | **0.063** | 0.118 | 0.084 | **0.071** | 0.167 |
| $\widehat{\theta}^{C_{ms}}$ | **0.032**$^{+}$ | *0.791* | *0.608* | **0.032**$^{+}$ | **0.063**$^{+}$ | *0.352* |
| $\widehat{\theta}^{R}_{0.95}$ | **0.055** | **0.063** | 0.084 | 0.071 | 0.089 | 0.155 |
| $\widehat{\theta}^{R}_{0.975}$ | **0.063** | 0.089 | 0.095 | 0.095 | 0.134 | 0.187 |
| $\widehat{\theta}^{R}_{0.99}$ | 0.095 | 0.138 | 0.152 | *0.155* | 0.212 | 0.257 |
| $\widehat{\theta}^{R}$ | 0.095 | **0.045**$^{+}$ | **0.014**$^{+}$ | 0.089 | **0.071** | 0.184 |
| $\widehat{\theta}^{ML}$ | **0.056** | 0.096 | **0.060** | **0.049** | **0.083** | **0.089** |
| $\widehat{\theta}^{D}_{b=5}$ | *0.134* | 0.122 | **0.032** | 0.134 | *0.305* | 0.187 |
| $\widehat{\theta}^{D}_{b=10}$ | 0.077 | 0.192 | 0.089 | 0.077 | 0.239 | 0.122 |
| $\widehat{\theta}^{D}_{b=20}$ | **0.063** | 0.226 | 0.122 | **0.063** | 0.210 | 0.100 |
| $\widehat{\theta}^{D}_{b=40}$ | 0.077 | 0.247 | 0.145 | 0.077 | 0.202 | 0.105 |
| $\widehat{\theta}^{S}_{b=5}$ | *0.134* | 0.118 | **0.032** | 0.134 | 0.305 | 0.187 |
| $\widehat{\theta}^{S}_{b=10}$ | 0.071 | 0.190 | 0.084 | 0.071 | 0.237 | 0.118 |
| $\widehat{\theta}^{S}_{b=20}$ | **0.055** | 0.224 | 0.118 | **0.055** | 0.205 | **0.089** |
| $\widehat{\theta}^{S}_{b=40}$ | **0.055** | 0.247 | 0.141 | **0.055** | 0.190 | **0.084**$^{+}$ |

Table 2: Results of the absolute bias: the values in bold correspond to the better performances with the best one ticked with the signal +; the largest values are marked with italics.

| abias | MM | ARUnif | ARCauchy | MAR | MCBEV | GARCH |
|---|---|---|---|---|---|---|
| $\widehat{\theta}^{C}_{0.95}$ | 0.045 | **0.002** | 0.061 | 0.031 | **0.017** | **0.031** |
| $\widehat{\theta}^{C}_{0.975}$ | **0.013** | **0.008** | 0.025 | **0.003**$^{+}$ | 0.049 | **0.037** |
| $\widehat{\theta}^{C}_{0.99}$ | **0.017** | 0.018 | **0.020** | 0.049 | 0.114 | 0.121 |
| $\widehat{\theta}^{C}$ | 0.073 | 0.023 | 0.100 | 0.064 | **0.002**$^{+}$ | **0.023**$^{+}$ |
| $\widehat{\theta}^{C_{ms}}$ | **0.002**$^{+}$ | *0.763* | *0.600* | **0.005** | 0.044 | 0.323 |
| $\widehat{\theta}^{R}_{0.95}$ | 0.034 | **0.004** | 0.047 | 0.023 | **0.031** | **0.028** |
| $\widehat{\theta}^{R}_{0.975}$ | **0.005** | 0.011 | **0.016** | **0.009** | 0.062 | 0.092 |
| $\widehat{\theta}^{R}_{0.99}$ | 0.022 | 0.021 | **0.025** | 0.053 | 0.126 | 0.174 |
| $\widehat{\theta}^{R}$ | 0.081 | **0.001**$^{+}$ | 0.099 | 0.070 | **0.003** | 0.063 |
| $\widehat{\theta}^{ML}$ | 0.012 | 0.064 | 0.043 | **0.006** | 0.057 | **0.045** |
| $\widehat{\theta}^{D}_{b=5}$ | *0.132* | 0.118 | **0.008**$^{+}$ | *0.132* | 0.304 | *0.185* |
| $\widehat{\theta}^{D}_{b=10}$ | 0.063 | 0.187 | 0.077 | 0.063 | 0.235 | 0.116 |
| $\widehat{\theta}^{D}_{b=20}$ | 0.030 | 0.220 | 0.110 | 0.030 | 0.202 | 0.083 |
| $\widehat{\theta}^{D}_{b=40}$ | **0.016** | 0.234 | 0.124 | 0.016 | 0.188 | 0.069 |
| $\widehat{\theta}^{S}_{b=5}$ | *0.132* | 0.118 | **0.008**$^{+}$ | *0.132* | *0.304* | *0.185* |
| $\widehat{\theta}^{S}_{b=10}$ | 0.063 | 0.187 | 0.077 | 0.063 | 0.235 | 0.116 |
| $\widehat{\theta}^{S}_{b=20}$ | 0.029 | 0.221 | 0.111 | 0.029 | 0.201 | 0.082 |
| $\widehat{\theta}^{S}_{b=40}$ | **0.010** | 0.240 | 0.130 | 0.010 | 0.182 | 0.063 |

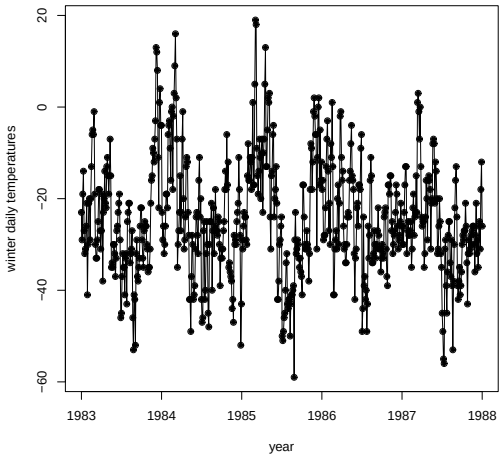


Figure 1: Negated daily minimum temperatures of winter months November-February in Wooster, Ohio (sample size $n = 602$).

## References

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

Chernick M.R., Hsing T., and McCormick W.P. (1991). Calculating the extremal index for a class of stationary sequences. *Advances in Applied Probability*, 23, 835–850.

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

Ferreira, M. (2016). The Lawrence-Lewis Pareto process: an extremal approach. *Electronic Journal of Applied Statistical Analysis*, 9(1), 76–89.

Ferreira, H. and Ferreira, M. (2012). On extremal dependence of block vectors. *Kybernetika*, 48(5), 988–1006.

Ferreira, H. and Ferreira, M. (2015). Estimating the extremal index through local dependence. Accepted for publication in Annales de lInstitut Henri Poincar.

Ferro C.A.T. and Segers J. (2003). Inference for clusters of extreme values. *Journal of the Royal Statistical Society: Series B*, 65, 545–556.

Fukutome, S., Liniger M.A., and Süveges M. (2014). Automatic threshold and run parameter selection: a climatology for extreme hourly precipitation in switzerland. *Theoretical and Applied Climatology*, 120(3), 403–416.

Gomes, M. (1993). On the estimation of parameters of rare events in environmental time series. In *V. Barnett and K. Turkman (Eds.), Statistics for the environment 2: Water Related Issues*, pages 225–241. Wiley.

Hall, A. (1996). Maximum term of a particular autoregressive sequence with discrete margins. *Comm. Statist. Theory Methods*, 25(4), 721–736.

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

Laurini F. and Tawn J. (2003). New estimators for the extremal index and other cluster characteristics. *Extremes*, 6, 189–211.

Laurini F. and Tawn J. (2012). The extremal index for GARCH(1,1) processes. *Extremes*, 15, 511–529.

Leadbetter, M.R. (1974). On extreme values in stationary sequences. *Zeitschrift fr Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 28(4), 289–303.

Leadbetter, M.R., Lindgren, G., and Rootzn, H. (1983). Extremes and Related Properties of Random Sequences and Processes. New York: Springer.

Nandagopalan S. (1990). Multivariate extremes and estimation of the extremal index. Ph.D. Thesis, University of North Carolina, Chapel Hill.

Northrop, P.J. (2015). An efficient semiparametric maxima estimator of the extremal index. *Extremes*, 18(4), 585–603.

O'Brien, G.L. (1987). Extreme values for stationary and Markov sequences. *Annals of Probability*, 15, 281–291.

Robert, C.Y., Segers J., and Ferro C. (2009). A sliding blocks estimator for the extremal index. *Electronic Journal of Statistics*, 3, 993–1020.

Smith, R.L. (1992). The extremal index for a Markov chain. *Journal of Applied Probability*, 29, 37–45.

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

Smith, R.L., Tawn, J.A., and Coles, S.G. (1997). Markov chain models for threshold exceedances. Biometrika, 84, 249–268.

Süveges, M. (2007). Likelihood estimation of the extremal index. *Extremes*, 10, 41–55.

Süveges, M. and Davison A.C. (2010). Model misspecification in peaks over threshold analysis. *Annals of Applied Statistics*, 4, 203–221.

Weissman I. and Novak S.Y. (1998). On blocks and runs estimators of the extremal index. *Journal of Statistical Planning and Inference*, 66(2), 281–288.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50, 1–25.