Likelihood estimation of the extremal index

Mária Süveges

Abstract The article develops the approach of Ferro and Segers (J.R. Stat. Soc., Ser. B 65:545, 2003) to the estimation of the extremal index, and proposes the use of a new variable decreasing the bias of the likelihood based on the point process character of the exceedances. Two estimators are discussed: a maximum likelihood estimator and an iterative least squares estimator based on the normalized gaps between clusters. The first provides a flexible tool for use with smoothing methods. A diagnostic is given for condition $D^{(2)}(u_n)$, under which maximum likelihood is valid. The performance of the new estimators were tested by extensive simulations. An application to the Central England temperature series demonstrates the use of the maximum likelihood estimator together with smoothing methods.

Keywords Central England temperature data · Clusters · Diagnostic · Extremal index · Extreme value theory · Gaps · Local likelihood

AMS 2000 Subject Classification 62G32 · 62P12 · 62M09 · 62F99

1 Introduction

In recent years there has been increasing interest in the grouping characteristics of extreme events. Applications include the temporal distribution of large insurance claims and of financial crashes, and incidences of catastrophic weather events. These motivate a search for reliable tools to describe these

M. Süveges (✉)
École Polytechnique Fédérale de Lausanne, IMA-FSB-EPFL,
Station 8, Lausanne CH-1015, Switzerland
e-mail: maria.suveges@epfl.ch

features, a key aspect being the estimation of the extremal index θ, which governs the clustering of the extremes of a univariate observational series.

In addition to the runs and blocks estimators of θ (Smith and Weissman 1994; Hsing 1993; Weissman and Novak 1998), two other estimators have recently been proposed: a two-threshold method (Laurini and Tawn 2003), and the intervals estimator (Ferro and Segers 2003). The first estimates the extremal index by means of the number of independent clusters observed in the sample, and requires a preliminary identification of clusters. The intervals estimator exploits the compound Poisson character of the point process of exceedances; with an appropriate normalization, this implies that the inter-exceedance times follow approximately a mixture distribution comprising a point mass at zero and an exponential distribution.

Both new approaches have drawbacks. The two-threshold method requires the choice of two declustering parameters, although it is less sensitive to the choice of the run length than is the runs estimation. Furthermore, in finance, where volatility-driven models are common, preliminary modeling of the time series is required to select the second parameter of the method, a lower threshold that separates low- and high-volatility periods. Like moment estimators in general, the intervals estimator is hard to use with smoothing methods, as would be necessary in investigating changes of the extremal index over time. Moreover, the use of the methods based on the distribution of the inter-exceedance times is inherently awkward, since zero inter-exceedance times cannot be observed in practice; thus, all the inter-exceedance times are attributed to the exponential part of the mixture.

The article follows Ferro and Segers (2003) by investigating the behaviour of likelihood methods for the observed separation of the exceedances of a process over a threshold u_n. We exploit the fact that in the usual limit, the number of observations below u_n between two consecutive exceedances follow the same point mass–exponential mixture distribution as the normalized inter-exceedance times of Ferro and Segers (2003). Consequently, a likelihood can be introduced that well describes processes with independent inter-exceedance times and yields a maximum likelihood estimator. The validity of these estimators can be extended to dependent processes satisfying the long-range approximate independence condition of Ferro and Segers (2003) and the condition $D^{(2)}(u_n)$ of Leadbetter and Nandagopalan (1989) and Chernick et al. (1991). A second method, an iterative weighted least squares procedure based on the exponential quantile–quantile plot of the normalized spacings, yields a self-consistent way to determine a limit separation over which clusters can be regarded as independent, and provides an estimate of the extremal index.

To assess the performance of the new estimators and to compare them to the runs, two-threshold and intervals methods, data were simulated from a number of different models with extremal index ranging from independence to strong short-range dependence. The simulations showed that the maximum likelihood estimator has lower bias and mean squared error than the runs, two-threshold and intervals methods in cases where clusters are well modeled by sequences of adjacent exceedances. Despite its somewhat higher root mean

squared error, the iterative least squares procedure showed remarkably low sensitivity to threshold choice and to the extremal dependence structure of the simulated models.

To demonstrate the use of the maximum likelihood estimator, we investigate possible changes in the extremal index of cold winter temperatures in the Central England temperature (CET) series, using local likelihood methods. To obtain confidence limits, we applied bootstrap techniques.

In Section 2, we introduce notation used throughout the article, and state the basic theorem and its corollaries on which the estimation methods are based. Section 3 discusses the problems with a likelihood for inter-exceedance times, introduces the gaps as a partial solution, and describes the two proposed estimators. Further, it proposes a way of checking the condition $D^{(2)}(u_n)$ preliminary to estimation, and simulations are described. Smoothing methods, comparison of the different estimators using processes with known changing extremal index and the application to the CET data are presented in Section 4. Discussion is given in Section 5.

2 Definitions and Preliminaries

Let $X_1, X_2, ...$ be a strictly stationary process with marginal distribution function F, and let $M_n = \max\{X_1, \ldots, X_n\}$. Define the independent sequence of variables $\hat{X}_1, \hat{X}_2, ...$ with the same marginal distribution F as $X_1, X_2, ...,$ and let $\hat{M}_n = \max\{\hat{X}_1, \ldots, \hat{X}_n\}$ denote the maximum of this sequence. Under appropriate mixing conditions, if there exists a nondegenerate limiting distribution G for the variable $(\hat{M}_n - b_n)/a_n$ with some sequences of constants $\{a_n > 0\}$ and $\{b_n\}$, then the normalized maximum $(M_n - b_n)/a_n$ of the dependent series also has a nondegenerate limit distribution G^*, and they are related to each other by $G^* = G^\theta$, where θ is the extremal index (Leadbetter 1983). If we map the sequence from $(0, n]$ into the interval $(0, 1]$ by dividing the times by n, the locations of exceedances over a high threshold u_n in a stationary process are in the limit distributed as the points of a homogeneous Poisson process on the interval $(0, 1]$. The rate for each threshold is constant, equal to $G^*((u_n - b_n)/a_n)$, and is related to θ via the relation between G^* and G. This characterization of the times of exceedances suggests an estimator of the extremal index (Ferro and Segers 2003): in a homogeneous Poisson process with rate λ, the intervals between two neighboring points follow an exponential distribution with parameter λ. With estimation in mind, a very convenient normalization of the occurrence times can be found. For this purpose, introduce the random variable

$$T(u_n) = \min\{n \geq 1 : X_{n+1} > u_n \mid X_1 > u_n\}$$

for the inter-exceedance times in the stationary sequence $\{X_i\}$. Ferro and Segers (2003) have proved the basic convergence theorem for this estimation method under the following condition.

Condition $\Delta^*(u_n)$ For any $A \in \mathcal{F}_{1,k}(u_n)$ with $P(A) \neq 0$, $B \in \mathcal{F}_{k+l,cr_n}(u_n)$ and $1 \leq k \leq cr_n - l$,

$$|P(B \mid A) - P(B)| \leq \alpha^*(cr_n, l)$$

for all $c > 0$, and there exists a sequence $l_n = o(n)$ for which $\alpha^*(cr_n, l_n) \to 0$ as $n \to \infty$ for all $c > 0$.

Theorem 1 (Ferro and Segers 2003) *If there exist sequences of integers $\{r_n\}$ and of thresholds $\{u_n\}$ such that*

(1) $r_n \to \infty$, $r_n \overline{F}(u_n) \to \tau$ and $P\{M_{r_n} \leq u_n\} \to e^{-\theta\tau}$ as $n \to \infty$ for some $\tau \in (0, \infty)$ and $\theta \in [0, 1]$, and
(2) *Condition $\Delta^*(u_n)$ is satisfied,*

then as $n \to \infty$, $P\{\overline{F}(u_n) T(u_n) > t\} \to \theta \exp(-\theta t)$ for $t > 0$.

This theorem states that in the limit an inter-exceedance time, normalized by $\overline{F}(u_n)$, tends in distribution to a variable following an exponential-point mass mixture: it is zero with probability $1 - \theta$, nonzero with probability θ, and conditionally on being nonzero, it follows an exponential distribution with mean θ^{-1}. The extremal index plays a double role: it determines both the proportion of zero and nonzero inter-exceedance times and the expected value of the normalized inter-cluster times. The first role corresponds to the observation of Leadbetter (1983) that the limiting mean cluster size is θ^{-1}, since thus a proportion $1 - \theta$ of the exceedances has to occur as secondary elements of clusters, and in the limit these have zero inter-exceedance times.

3 Likelihood and Estimation

3.1 Likelihood

Suppose that we observe a random sequence $X_1, X_2, \ldots, X_n$ that satisfies $\Delta^*(u_n)$. Suppose that N observations exceed the threshold u_n, let the collection of indices $\{j : X_j > u_n\}$ denote the locations of the exceedances, and let T_i denote the ith inter-exceedance time, that is, $T_i = j_{i+1} - j_i$, $i = 1, \ldots, N - 1$. If the point process of the exceedances were well approximated by a compound Poisson process, and the inter-exceedance times were independent, a simple likelihood could be proposed. Two problems with this approach are:

(a) Although the distances between the exceedances of the same cluster are zero in the limit process, the usual distance estimate as given by T_i provides nonzero values for any n. Therefore, each interval between exceedances is nonzero and is assigned to the exponential part of the likelihood, causing likelihood-based estimators to be biased toward independence.

(b) The inter-exceedance times are dependent, and the use of the independent likelihood is unfounded.

Problem (a) can be solved by the introduction of a new variable. Let $S(u_n)$ denote $T(u_n) - 1$, and call it a gap of the exceedances to distinguish it from inter-exceedance times. Theorem 1 can be stated for the gaps without change; the proof follows that of Theorem 1, and since it requires only a slight modification by considering $P\{\overline{F}(u_n)(T(u_n) - 1) > t\}$, we do not present it here. Thus, we can propose an independent log-likelihood in terms of $\overline{F}(u_n)S(u_n)$ instead of $\overline{F}(u_n)T(u_n)$:

$$\ell(\theta; \{S_i\}) = (N - 1 - N_C) \log(1 - \theta) + 2N_C \log \theta - \theta \sum_{i=1}^{N-1} \overline{F}(u_n)S_i, \quad (1)$$

where $N_C = \sum_{i=1}^{N-1} I(S_i \neq 0)$.

For problem (b), the dependence of the spacing process, we have to check how the validity of the likelihood can be extended. Modeling the exceedance distances by $S_i(u_n)$ is equivalent to modeling the sequences of adjacent exceedances in the observed series by the clusters of the limiting point process. The condition for this model to be valid is $D^{(2)}(u_n)$ of Chernick et al. (1991) that is based on Leadbetter and Nandagopalan (1989).

Condition $D^{(2)}(u_n)$ is said to be satisfied if

$$nP\{X_j > u_n, X_{j+1} \leq u_n, M_{j+2,r_n} > u_n\} \to 0 \text{ as } n \to \infty,$$

with the sequence $\{u_n\}$ as in the long-range asymptotic independence condition, and a sequence of block sizes $\{r_n\}$ such that $n/r_n \to \infty$; moreover, simultaneously in every block, the asymptotic long-range independence condition $D(u_n)$ of Leadbetter et al. (1983) is satisfied.

Condition $D^{(2)}(u_n)$ ensures that the probability of again exceeding the high threshold u_n after dropping below it within a cluster tends to zero, and provides thus a sufficient condition for the contiguous sequence model to be adequate. From the full log-likelihood (1), we can derive the maximum likelihood estimator

$$\hat{\theta} = \frac{\sum_{i=1}^{N-1} qS_i + N - 1 + N_C - \left[\left(\sum_{i=1}^{N-1} qS_i + N - 1 + N_C\right)^2 - 8N_C \sum_{i=1}^{N-1} qS_i\right]^{1/2}}{2 \sum_{i=1}^{N-1} qS_i}, \quad (2)$$

where q denotes the estimated value of $\overline{F}(u)$.

When $N_C = N - 1$, the likelihood function (1) is monotone increasing within the interval $[0, 1]$, with maximum attained at $\theta = 1$. The maximum likelihood estimator can be then defined by $\hat{\theta} = 1$; this is equivalent to allowing $N_C = N - 1$ in (2).

Another estimator exploits the fact that, as was noted by Ferro (2003), the standard exponential quantile-quantile plot of $\overline{F}(u_n)S(u_n)$ can be fitted by a

broken-stick model; one segment is composed of zeros and the other is a line with slope θ^{-1}, intersecting each other at $(-\log\theta, 0)$. In practice, we have found that the best way to use this structure is the following iterative weighted least squares procedure:

(1) Select the nonzero gaps $\{s_i^*\}$ from the collection of gaps, scale them by their mean and order them to make a sample $\{\chi_i : i = 1, \ldots, N_C\}$, to compare to standard exponential quantiles $x_i = -\log(1 - i/(N_C + 1))$ for $i = 1, \ldots, N_C$;

(2) Fit a weighted least squares model to the points $\{x_i, \chi_i\}$ with weights given by $w_i^{-1} = \sum_{j=N-i}^{N-1} j^{-2}$ to obtain the estimated intersect $\hat{\alpha}$ and slope $\hat{\beta}$;

(3) Find an estimate of the extremal index by $\hat{\theta}_{LS} = \min\left\{\exp\left(\hat{\alpha}/\hat{\beta}\right), 1\right\}$; and finally

(4) Choose the largest $\lfloor\hat{\theta}_{LS}(N-1)\rfloor$ spacings from the original collection $\{s_i^*\}$, and after ordering, compare them to the sample used to obtain $\hat{\theta}_{LS}$. If they are different, apply steps (2)–(4) for the points $\{(x_i, \chi_i) : i \geq N - \lfloor\hat{\theta}_{LS}(N-1)\rfloor\}$. If they are identical, accept $\hat{\theta}_{LS}$ obtained in (3) as the estimated extremal index.

This estimator works by finding a self-consistent estimator for θ: a first guess for θ is given by using the exponential nature of the inter-cluster distances, then, making use of the fact that a proportion of $1 - \theta$ of the observed exceedances should be "secondary" cluster elements, it determines a new set of inter-cluster gaps obtained from the first guess for θ. Then a second estimate is given, based on the exponential distribution of the new gap set. The iteration is continued until it reaches a stable point, where neither the estimate nor the set of gaps change further. That is to say, since there are two roles to the extremal index, first by determining the mean cluster size, second, by determining the distribution of the inter-cluster times, this procedure uses one to correct the estimate by the other, and yields the estimate where these two agree.

3.2 Condition $D^{(2)}(u_n)$

The crucial condition $D^{(2)}(u_n)$ for the validity of the maximum likelihood estimator (2) may be checked by calculating

$$p(u, r) = \frac{\sum_{j=1}^{n} I(X_j > u, X_{j+1} \leq u, M_{j+2,r} > u)}{\sum_{j=1}^{n} I(X_j > u)}$$

for the observed sequence $X_1, \ldots, X_n$, where I is the indicator function. Given u and r, we can compute the proportion of the anti-$D^{(2)}(u_n)$ events $\{X_{j+1} \leq u, M_{j+2,r} > u \mid X_j > u\}$ among the exceedances for a range of thresholds and block sizes.

Examples are given in Fig. 1; the upper left panel refers to an AR(2) process with $\phi_1 = 0.93$ and $\phi_2 = -0.86$, which does not satisfy condition
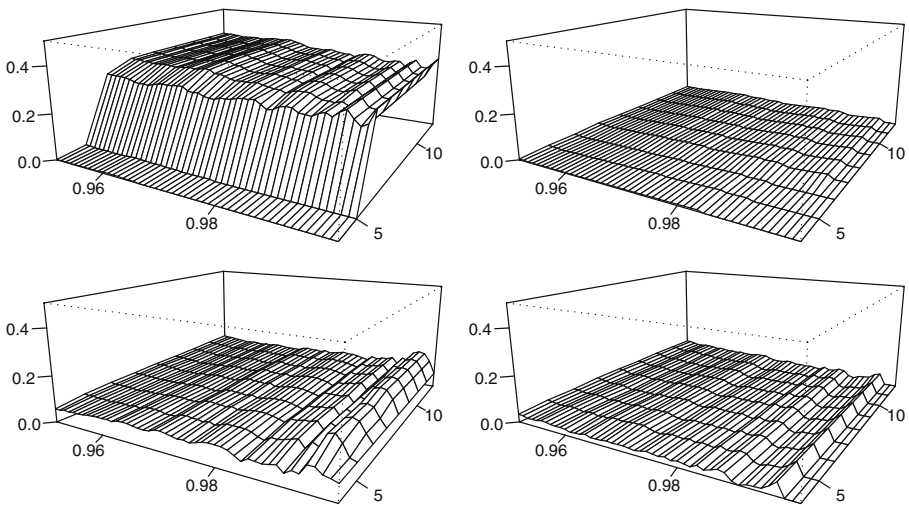
Fig. 1 The observed proportion of anti-$D''(u_i)$ events for the oscillating AR(2) (*upper left panel*), moving maximum process (*upper right*), Markov chain (*lower left*) and ARMA(1,4) (*lower right*). *Foreground horizontal axis*: $F(u)$, *right-hand side horizontal axis*: block size r, *vertical axis*: $p(u, r)$

$D^{(2)}(u_n)$; the upper right panel shows a moving maximum sequence $\max\{Z_{i-3}, Z_{i-2}, Z_{i-1}, Z_i\}$, Z_i being an independent, identically distributed random sequence, which satisfies $D^{(2)}(u_n)$. The two lower panels contain simulated examples used for the demonstration of the behaviour of the maximum likelihood estimator in Section 3.3; these are a Markov process with standard Gumbel margins and symmetric logistic joint distribution with parameter $r = 2$ (Smith 1992), and an ARMA(1,4) process with innovations following a Pareto distribution, having parameters $\phi = 0.82$, $\vartheta_1 = -0.11$ and $\vartheta_i = 0.13$ for $i = 2, 3, 4$.

One aspect of the plots is the trend of the proportions. $D^{(2)}(u_n)$ is a limit condition satisfied if there exists a path (u_i, r_j) with $u_i \to \infty$ and $r_j \to \infty$ according to condition $D^{(2)}(u_n)$, for which $p(u_i, r_j) \to 0$. Finding such a direction on the plots amounts to finding that there is a threshold where the contiguous cluster model will be a good approximation. On the other hand, even if there is no direction with marked downward trend, an impression of the bias can be gained by the proportion of anti-$D^{(2)}(u_n)$ events, since these give information on how many clusters are misidentified as two or more clusters; if such clusters are few, we might decide to accept this relatively small upward bias despite the possible failure of $D^{(2)}(u_n)$. The four panels of Fig. 1 illustrate these types of behaviour: for the AR(2) process, there is no direction of downward trend, and we can expect a very high bias; for the moving maxima sequence, $D^{(2)}(u_n)$ is satisfied and the maximum likelihood estimator works well; for the ARMA(1,4) sequence, there seems to be a constant but low proportion of anti-$D^{(2)}(u_n)$ events independently of (u, r), and as we shall see in Subsection 3.3, the maximum likelihood estimator yields good results; and for the Markov

chain model, the proportion is high enough to cause a bias of around 20% in the maximum likelihood estimator $\hat{\theta}$ (cf. Fig. 2).

3.3 Simulation Studies

We compared the performances of the new estimators to the intervals, runs, and two-threshold estimators on examples from three families of simple models: ARMA, MARMA, and Markov chain processes. We simulated each process with parameter combinations yielding exceedance sequences with different extremal indexes ranging from near-independence to strong dependence. Each parameter combination in every process was used to obtain $R = 500$ independent repetitions of sequences with lengths $n = 2,000$ and $30,000$. For each simulated process, estimation was performed on the 500 repetitions using the different estimators with four different thresholds corresponding to empirical exceedance probabilities 0.95, 0.975, 0.9825 and 0.99. The variance and bias of the estimators were assessed by calculating the root mean squared error and median relative bias $B_{method} = \left(\text{median}_{i=1,\ldots,R}\{\hat{\theta}_{method,i}\} - \theta \right)\theta^{-1}$ and plotting them against thresholds. The results can be summarized as follows:

1. The simulations justified the role of condition $D^{(2)}(u_n)$: for any process that satisfies this condition, the maximum likelihood estimator performs better than the intervals or iterative least squares methods in terms of both bias and of mean squared error. The results for two processes are shown in
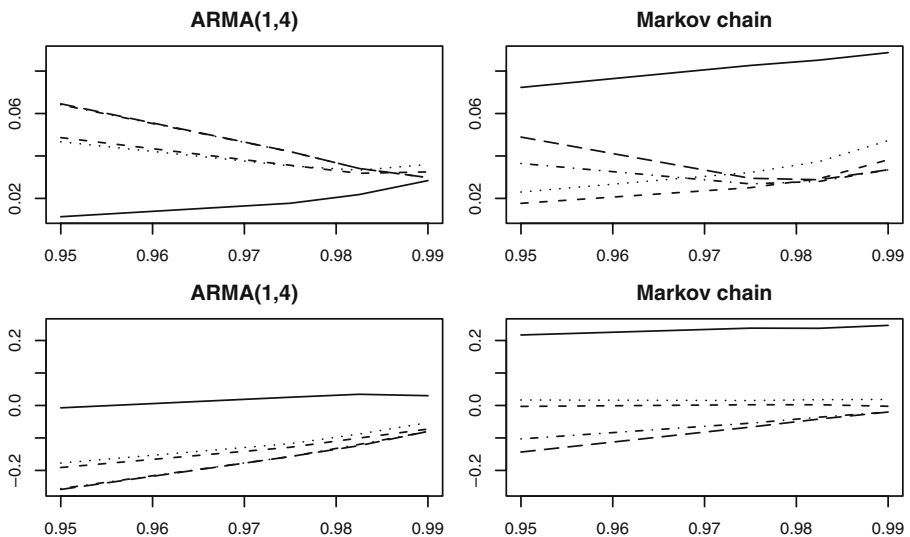

Fig. 2 Comparison of the maximum likelihood (*solid*) and iterative least squares (*dashed*) estimators to the intervals (*dotted*), 2-threshold (*dash-dotted*) and runs estimators (*long-dashed*) on an ARMA(1,4) process and on a Markov chain with Gumbel margins and symmetric logistic bivariate joint distribution. For the 2-threshold method, run length 8 and lower threshold $F(l_n) = 0.5$ were chosen. For the runs method, run length 8 was used. The number of observations was $n = 30,000$

Fig. 2; these are the ARMA(1,4) and the Markov chain given in detail in Section 3.2, for which the diagnostic plots to check $D^{(2)}(u_n)$ are presented in Fig. 1.

2. Like the intervals estimator, the iterative weighted least squares estimator does not require condition $D^{(2)}(u_n)$, and its performance is very close to that of the intervals method. As the iterative least squares method uses only the longer separations for the estimation, its variance is larger for short series with high thresholds, but since it is less sensitive to the threshold choice (its behaviour changed little for any threshold in the range $\overline{F}(u_n) \geq 0.05$), this can be compensated by choosing a relatively low threshold.

3. Simulations from processes with independent observations for which $\theta = 1$ confirmed that with the extension of the likelihood described in Section 3, the application of nonparametric bootstrap methods can be used for inference for possibly independent data sets too.

4 Local Extremal Index

To demonstrate the use of our maximum likelihood estimator, we examine now the possibility to apply them to detect nonstationarity in data. Stationarity is a condition only rarely observed in practice for long time series; a fixed type of parametric model with slowly varying parameters more often yields a useful description of the data. The point process description of the extremes of Section 2 and the likelihood model presented in Section 3 offer a way to recognize problems related to nonstationarity: likelihood models can easily be used together with smoothing methods.

4.1 Simulation Studies

We simulated ARMA(p, q) sequences of various extremal indexes, and attached them to form processes with abruptly changing extremal characteristics. All the ARMA(p, q) sequences satisfied condition $D^{(2)}(u_n)$, so that the maximum likelihood estimator was expected to behave well.

To compare the maximum likelihood, the intervals, and the iterative least squares estimators, we estimated the extremal index of the series with a smoothing interval of length 7,301 with thresholds corresponding to $F(u_1) = 0.95$ and $F(u_2) = 0.98$. For every estimation method, θ was taken to be constant for the whole window. A weighted form of the maximum likelihood estimator was used with weights derived from a truncated normal kernel, and to estimate the variance of the method, we applied a parametric bootstrap consisting of the simulation of homogeneous Poisson processes and finding the 95% pointwise percentile intervals.

The intervals and the iterative least squares estimators allow only for the use of equal weights. For them, the bootstrap method proposed by Ferro and Segers (2003) was applied to obtain confidence intervals.

The results on one process are shown in Fig. 3; the two processes composing the estimated sequence were the same as our first example in Section 3 ($\theta = 0.25$) for the first part and another ARMA(1,4) with coefficients $\phi = 0.38$, $\vartheta_1 = 0.74$, $\vartheta_2 = 0.72$, $\vartheta_3 = 0.05$, $\vartheta_4 = 0.01$ ($\theta = 0.13$) for the second. The figure reflects well the generally observed characteristics. The wobbles in the estimates with intervals and iterative least squares at locations where the true extremal index was constant are due to the use of constant weights that allocate too much influence to observations entering or leaving the window. In most cases it is worthwhile to use relatively low thresholds, around the quantile $F^{-1}(0.95)$.

4.2 Central England Temperature Series

We use the Central England Temperature data to assess possible changes in the extremal index of unusually cold winter temperatures. This series of daily mean temperatures, representative of a roughly triangular area of the United Kingdom enclosed by Preston, London and Bristol, stretches over 233 years and provides excellent test data for the application of kernel methods. Data from 1 January 1772 through 31 December 2004 were used.

Since the series has both a seasonal component and a trend, a two-step procedure was applied to obtain a series having roughly the same range
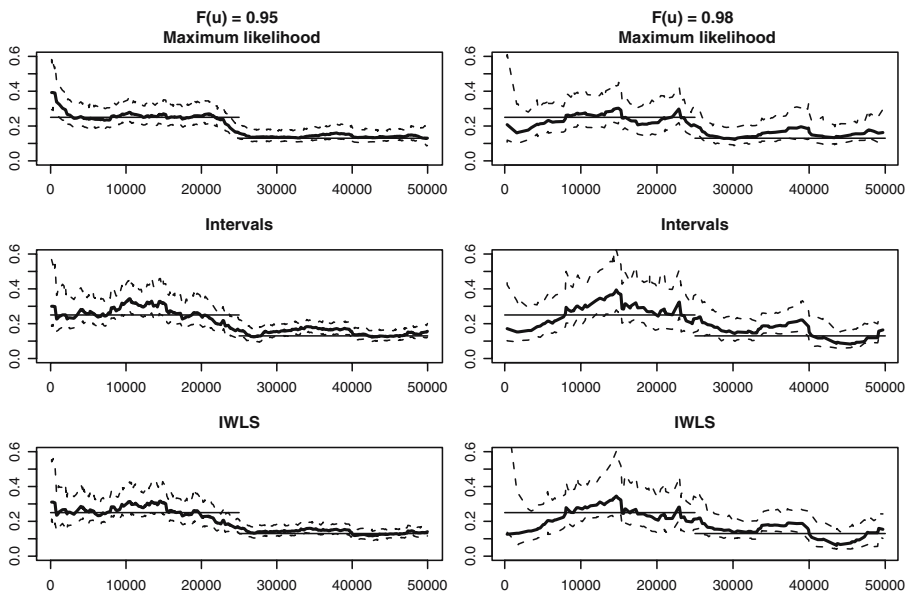

Fig. 3 Comparison of the weighted MLE, intervals and iterative least squares estimators with thresholds given by $F(u_1) = 0.95$ (*left panels*) and $F(u_2) = 0.98$ (*right*). The estimators are the weighted maximum likelihood (*top*), the intervals (*middle*), and the iterative least squares (*bottom*). *Heavy solid lines*: estimates, *thin dashed*: 95% percentile bootstrap CI, *thin solid*: true extremal index

over the whole period, enabling us to determine common thresholds. Yearly seasonal variation in the median and in the dispersion was approximated by a cubic spline-smoothed median and median absolute deviation taken for each day of the year over the 233 years of the series, then used to center and scale the data. Any longer-term trend, which seems to become non-negligible in the last 50 years, was removed by centering and scaling with cubic spline-smoothed 10-year medians and median absolute deviations. At the ends of the series, we repeated the first 5 years before 1 January 1772 and the last five after 31 December 2004. To apply standard methods for maxima, the series was then negated, the sequence of gaps calculated by choosing a threshold and applying the procedure outlined in Section 3; finally, we dropped the gaps ending in months from March through November.

The simulation studies suggested that above a threshold u for which $F(u) = 0.95$, the estimators behave well, and, as was shown directly by fitting generalized Pareto models to exceedances in the series $\{y_i\}$ for a range of thresholds, this gives acceptable thresholds for a valid extreme-value approximation. The threshold value $u = 2.0$ was chosen; this corresponds to empirical exceedance probability $q \approx 0.023$.

To check condition $D^{(2)}(u_n)$, we divided the series into four periods 1772–1830, 1830–1889, 1889–1948 and 1948–2004, and we computed the surface $p(u, r)$ with the values $F(u) = 0.95...0.995$ and $r = 4, 5, 6, 7, 12, 20$ for each period. Although the proportion of anti-$D^{(2)}(u_n)$ events is nearly 0.1, it remains constant for all the plane (u, r), so, for the sake of estimation, we supposed that the resulting bias will remain approximately constant and relatively small, thus allowing us to detect any nonstationarity. Comparisons of the results to the intervals and the iterative weighted least squares method might yield information on the bias.

4.3 Smoothing of the CET

Corresponding to a description by a fixed model with parameters varying with time, we suppose that there exists a limiting Poisson process of the exceedances for the Central England temperature data, but it is inhomogeneous, that is, the parameter θ is changing over time: at any point t the limiting distribution of the inter-exceedance spacings is $P\left\{\overline{F}(u_n; t) S(u_n) > s\right\} \to \theta(t) \exp(-\theta(t)s)$ for $s > 0$ as $n \to \infty$, where $\theta(t)$ is the extremal index, a smooth function of the time. We assume moreover that the change is slow in the sense that there is a finite time interval around any point t_i where this process can be well approximated by a homogeneous one with constant $\theta^*(t_i)$, and that $\theta^*(t_i) \approx \theta(t_i)$. A local constant estimator of the extremal index $\theta(t_i)$ can then be constructed by fitting a weighted likelihood model within the homogeneity intervals around each point. In some cases, better fit may be expected from a local polynomial model which traces more faithfully the changes in the estimated parameter (Fan and Gijbels 1996); however, in this case of nonstationarity the inhomogeneous Poisson process nature of the cluster locations is not assured by any theoretical result.

To compare the maximum likelihood, the intervals, and the iterative least squares estimators, we estimated the extremal index of the series with two different smoothing time intervals $h = 3{,}001$ and $7{,}301$ (approximately 33 and 80-year periods) with threshold 2.0. To model the temporal behaviour of the extremal index we applied a logit-transformed local linear form for θ as a function of time. We used direct optimization of the locally weighted log-likelihood

$$\ell^{(W)}(t_{i_0}) = \sum_{i=1}^{K} w(t_i)\ell(\theta(t_i)), \tag{3}$$

where K is the number of clusters in the interval centered on the cluster at time t_{i_0}, t_i are the locations of clusters, $\ell(\theta(t_i))$ is the likelihood contribution from the cluster beginning at t_i, and

$$\theta(t_i) = \frac{\exp\{\alpha_{t_{i_0}} + \beta_{t_{i_0}}(t_i - t_{i_0})\}}{1 + \exp\{\alpha_{t_{i_0}} + \beta_{t_{i_0}}(t_i - t_{i_0})\}}, \tag{4}$$

with $\alpha_{t_{i_0}}$ and $\beta_{t_{i_0}}$ as the parameters of the model in the smoothing interval centered at t_{i_0}. The weights $w(t_i)$ were given by a normal kernel. For this setup, a parametric bootstrap based on the fit $\{\hat{\alpha}_{t_{i_0}}, \hat{\beta}_{t_{i_0}}\}$ with 500 repetitions was used to assess the variability of the estimation. For each interval, we simulated the inhomogeneous Poisson process with intensity determined by $\theta(t_i)$, repeated the estimation for the samples, and took the 0.025- and 0.975-quantiles. The intervals and the iterative weighted least squares procedure was applied the same way as in Section 4.1.

Figure 4 shows the results obtained by local linear smoothing and intervals estimation for the 7,301-day smoothing interval and threshold 2.0. Results for the intervals and the iterative least squares estimations were fairly similar. The bias caused by the possible failure of $D^{(2)}(u_n)$ appears to be small. After about 1850, the two estimates are much the same, since their confidence bands are
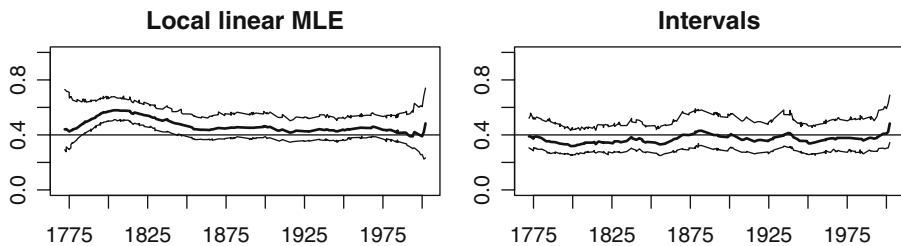

Fig. 4 Comparison of the weighted local linear MLE and intervals estimators with threshold $u_{(2)} = 2.0$. The *left panel* shows the extremal index for 7,301-day smoothing intervals by the local linear maximum likelihood estimator, the *right panel* are the estimates by the intervals method with the same window. *Heavy solid lines*: estimates, *thin solid*: 0.95 percentile bootstrap CI

strongly overlapping. Before 1850, there is a difference between the maximum likelihood and the intervals estimate: the maximum likelihood method shows a pronounced peak of $\theta > 0.6$ where the intervals (and the iterative least squares) give nearly constant values around $\theta = 0.4$. The difference might be due to the use of equal weights; using non-equal weights accentuates more the local characteristics of the time series around the center of the estimation window. With a long window, a local feature in the series is masked by the equally weighted farther periods in the intervals and the least squares estimation; in investigations with shorter window, these estimators too showed a peak here, though with a lower maximum. Another plausible explanation would be the condition $D^{(2)}(u_n)$ failing temporarily (indicating nonstationarity at least in this period), but this was not confirmed by the analysis of the anti-$D^{(2)}(u_n)$ events.

4.4 Test for Constant Extremal Index

For the period 1772–1980 both the intervals and the iterative least squares methods suggested the possibility of a constant extremal index with a value slightly below 0.5. For the local linear maximum likelihood estimation, this was checked by a bootstrap procedure: for threshold 2.0, we resampled the winter inter-cluster spacings and cluster sizes with equal probabilities, and performed weighted local linear maximum likelihood estimation on 4,000 repetitions, resulting in a collection of extremal index curves. These were then smoothed and their values on a common time grid were calculated to obtain pointwise and overall percentile confidence bands for the time-dependent curves. We show the result in Fig. 5. Although there is some suggestion of a weak downward tendency, an overall extremal index value of 0.5 seems plausible for most of the observed period, even for the period 1800–1830. Another peculiar interval is the few decades beginning around 1980, when the estimated extremal index has a strong upward tendency that is shown by all three estimation methods (and even with all window lengths tried), though it is not significant in Fig. 5.
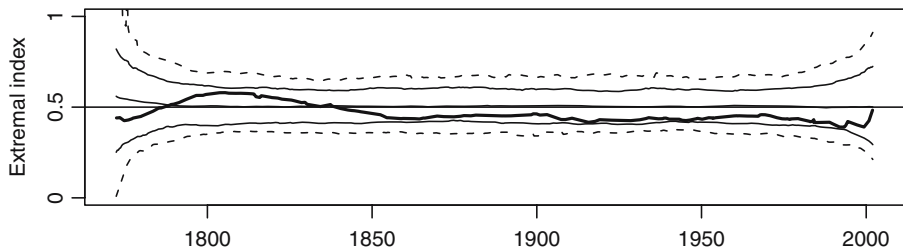

Fig. 5 Check of the hypothesis of constant extremal index with the threshold $u_{(2)} = 2.0$. *Heavy solid line*: estimated extremal index for the true sequence, *thin solid lines*: median, 0.025- and 0.975-percentile bootstrap confidence interval, *thin dashed lines*: 95% overall confidence interval, *thin straight line* at $\theta = 0.5$: plausible constant value for the extremal index

5 Discussion

The article proposes new estimators for the extremal index based on inter-exceedance times. These are a maximum likelihood estimator, and an iterative weighted least squares estimator based on the exponential quantile-quantile plot. Their good properties can be summarized as follows:

–   The maximum likelihood estimator is of closed form and so is very easy to implement in practice, while the iterative least squares method uses only standard statistical techniques, without the need for preliminary declustering or checking $D^{(2)}(u_n)$.
–   In addition to its simplicity, the maximum likelihood estimator has smaller bias and root mean squared error than the intervals, the two-threshold, and the runs methods on the class of processes satisfying condition $D^{(2)}(u_n)$.
–   As the maximum likelihood estimator is derived from a regular likelihood model, it can be used with smoothing methods, providing a tool for checking changes in clustering.
–   In the simulations with extremal index varying with time, the maximum likelihood estimator could follow the changes in θ more faithfully and with smaller variance.
–   In all simulations of stationary processes, the iterative least squares method showed the lowest sensitivity to threshold choice and very small bias.

Beside the advantages, there are some drawbacks:

–   The validity of the likelihood model is restricted by the condition $D^{(2)}(u_n)$.
–   Sensitivity to threshold choice persists for our estimators, although it is lower than for runs and two-threshold methods. The same is true, however, for any extremal analysis, and as in general, no simple solution can be proposed.

Future developments are still possible. Similar maximum likelihood models could be built on truncating the exponential part at a separation M and considering every spacing below M as within-cluster time. This would obviously require conditions similar to $D^{(2)}(u_n)$, which is the first of the hierarchy of conditions $D^{(k)}(u_n)$ of Chernick et al. (1991), and which might be checked by similar plots as $D^{(2)}(u_n)$. This extension could be regarded as a framework comprising a large class of estimators, included the runs estimator that can be derived as a maximum likelihood estimator from a penultimate approximation to the likelihood (1), and the intervals estimator, which is a moment estimator. Extensions to multivariate clustering can be imagined based on Nandagopalan (1994).

References

Chernick, M.R., Hsing, T., McCormick, W.P.: Calculating the extremal index for a class of stationary sequences. Adv. Appl. Probab. **23**, 835–850 (1991)

Fan, J., Gijbels, I.: Local Polynomial Modelling and Its Applications. Chapman & Hall, London (1996)

Ferro, C.A.T.: Statistical methods for clusters of extreme values. Ph.D. thesis, Lancaster University (2003)

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

Hsing, T.: Extremal index estimation for a weakly dependent stationary sequence. Ann. Stat. **21**, 2043–2071 (1993)

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

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

Leadbetter, M.R., Nandagopalan, S.: On exceedance point processes for stationary sequences under mild oscillation restrictions. Lect. Notes Stat. **51**, 69–80 (1989)

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

Nandagopalan, S.: On the multivariate extremal index. J. Res. Natl. Inst. Stand. Technol. **99**, 543–550 (1994)

Smith, R.L.: The extremal index for a Markov chain. J. Appl. Probab. **29**, 37–45 (1992)

Smith, R.L., Weissman, I.: Estimating the extremal index. J. R. Stat. Soc., Ser. B **56**, 515–528 (1994)

Weissman, I., Novak, S.Yu.: On blocks and runs estimators of the extremal index. J. Stat. Plan. Inference **66**, 281–288 (1998)