Automatic declustering of extreme values via an estimator for the extremal index

Christopher A. T. Ferro[†]
Johan Segers[‡]

May 20, 2002

Abstract. Inference for clusters of extreme values of a time series typically requires the identification of independent clusters of exceedances over a high threshold. The choice of declustering scheme often has a significant impact on estimates of cluster characteristics. We propose an automatic declustering scheme that is justified by an asymptotic result for the arrival times between threshold exceedances. The scheme relies on the extremal index, which we show may be estimated prior to declustering. The scheme also supports a bootstrap procedure for assessing the variability of estimates.

Keywords: automatic declustering, bootstrap, extremal index, extreme values, inter-arrival times

[†]Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, UK. E-mail: c.ferro@lancaster.ac.uk
[‡]Eurandom, P.O.Box 513, NL-5600 MB Eindhoven, The Netherlands.
E-mail: segers@eurandom.tue.nl

1 Introduction

Let $\{\xi_n\}_{n\geq1}$ be a strictly stationary sequence of random variables with marginal distribution function F, finite or infinite right end-point $\omega = \sup\{x : F(x) < 1\}$ and tail function $\bar{F} = 1 - F$. For integers $0 \leq k < l$ and $n \geq 1$, put

$$M_{k,l} = \max\{\xi_i : i = k+1, \ldots, l\} \quad \text{and} \quad M_n = M_{0,n}.$$

We say that the process has extremal index $\theta \in [0,1]$ if for each $\tau > 0$ there exists a sequence $\{u_n\}_{n\geq1}$ such that, as $n \to \infty$,

(i) $n\bar{F}(u_n) \to \tau$,

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

see Leadbetter *et al.* (1983). If $\theta = 1$ then exceedances of an increasing threshold occur singly in the limit; if $\theta < 1$ then exceedances tend to cluster in the limit. We consider the problem of making inferences about characteristics of such clusters. The extremal index is one such characteristic, which has an interpretation due to Leadbetter (1983) as the reciprocal of the mean cluster size.

Theorem 4.5 of Hsing (1987) shows that clusters of exceedances may be considered independent in the limit. Consequently, a common approach to inference is to identify independent clusters of exceedances above a high threshold, evaluate for each cluster the characteristic of interest, and form estimates from these values. The methods used to identify clusters define different estimators, the two most common being blocks and runs declustering used by Leadbetter *et al.* (1989). Runs declustering, for example, assumes that exceedances belong to the same cluster if they are separated by fewer than a certain number (the run length) of values below the threshold. Hsing (1991) points out that a problem with these estimators is the selection of the declustering parameters, which is largely arbitrary: the choice of block length or run length usually has a significant influence on the estimate of the cluster characteristic. Estimation of the extremal index is developed by Leadbetter and Nandogopalan (1989), Smith and Weissman (1994) and Weissman and Novak (1998) among others; for an application in finance see Longin (2000). Examples of other cluster characteristics are the cluster maximum, which is the focus of peaks-over-threshold (POT) modelling (reviewed by Davison and Smith, 1990), and the 'excess height statistic' used by Leadbetter (1995) to monitor ozone levels.

We investigate the point process of exceedance times (see Hsing *et al.*, 1988) and find that the asymptotic distribution of the inter-arrival times belongs to a one-dimensional parametric family of distributions indexed by the extremal index. This result provides a limiting argument for one particular declustering scheme, characterised by the extremal index. We can estimate the extremal index without declustering by equating theoretical moments of the limiting distribution to their empirical counterparts. In this way we define an automatic declustering scheme that does not require a subjective choice of auxiliary parameter. Furthermore, the declustering scheme supports a bootstrap procedure for obtaining confidence intervals on estimates of cluster characteristics that accounts for the uncertainty in the scheme's estimation.

We derive the asymptotic distribution of times between threshold exceedances in Section 2 and present our estimator for the extremal index in Section 3. We prove consistency of the estimator for m-dependent processes. In Section 4 we define the declustering scheme and bootstrap procedure. We investigate the performance of our extremal-index estimator with a simulation study in Section 5, and conclude in Section 6 with an application to a series of daily minimum temperatures recorded at Wooster, Ohio.

2 Inter-arrival times

The statistical development in this paper rests on the limiting distribution of the times between exceedances of a threshold u by the process $\{\xi_n\}_{n\geq 1}$. Let $T(u)$ be a random variable with distribution

$$T(u) \stackrel{d}{=} \min\{n \geq 1 : \ \xi_{n+1} > u\} \quad \text{given } \xi_1 > u,$$

that is,

$$P\{T(u) = n\} = P(M_{1,n} \leq u, \xi_{n+1} > u \mid \xi_1 > u) \quad \text{for } n \geq 1,$$

or, alternatively,

$$P\{T(u) > n\} = P(M_{1,n+1} \leq u \mid \xi_1 > u) \quad \text{for } n \geq 1.$$

We compute the asymptotic distribution of $T(u)$. The case when the random variables ξ_n, $n \geq 1$, are independent, is straightforward. Clearly

$$P\{T(u) > n\} = F(u)^n \quad \text{for } n \geq 1,$$

so that, for $x > 0$,

$$P\{\bar{F}(u)T(u) > x\} = P\{T(u) > \lfloor x/\bar{F}(u) \rfloor\} = \exp\{\lfloor x/\bar{F}(u) \rfloor \log F(u)\}.$$

Finally, if F has no atom at its end-point ω then, since $\log(1+\epsilon) \sim \epsilon$ as $\epsilon \to 0$, we have

$$\lim_{u \uparrow \omega} P\{\bar{F}(u)T(u) > x\} = e^{-x}, \quad \text{for } x > 0.$$

So $\bar{F}(u)T(u)$ is asymptotically standard exponentially distributed. This agrees with the result (see Hsing *et al.*, 1988) hat the point process of exceedance times has a Poisson-process limit.

Now consider the general case with extremal index $\theta \in [0,1]$. The corresponding point-process limit for the exceedance times is compound Poisson (see Hsing *et al.*, 1988). This leads us to expect that the limit distribution of the inter-arrival times will be a mixture of an exponential distribution and a point-mass on zero. This is indeed the case, as described in Theorem 2.1 below. For real u and integers $1 \leq k \leq l$, let $\mathcal{F}_{k,l}(u)$ be the σ-field generated by the events $\{\xi_i > u\}$, $k \leq i \leq l$. Define the mixing coefficients

$$\alpha_{n,q}(u) = \max_{1 \leq k \leq n-q} \sup |P(B \mid A) - P(B)|,$$

where the supremum is over all $A \in \mathcal{F}_{1,k}(u)$ with $P(A) > 0$ and all $B \in \mathcal{F}_{k+q,n}(u)$.

Theorem 2.1 *Let the positive integers r_n, $n \geq 1$, and the thresholds u_n, $n \geq 1$, be such that*

$$r_n \to \infty, \qquad r_n\bar{F}(u_n) \to \tau, \quad \text{and} \quad P(M_{r_n} \leq u_n) \to e^{-\theta\tau},$$

for some $\tau \in (0,\infty)$ and $\theta \in [0,1]$. If there exist positive integers $q_n = o(r_n)$ such that $\alpha_{cr_n,q_n}(u_n) = o(1)$ for all $c > 0$, then

$$P\{\bar{F}(u_n)T(u_n) > t\} \to \theta e^{-\theta t}, \quad \text{for } t > 0.$$

The proof is contained in Appendix A. Note that the mixing condition used here is similar to that of Weissman and Novak (1998).

Theorem 2.1 says that

$$\bar{F}(u)T(u) \xrightarrow{d} T_\theta \quad \text{as } u \uparrow \omega, \tag{1}$$

where T_θ is a random variable distributed according to the mixture distribution

$$(1-\theta)\epsilon_0 + \theta\mu_\theta, \tag{2}$$

ϵ_0 is the degenerate probability distribution at 0 and μ_θ is the exponential distribution with mean θ^{-1}. Note the dual role of the extremal index: θ is both the proportion of non-zero inter-arrival times and the reciprocal of the mean of the non-zero inter-arrival times. These features are evident in Fig. 1, which was constructed from the times between the largest 1000 values in a sequence of length 10000 generated from a stationary process with extremal index $\theta = 0.5$.

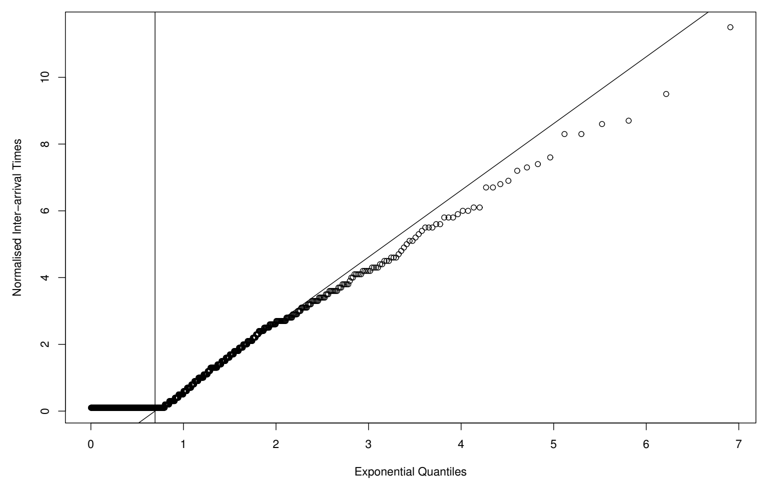

Figure 1: Quantile-quantile plot of normalised inter-arrival times against unit-exponential quantiles. The vertical line indicates the $(1-\theta)$-quantile, the sloping line has gradient θ^{-1}, and $\theta = 0.5$.

3 Extremal-index estimation

In this section we describe an estimator for the extremal index that is based on the limit result (1). Suppose that we have a sample $\xi_1, \ldots, \xi_n$ and a high threshold u. Let $N = N_n(u) = \sum_{i=1}^n I(\xi_i > u)$ be the number of observations exceeding u, and let

$$1 \leq S_1 < \cdots < S_N \leq n$$

be the exceedance times. The observed inter-arrival times are

$$T_i = S_{i+1} - S_i, \quad \text{for } i = 1, \ldots, N-1.$$

3.1 Intervals estimator

The second moment of T_θ is $E(T_\theta^2) = 2/\theta$, which we estimate to obtain a first estimator for θ. Let $\bar{F}_n(u)$ be an estimator for $\bar{F}(u)$. Then

$$\bar{\theta}_n(u) = \frac{2(N-1)}{\bar{F}_n(u)^2 \sum_{i=1}^{N-1} T_i^2}.$$

For example, if $\bar{F}_n(u) = N/n$, then we have

$$\bar{\theta}_n(u) = \frac{2n^2(N-1)}{N^2 \sum_{i=1}^{N-1} T_i^2}.$$

We can improve on this however. The first moment of T_θ is one so that θ is related to the coefficient of variation, ν, of the inter-arrival times by

$$1 + \nu^2 = E(T_\theta^2)/\{E(T_\theta)\}^2 = 2\theta^{-1}.$$

In particular, the exceedance times are over-dispersed (there is clustering in the limit) if and only if the extremal index is less than one. This relationship motivates another estimator for θ,

$$\widehat{\theta}_n(u) = \frac{2\left(\sum_{i=1}^{N-1} T_i\right)^2}{(N-1)\sum_{i=1}^{N-1} T_i^2}, \tag{3}$$

where we do not need to estimate $\bar{F}(u)$.

Finally, let us make one more improvement to our estimator by considering the penultimate approximation to the limiting mixture distribution (2). If we set $r_n = n$ in the proof of Theorem 2.1, then we see that the distribution of the inter-arrival times satisfies

$$P\{T(u_n) > n\} = \theta F(u_n)^{\theta n} + o(1).$$

We can derive an estimator for θ based on this relationship. Let T denote a random variable on the positive integers whose distribution is given by

$$P(T > n) = \theta p^{\theta n}, \quad \text{for } n = 1, 2, \dots \tag{4}$$

where $\theta \in (0, 1]$ and $p \in (0, 1)$ may be thought of as $F(u_n)$.

First, note that

$$\begin{aligned} \frac{2\{E(T)\}^2}{E(T^2)} &= \frac{2\{1-(1-\theta)p^\theta\}^2}{2\theta p^\theta + \theta p^\theta(1-p^\theta) + (1-p^\theta)^2} \\ &= \theta + \theta(2 - 3\theta/2)(1-p) + O\left\{(1-p)^2\right\}, \quad \text{as } p \to 1, \end{aligned}$$

which follows from

$$\begin{aligned} E(T-1) &= \sum_{n=1}^{\infty} P(T > n) = \theta p(1-p)^{-1}, \\ E\left\{\frac{T(T-1)}{2}\right\} &= \sum_{n=1}^{\infty} nP(T > n) = \theta p(1-p)^{-2}. \end{aligned}$$

Therefore, the first-order bias of $\widehat{\theta}_n(u)$ at a threshold u is approximately $\theta(2 - 3\theta/2)\bar{F}(u)$. In contrast, the relationship

$$\frac{2\{E(T-1)\}^2}{E\{(T-1)(T-2)\}} = \theta$$

motivates the estimator

$$\widetilde{\theta}_n(u) = \frac{2\left\{\sum_{i=1}^{N-1}(T_i - 1)\right\}^2}{(N-1)\sum_{i=1}^{N-1}(T_i-1)(T_i-2)}, \tag{5}$$

whose first-order bias is approximately zero. We call $\widetilde{\theta}_n(u)$ the intervals estimator. Compared to estimator (3), the estimators for the first and second moments from which the intervals estimator is constructed have been shrunk towards zero. This is intuitively appealing. The smallest observed inter-arrival times are positive, whereas the limiting distribution (2) models them as zero. The intervals estimator ensures that the contributions from the smallest inter-arrival times are indeed zero; the larger inter-arrival times are relatively unaffected.

Consistency of the intervals estimator for m-dependent processes is stated in the following theorem, the proof of which is contained in Appendix B. Recall that, for a positive integer m, the sequence $\{\xi_n\}_{n\geq 1}$ is m-dependent if, for all positive integers k, the σ-fields $\sigma(\xi_i : 1 \leq i \leq k)$ and $\sigma(\xi_i : i \geq k+m)$ are independent.

Theorem 3.1 *Let the positive integers r_n, $n \geq 1$, and the thresholds u_n, $n \geq 1$, be such that*

$$r_n = o(n), \quad r_n \to \infty, \quad r_n \bar{F}(u_n) \to \tau, \quad \text{and} \quad P(M_{r_n} \leq u_n) \to e^{-\theta\tau},$$

for some $\tau \in (0,\infty)$ and $\theta \in (0,1]$. If $\{\xi_n\}_{n\geq 1}$ is m-dependent, then $\widetilde{\theta}_n(u_n) \xrightarrow{p} \theta$.

3.2 Maximum likelihood estimation

We have described two models (the limiting form (2) and the penultimate approximation (4)) for the inter-arrival times. In this section we examine the possibility of using these models to construct maximum likelihood estimators for the extremal index. This contrasts with the moment-based intervals estimator (5). We shall construct the likelihoods under the assumption that the inter-arrival times are independent since we have models for their marginal distribution only. This is an incorrect assumption but does not affect the validity of maximum likelihood point estimates.

If we write $t_i = NT_i/n$ then the log-likelihood from the limiting model (2) is

$$\sum_{i=1}^{N-1} \log\left\{(1-\theta)^{I(t_i=0)}\left(\theta^2 e^{-\theta t_i}\right)^{I(t_i>0)}\right\} = 2(N-1)\log\theta - \theta\sum_{i=1}^{N-1} t_i$$

since the observed inter-arrival times are always strictly positive. The resulting maximum likelihood estimator for θ is $\min\{1, 2/\bar{t}\}$, where $\bar{t}$ is the mean of the t_i. This estimator tends in distribution to one as n increases: a failure arising from the model assigning all of the inter-arrival times to the exponential component of the mixture distribution. It is possible to circumvent this problem by grouping a certain number of the smallest inter-arrival times and treating them as though they were equal to zero. This solution requires the choice of an auxiliary parameter however. The choice is arbitrary and can have a significant impact on estimates so that we are no better off than if we had used runs or blocks declustering.

The log-likelihood from the penultimate model (4) is

$$m_1 \log(1-\theta p^\theta) + \{\log\theta + \log(1-p^\theta)\}\sum_{i\geq 2} m_i + \theta \log p \sum_{i\geq 2}(i-1)m_i,$$

where m_i is the number of inter-arrival times equal to i. Maximum likelihood estimates may be found by numerical optimisation. Unfortunately, this likelihood performs poorly in practice. The reason for this is that the distribution (4) is a good model for the large inter-arrival times only; at the same time, for p close to one, m_1 has a strong influence on the likelihood. Again, grouping small inter-arrival times and modelling only the larger times is a possible, but unattractive, solution.

The problems encountered above make the likelihood approach difficult to implement so that we prefer the intervals estimator.

4 Automatic declustering and bootstrapping

While we have shown how the extremal index may be estimated without recourse to declustering, estimating other cluster characteristics may require clusters to be identified. All declustering schemes proposed in the literature require an auxiliary parameter, the choice of which is largely arbitrary. In this section we explain how the limiting distribution (2) may be used to identify clusters without making an arbitrary choice. As we shall see, declustering also supports a bootstrap procedure for assessing estimation uncertainty; we describe a procedure to compute confidence limits for estimates of general cluster characteristics. Importantly, this procedure accounts for the uncertainty in the choice of declustering scheme.

Recall that the limiting process of exceedance times is a compound Poisson process so that we may categorise inter-arrival times into two types: independent inter-cluster times (between clusters), and independent sets of intra-cluster times (within clusters). As mentioned below (2), the extremal index is the proportion of inter-arrival times that may be regarded as inter-cluster times. Suppose that we observe N exceedance times, $S_1 < \cdots < S_N$, and $T_i = S_{i+1} - S_i$ for $i = 1, \ldots, N-1$ are the inter-arrival times. Then we can assume that the largest $C - 1 = \lfloor \theta N \rfloor$ inter-arrival times are approximately independent inter-cluster times that divide the remainder into approximately independent sets of intra-cluster times. To be precise, if $T_{(C)}$ is the C-th largest inter-arrival time and T_{i_j} is the j-th inter-arrival time to exceed $T_{(C)}$, then $\{T_{i_j}\}_{j=1}^{C-1}$ is a set of approximately independent inter-cluster times. (In the case of ties, decrease C until $T_{(C-1)}$ is strictly greater than $T_{(C)}$.) Let $\mathcal{T}_j = \{T_{i_{j-1}+1}, \ldots, T_{i_j - 1}\}$, where $i_0 = 0$, $i_C = N$ and $\mathcal{T}_j = \emptyset$ if $i_j = i_{j-1} + 1$. Then $\{\mathcal{T}_j\}_{j=1}^{C}$ is a collection of approximately independent sets of intra-cluster times. Furthermore, each set $\mathcal{T}_j$ has associated with it a set of threshold exceedances, $\mathcal{C}_j = \{\xi_k : k \in \mathcal{S}_j\}$, where $\mathcal{S}_j = \{S_{i_{j-1}+1}, \ldots, S_{i_j}\}$.

This interpretation justifies a decomposition of the observed process into C clusters, where the j-th cluster comprises the exceedances $\mathcal{C}_j$. This is equivalent to runs declustering with run length $T_{(C)}$. In practice, C is defined by replacing θ with the intervals estimate (5) so that we have an entirely automatic declustering procedure, justified by the limiting theory.

Suppose now that we are interested in making inferences about a cluster functional, H. This could be the 'excess height statistic' of Leadbetter (1995) for example. We can evaluate the functional for each cluster to obtain values $\{H_j\}_{j=1}^{C}$ that may be used to estimate properties of H. Denote such estimates by $\bar{H}$. For example, we might estimate the expectation of H by $\bar{H} = C^{-1}(H_1 + \cdots + H_C)$.

We can use the bootstrap to obtain confidence limits on such estimates. Given the decomposition of the process of exceedances described above, we recommend the following procedure:

(a) Resample with replacement $C-1$ inter-cluster times from $\{T_{i_j}\}_{j=1}^{C-1}$.

(b) Resample with replacement C sets of intra-cluster times (some of which may be empty) and associated exceedances from $\{(\mathcal{T}_j, \mathcal{C}_j)\}_{j=1}^{C}$.

(c) Intercalate these inter-arrival times and clusters to form a bootstrap replication of the process.

(d) Compute N, $\widehat{\theta}$ and C for the bootstrap process and decluster accordingly.

(e) Compute $\bar{H}$ from the declustered bootstrap process.

Forming B such bootstrap processes yields collections of estimates, $\widehat{\theta}_{(1)}, \ldots, \widehat{\theta}_{(B)}$ and $\bar{H}_{(1)}, \ldots, \bar{H}_{(B)}$, that may be used to approximate the distributions of the original point estimates, $\widehat{\theta}$ and $\bar{H}$. In particular, the empirical α- and $(1-\alpha)$-quantiles of each sample define $(1-2\alpha)$-confidence intervals. We demonstrate this procedure for the data-set in Section 6, and make use of it in the simulation study of the following section.

5 Simulation study

In this section we investigate the performance of the intervals estimator (5) for the extremal index and compare it to the performance of the runs estimator, which may be written in terms of the inter-arrival times as

$$\widehat{\theta}_n(u;r) = N^{-1}\left\{\sum_{i=1}^{N-1} I(T_i > r) + 1\right\}$$

when the run length is r. We simulate data from two stationary processes: a max-autoregressive process and a first-order Markov chain with extreme-value transition distribution.

Choose $\theta \in (0,1]$. Let W_n, $n \geq 1$, be independent unit Fréchet random variables and put

$$\xi_1 = W_1/\theta \quad \text{and} \quad \xi_n = \max\{(1-\theta)\xi_{n-1}, W_n\}, \quad \text{for } n \geq 2. \tag{6}$$

Then $\{\xi_n\}_{n\geq 1}$ is a max-autoregressive process with extremal index θ. For the Markov chain, let $\beta \in (0,1]$ and

$$P(\xi_1 \leq x_1, \xi_2 \leq x_2) = \exp\left\{-\left(x_1^{-1/\beta} + x_2^{-1/\beta}\right)^{\beta}\right\}.$$

This is a bivariate extreme-value distribution with symmetric logistic dependence structure; the extremal index of $\{\xi_n\}_{n\geq 1}$ for specific values of β may be found by simulation (see Smith *et al.*, 1997). We simulate 1000 sequences of length 5000 from both of these processes for each of three extremal indices: 0.25, 0.5 and 0.75. (This corresponds to $\beta = 0.43, 0.64$ and 0.82 for the Markov chain.) For each sequence, we compute the intervals estimator and three runs estimators (with run lengths 1, 5 and 9) at a range of thresholds chosen so that there are N exceedances, with N ranging from 10 to 1000. Note that the intervals estimators can give estimates that are greater than one. Hereafter we set estimates equal to one if this is the case. We report the root-mean-square error (rmse) of the point estimates and the coverage probability of bootstrapped confidence limits (computed with $B = 1000$) in each case.

Fig. 2 shows the empirical root-mean-square errors and coverage probabilities for the four estimators applied to the max-autoregressive processes. The sensitivity of the runs estimator to run length, threshold and the true extremal index is clear; the intervals estimator is robust to these factors. This robustness is particularly evident at lower thresholds, where the coverage probability for the intervals estimator remains close to the nominal value. Nevertheless, for high thresholds, the runs estimators appear to outperform the intervals estimator. Fig. 3 presents the results for the Markov chains. Similar comments can be made here with regard to the robustness of the intervals estimator. Note, however, that not every choice of run length for the runs estimator leads to performance superior at high thresholds to that of the intervals estimator. This further demonstrates the benefit of the intervals estimator.

6 Data example

We conclude with an application of the intervals estimator and intervals declustering to a time-series of negated daily minimum temperatures, recorded to the nearest degree Fahrenheit, at Wooster, Ohio. See Smith *et al.* (1997) for a description of the data and additional analysis. We shall estimate the extremal index and mean cluster excess of the series at thresholds $u = 0.5, 1.5, \ldots, 14.5$. For some threshold u and some cluster $\{\xi_k : k \in \mathcal{S}\}$ of exceedances, the cluster excess is defined by $\sum_{k\in\mathcal{S}}(\xi_k - u)$, and is an indicator for the severity of a cold period.

The intervals estimates of the extremal index are shown in Fig. 4 with bootstrap 95% confidence limits. For comparison, runs estimates with run length five days are also shown. As expected, the variance of the intervals estimator is greater than that of the runs estimator, which does not account for uncertainty in the run length. Both plots support an extremal index of about 0.6, which is similar to the values found by Smith *et al.* (1997).

The estimates of the mean cluster excess are presented in the left-hand plot of Fig. 5. The results are not clear-cut although the graph appears to level out at a threshold of ten degrees and a mean cluster excess of about six degrees.

Finally, returning to the bootstrap procedure of Section 4, notice that each bootstrapped process has an associated run length, $r_{(b)} = T_{(C_{(b)})}$, where $C_{(b)} = \lceil \widehat{\theta}_{(b)} N_{(b)} \rceil$ and $N_{(b)}$ is the number of exceedances in the process. At each threshold, the collection $\{r_{(b)}\}_{b=1}^{B}$ describes the uncertainty in the declustering scheme and indicates the extent of the uncertainty ignored by using runs declustering with an arbitrary run length. The right-hand plot in Fig. 5 summarises the run length distribution. A run length of about one week is suggested with the interpretation that cold periods separated by more than one week can be considered independent.

Acknowledgements

The authors would like to thank Jonathan Tawn for helpful comments and The Bernoulli Society for organising the Gothenburg meeting (December 10–16, 2001) of the Séminaires Européens de Statistiques, which catalysed the work presented in this paper. The authors are pleased to acknowledge the financial support of The Engineering and Physical Sciences Research Council (CATF), The Fund for Scientific Research, Flanders (JS) and The Department

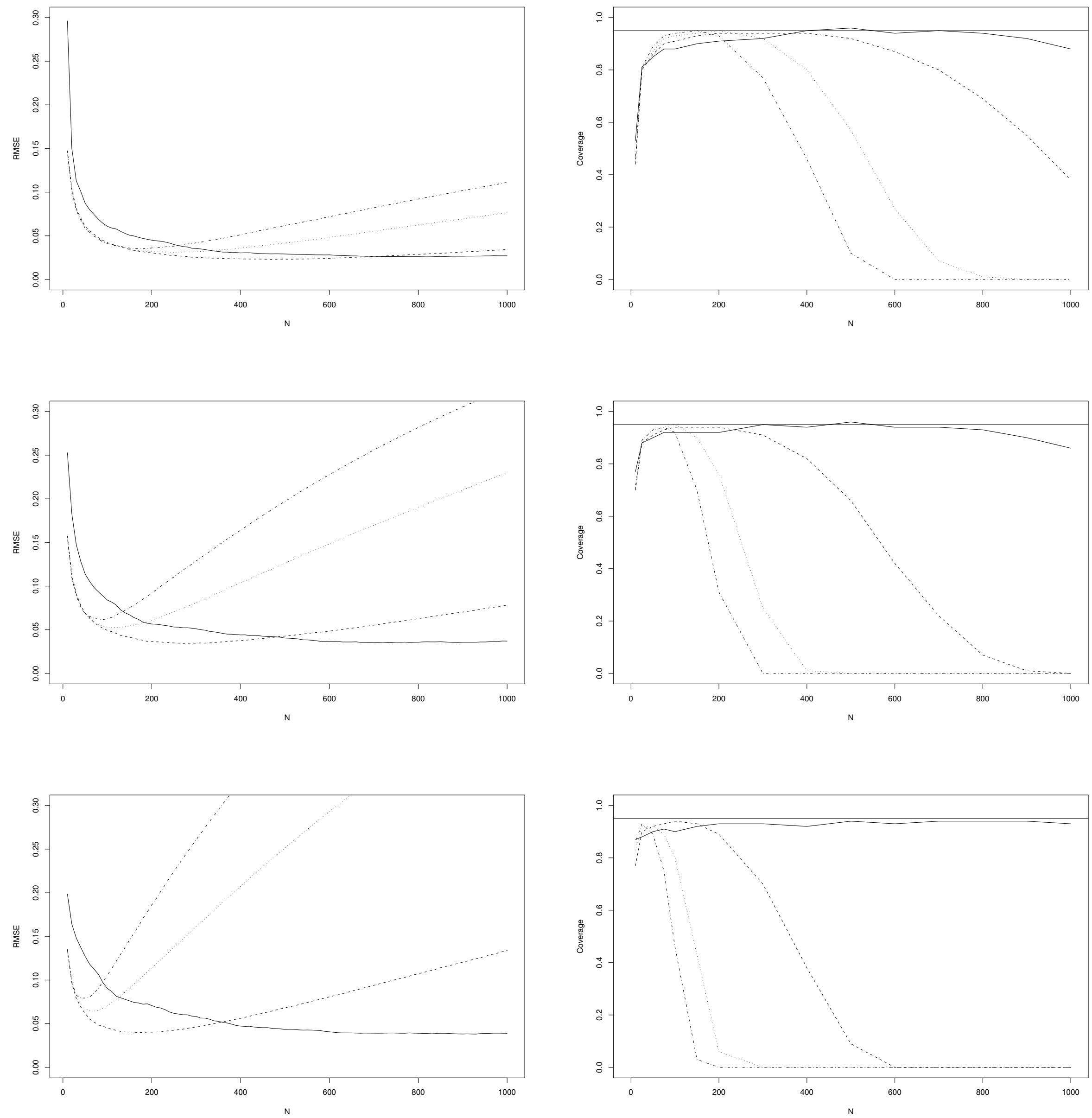

Figure 2: Root-mean-square errors (left) and coverage probabilities (right) for the intervals estimator (—) and runs estimators with run lengths 1 (- - -), 5 (· · ·) and 9 (-·-·-) applied to max-autoregressive sequences with extremal indices 0.25 (top), 0.5 (middle) and 0.75 (bottom). The nominal coverage probability (0.95) is indicated by a horizontal line.

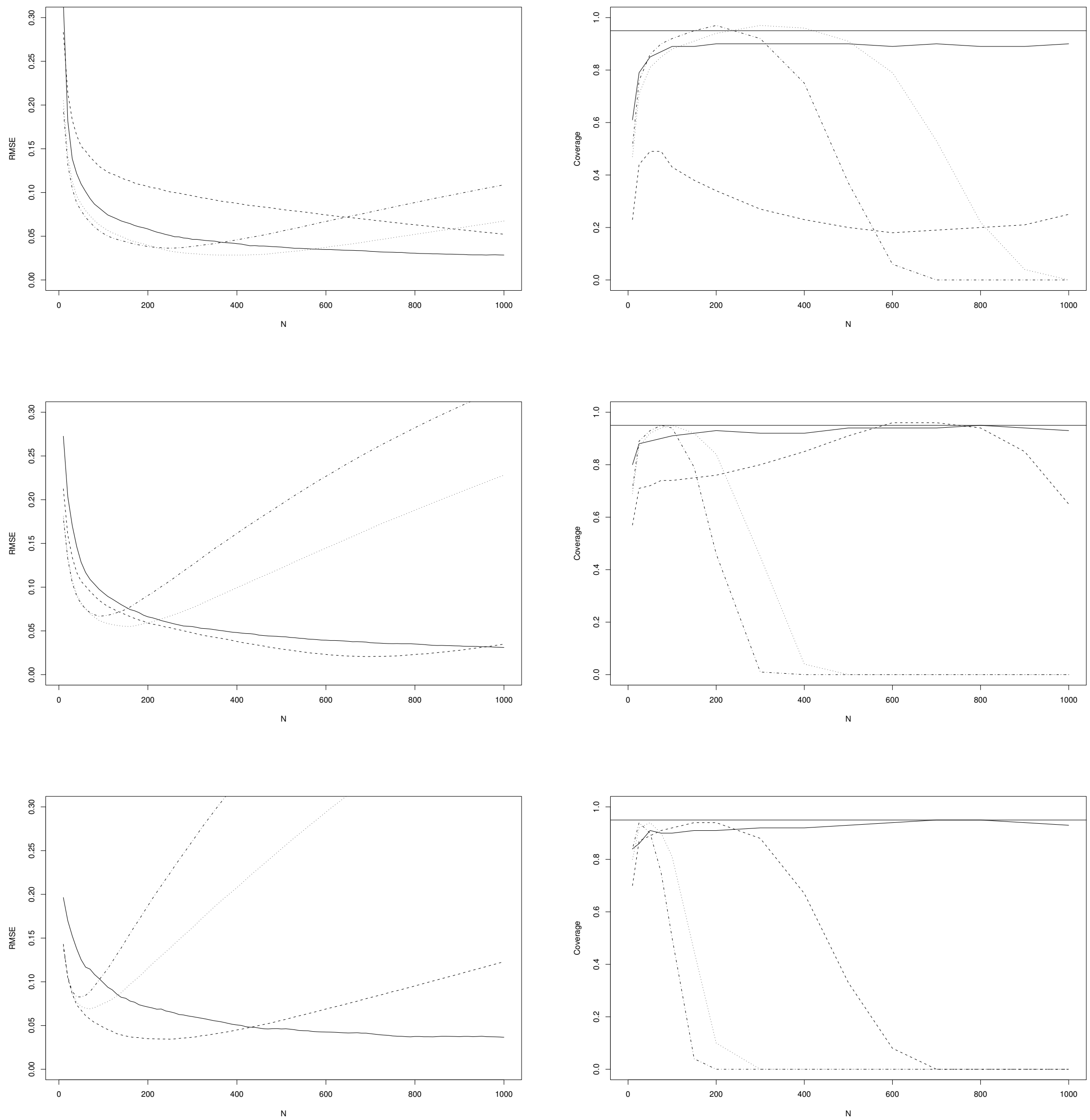

Figure 3: Root-mean-square errors (left) and coverage probabilities (right) for the intervals estimator (—) and runs estimators with run lengths 1 (- - -), 5 (· · ·) and 9 (-·-·-) applied to Markov chain sequences with extremal indices 0.25 (top), 0.5 (middle) and 0.75 (bottom). The nominal coverage probability (0.95) is indicated by a horizontal line.

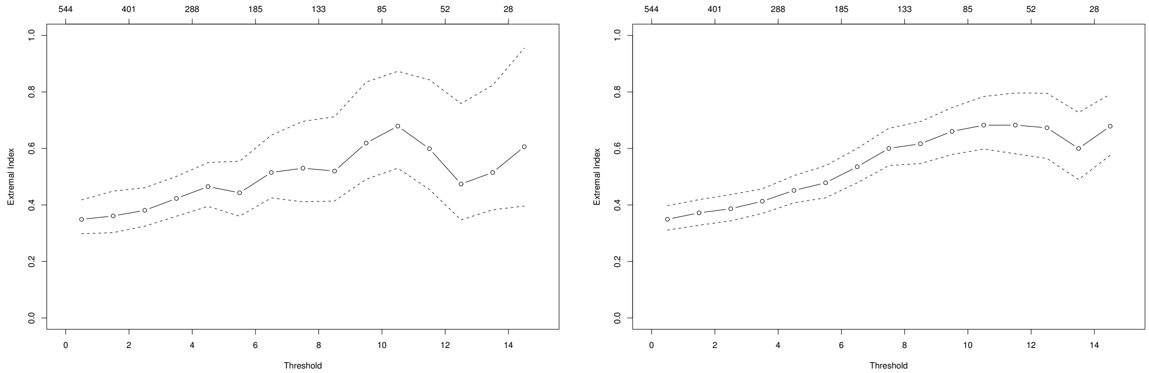

Figure 4: Estimates (–○–) of the extremal index for the Wooster temperature series. The intervals estimator (left) and the runs estimator with run length five (right) are shown with bootstrapped 95% confidence limits (- - -). The number of exceedances at each threshold is indicated on the upper axis.

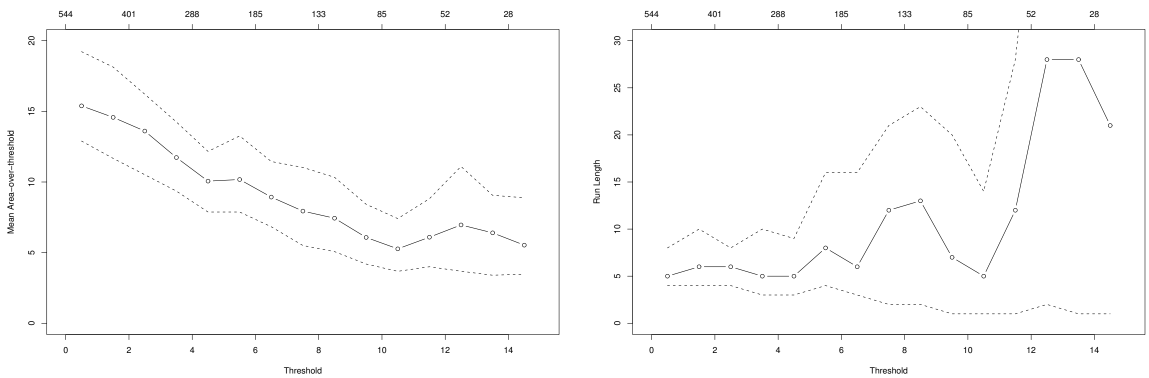

Figure 5: Estimates (–○–) of mean cluster excess (left) and run length (right) for the Wooster temperature series. Bootstrapped 95% confidence limits (- - -) are also shown. The number of exceedances at each threshold is indicated on the upper axis.

of Mathematics and Statistics at Lancaster University for sponsoring a visit of JS.

A Proof of Theorem 2.1

Since both $q_n = o(r_n)$ and $\alpha_n = \alpha_{r_n,q_n}(u_n) = o(1)$, we can find positive integers p_n such that

$$p_n = o(r_n), \quad r_n\alpha_n = o(p_n) \quad \text{and} \quad q_n = o(p_n).$$

(Take for instance $p_n = \lfloor \{r_n \max(q_n, r_n\alpha_n)\}^{1/2} \rfloor$.) Since $F(u_n)^{r_n} \to e^{-\tau}$, all conditions of Corollary 2.3 of O'Brien (1987) are fulfilled. We derive

$$P(M_{r_n} \leq u_n) - F(u_n)^{r_n P(M_{1,p_n} \leq u_n | \xi_1 > u_n)} \to 0.$$

But $P(M_{r_n} \leq u_n) \to e^{-\theta\tau}$ by assumption, so necessarily

$$P(M_{1,p_n} \leq u_n \mid \xi_1 > u_n) \to \theta.$$

Now we can proceed with the proof of the stated limit relation itself. Put $k_n = \lfloor t/\bar{F}(u_n) \rfloor$. We have

$$\begin{aligned} P\{\bar{F}(u_n)T(u_n) > t\} &= P\{T(u_n) > k_n\} \\ &= P(M_{1,1+k_n} \le u_n \mid \xi_1 > u_n). \end{aligned}$$

For n large enough so that $q_n < p_n$, we have

$$\begin{aligned} P(M_{p_n,p_n+q_n} > u_n \mid \xi_1 > u_n) &\le P(M_{p_n,p_n+q_n} > u_n) + \alpha_n \\ &\le q_n\bar{F}(u_n) + \alpha_n \to 0, \end{aligned}$$

so that

$$\begin{aligned} &P(M_{1,1+k_n} \le u_n \mid \xi_1 > u_n) \\ &\quad = P(M_{1,p_n} \le u_n,\, M_{p_n+q_n,k_n} \le u_n \mid \xi_1 > u_n) + o(1). \end{aligned}$$

But $k_n \sim t\tau^{-1}r_n = O(r_n)$, so that by assumption

$$\begin{aligned} &P(M_{1,p_n} \le u_n,\, M_{p_n+q_n,k_n} \le u_n \mid \xi_1 > u_n) \\ &\quad = P(M_{p_n+q_n,k_n} \le u_n \mid M_{1,p_n} \le u_n,\, \xi_1 > u_n)P(M_{1,p_n} \le u_n \mid \xi_1 > u_n) \\ &\quad = \{P(M_{p_n+q_n,k_n} \le u_n) + o(1)\}\{\theta + o(1)\}. \end{aligned}$$

Now also $s_n = k_n - (p_n + q_n) \sim t\tau^{-1}r_n = O(r_n)$. Apply Corollary 2.3 of O'Brien (1987) again to deduce that

$$\begin{aligned} P(M_{p_n+q_n,k_n} \le u_n) &= P(M_{s_n} \le u_n) \\ &= F(u_n)^{s_n P(M_{1,p_n} \le u_n \mid \xi_1 > u_n)} + o(1) \to e^{-\theta t}. \end{aligned}$$

□

B Proof of Theorem 3.1

Denote $\bar{F}(u_n)$ by p_n. To avoid ambiguities, let $\widetilde{\theta}_n(u_n)$ be an arbitrary number, say 1, if either $N \in \{0, 1\}$ or if $N \ge 1$ but no T_i is larger than two.

Lemma B.1 *There exist $C > 0$, $0 < \gamma < 1$ and positive integer n_0 such that*

$$\begin{aligned} (i) \qquad P(M_k \le u_n) &\le C\gamma^{k/r_n} \\ (ii) \quad P(M_{1,k} \le u_n \mid \xi_1 > u_n) &\le C\gamma^{k/r_n} \end{aligned}$$

for all integer $k \ge 1$ and $n \ge n_0$.

Proof. (i) Fix positive integer k and write $k = qm+j$ where $q = 0, 1, 2, \ldots$ and $j = 0, 1, \ldots, m-1$. Divide the sample of size k into q blocks of size m and one block of size j. By m-dependence, we get

$$P(M_k \le u_n) \le (1 - p_n)^q \le (1 - p_n)^{-j/m} \left\{(1 - p_n)^{r_n/m}\right\}^{k/r_n}.$$

Since $1-p_n \to 1$ and $(1-p_n)^{r_n} \to e^{-\tau} < 1$, we can find $0 < \gamma < 1$ and a positive integer n_0 such that for all integer $n \geq n_0$, we have

$$\begin{aligned} (1-p_n)^{-j/m} &\leq (1-p_n)^{-1} \leq 2 \quad \text{for } j = 0, 1, \ldots, m-1, \\ (1-p_n)^{r_n/m} &\leq \gamma. \end{aligned}$$

Hence for all integer $n \geq n_0$, we have $P(M_k \leq u_n) \leq 2\gamma^{k/r_n}$.

(ii) For integer $k \geq m+1$, we have by m-dependence and stationarity

$$P(M_{1,k} \leq u_n \mid \xi_1 > u_n) \leq P(M_{m,k} \leq u_n) = P(M_{k-m} \leq u_n).$$

Thanks to (i), the latter is bounded by $2\gamma^{-m}\gamma^{k/r_n}$. For $k \leq m$, we have $2\gamma^{-m}\gamma^{k/r_n} \geq 1$. Put $C = 2\gamma^{-m}$. □

Lemma B.2 *If the positive integers s_n satisfy $s_n \sim \lambda r_n$ for some $0 < \lambda < \infty$, then*

$$P(M_{s_n} \leq u_n) \to e^{-\theta\lambda\tau} \quad \textit{and} \quad P(M_{1,s_n} \leq u_n \mid \xi_1 > u_n) \to \theta e^{-\theta\lambda\tau}.$$

Proof. Similar to the proof of Theorem 2.1. □

By convention, we set $\sum_{i=1}^{N-1}(T_i - 1) = 0$ if $N \in \{0, 1\}$.

Lemma B.3 *We have $N \overset{p}{\sim} np_n$ and $\sum_{i=1}^{N-1}(T_i - 1) \overset{p}{\sim} n$.*

Proof. Clearly $E(N) = np_n$. By stationarity and m-dependence, also $\text{var}(N) \leq 2mnp_n$. Together, $E[\{(np_n)^{-1}N - 1\}^2] \leq 2m(np_n)^{-1} \to 0$, which implies $N \overset{p}{\sim} np_n$. In particular, we have $N \xrightarrow{p} \infty$, so that $P(N \geq 2) \to 1$.

Next, $\sum_{i=1}^{N-1}(T_i - 1) = S_N - S_1 - (N-1) = n - (n - S_N) - S_1 - (N-1)$ if $N \geq 2$. For $c > 0$, we have by Lemma B.1

$$P(n - S_N \geq cr_n \text{ and } N \geq 2) \leq P(M_{\lfloor cr_n \rfloor} \leq u_n) \leq C\gamma^{\lfloor cr_n \rfloor / r_n}$$

so that $n - S_N = O_p(r_n) = o_p(n)$. Similarly, $S_1 = O_p(r_n) = o_p(n)$. From the first part, we already know that $N = o_p(n)$. □

It remains to deal with the term $\sum_{i=1}^{N-1}(T_i - 1)(T_i - 2)$ (which we also set equal to zero if $N \in \{0, 1\}$). Define

$$X_n = \sum_{i=1}^{n-1}\sum_{j=1}^{n-i}(j-1)I(\xi_i > u_n \geq M_{i,j}).$$

Observe that if $N \geq 2$, then

$$\begin{aligned} X_n &= \sum_{i=1}^{N-1}\sum_{j=1}^{T_i - 1}(j-1) + \sum_{j=1}^{n-S_N}(j-1) \\ &= \frac{1}{2}\sum_{i=1}^{N-1}(T_i - 1)(T_i - 2) + \frac{1}{2}(n - S_N)(n - S_N - 1). \end{aligned}$$

In particular, we have

$$X_n = \frac{1}{2}\sum_{i=1}^{N-1}(T_i - 1)(T_i - 2) + O_p(r_n^2).$$

Hence, to prove the Theorem, it is sufficient to show that $n^{-1}p_n X_n \xrightarrow{p} \theta^{-1}$. To this end, we need only show that

$$n^{-1}p_n E(X_n) \to \theta^{-1} \quad \text{and} \quad n^{-2}p_n^2 E(X_n^2) \to \theta^{-2}.$$

Lemma B.4 $n^{-1}p_n E(X_n) \to \theta^{-1}$

Proof. By stationarity of the process $\{\xi_n\}_{n\geq 1}$, we obtain

$$\begin{aligned} E(X_n) &= \sum_{i=1}^{n-1}\sum_{j=1}^{n-i}(j-1)P(\xi_i > u_n \geq M_{i,i+j}) \\ &= p_n \sum_{j=1}^{n}(j-1)(n-j)P(M_{1,1+j} \leq u_n \mid \xi_1 > u_n). \end{aligned}$$

Rewrite the sum as an integral to find

$$\begin{aligned} E(X_n) &= p_n \int_0^n (\lceil s\rceil - 1)(n - \lceil s\rceil)P(M_{1,1+\lceil s\rceil} \leq u_n \mid \xi_1 > u_n)\,\mathrm{d}s \\ &= p_n r_n \int_0^{n/r_n} (\lceil r_n s\rceil - 1)(n - \lceil r_n s\rceil)P(M_{1,1+\lceil r_n s\rceil} \leq u_n \mid \xi_1 > u_n)\,\mathrm{d}s. \end{aligned}$$

Hence we have

$$n^{-1}p_n E(X_n) = p_n r_n \int_0^{n/r_n} p_n(\lceil r_n s\rceil - 1)(1 - \lceil r_n s\rceil/n)P(M_{1,1+\lceil r_n s\rceil} \leq u_n \mid \xi_1 > u_n)\,\mathrm{d}s.$$

By Lemmas B.1 and B.2, the dominated convergence theorem yields

$$n^{-1}p_n E(X_n) \to \tau^2 \int_0^\infty s\theta e^{-\theta\tau s}\,\mathrm{d}s = \theta^{-1}.$$

□

Lemma B.5 $n^{-2}p_n^2 E(X_n^2) \to \theta^{-2}$

Proof. For integers $n \geq 1$, $1 \leq i \leq n-1$ and $1 \leq j \leq n-i$, abbreviate

$$I_{i,j} = I_{i,j}^{(n)} = I(\xi_i > u_n \geq M_{i,i+j}).$$

We have

$$E(X_n^2) = \sum_{i=1}^{n-1}\sum_{j=1}^{n-i}\sum_{k=1}^{n-1}\sum_{l=1}^{n-k}(j-1)(l-1)E(I_{i,j}I_{k,l}).$$

The expectations can be computed as follows:

$$\begin{aligned} E(I_{i,j}) &= E(I_{1,j}), \\ E(I_{i,j}I_{i,l}) &= E(I_{1,j\vee l}), \\ E(I_{i,j}I_{k,l}) &= 0, \quad \text{for } i < k \le i+j, \\ E(I_{i,j}I_{k,l}) &= E(I_{1,j})E(I_{1,l}), \quad \text{for } k \ge i+j+m. \end{aligned}$$

Hence we can write $E(X_n^2) = A_n + B_n + C_n$ where

$$\begin{aligned} A_n &= \sum_{i=1}^{n-1}\sum_{j=1}^{n-i}\sum_{l=1}^{n-i}(j-1)(l-1)E(I_{1,j\vee l}), \\ B_n &= 2\sum_{i=1}^{n-1}\sum_{j=1}^{n-i}\sum_{k=i+j+1}^{i+j+m-1}\sum_{l=1}^{n-k}(j-1)(l-1)E(I_{1,j}I_{1,l}), \\ C_n &= 2\sum_{i=1}^{n-1}\sum_{j=1}^{n-i}\sum_{k=i+j+m}^{n-1}\sum_{l=1}^{n-k}(j-1)(l-1)E(I_{1,j})E(I_{1,l}). \end{aligned}$$

The terms A_n and B_n. We have

$$\begin{aligned} A_n &\le n\sum_{j=1}^{n-1}\sum_{l=1}^{n-1}(j-1)(l-1)E(I_{1,j\vee l}), \\ B_n &\le 2mn\sum_{j=1}^{n-1}\sum_{l=1}^{n-1}(j-1)(l-1)E(I_{1,j\vee l}). \end{aligned}$$

Changing sums gives

$$\sum_{j=1}^{n-1}\sum_{l=1}^{n-1}(j-1)(l-1)E(I_{1,j\vee l}) \le 2\sum_{j=1}^{n-1}(j-1)\sum_{l=j}^{n-1}(l-1)E(I_{1,l}) = \sum_{l=1}^{n-1}(l-1)^2 lE(I_{1,l}).$$

By Lemma B.1, we have $E(I_{1,l}) = P(\xi_1 > u_n \ge M_{1,1+l}) \le p_n C\gamma^{l/r_n}$. Hence both $n^{-1}A_n$ and $n^{-1}B_n$ are bounded by some positive constant times $p_n\sum_{l=1}^{\infty}(l-1)^2 l\gamma^{l/r_n}$. As $n\to\infty$, this expression is $O\left\{p_n(1-\gamma^{1/r_n})^{-4}\right\} = O(p_n^{-3})$. Therefore, both A_n and B_n are $O(np_n^{-3}) = o(n^2p_n^{-2})$.

The term C_n. Changing the order of summation, we have

$$\begin{aligned} C_n &= 2\sum_{j=1}^{n}\sum_{l=1}^{n}(j-1)(l-1)E(I_{1,j})E(I_{1,l})\sum_{i=1}^{n-j}\sum_{k=i+j+m}^{n-l}1 \\ &= \sum_{j=1}^{n}\sum_{l=1}^{n}(j-1)(l-1)(n-l-j-m+1)(n-l-j-m)_+E(I_{1,j})E(I_{1,l}). \end{aligned}$$

Write the sum as an integral

$$\begin{aligned}C_n &= \int_0^n \int_0^n (\lceil s\rceil - 1)(\lceil t\rceil - 1)(n - \lceil s\rceil - \lceil t\rceil - m + 1) \\ &\qquad \times (n - \lceil s\rceil - \lceil t\rceil - m)_+ E(I_{1,\lceil s\rceil}) E(I_{1,\lceil t\rceil})\, \mathrm{d}s\, \mathrm{d}t \\ &= r_n^2 \int_0^{n/r_n} \int_0^{n/r_n} (\lceil r_n s\rceil - 1)(\lceil r_n t\rceil - 1)(n - \lceil r_n s\rceil - \lceil r_n t\rceil - m + 1) \\ &\qquad \times (n - \lceil r_n s\rceil - \lceil r_n t\rceil - m)_+ E(I_{1,\lceil r_n s\rceil}) E(I_{1,\lceil r_n t\rceil})\, \mathrm{d}s\, \mathrm{d}t.\end{aligned}$$

Since $E(I_{1,\lceil r_n s\rceil}) = p_n P(M_{1,1+\lceil r_n s\rceil} \leq u_n \mid \xi_1 > u_n)$, we can apply Lemmas B.1 and B.2, giving

$$n^{-2} p_n^2 C_n \to \tau^4 \int_0^\infty \int_0^\infty st\theta e^{-\theta\tau s} \theta e^{-\theta\tau t}\, \mathrm{d}s\, \mathrm{d}t = \theta^{-2}$$

by dominated convergence. □

References

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

Hsing, T. (1987) On the characterization of certain point processes. *Stochastic Process. Appl.*, **26**, 297–316.

Hsing, T. (1991) Estimating the parameters of rare events. *Stochastic Process. Appl.*, **37**, 117–139.

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

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

Leadbetter, M. R. (1995) On high-level exceedance modeling and tail-inference. *J. Statist. Plann. Inference*, **45**, 247–260.

Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983) *Extremes and Related Properties of Random Sequences and Processes*, ch. 3. New York: Springer.

Leadbetter, M. R. and Nandagopalan, S. (1989) On exceedance point processes for stationary sequences under mild oscillation restrictions. In *Extreme Value Theory, Lecture Notes in Statistics 51* (eds J. Hüsler and R. D. Reiss), pp. 69–80. New York: Springer.

Leadbetter, M. R., Weissman, I., de Haan, L. and Rootzén, H. (1989) On clustering of high values in statistically stationary series. *Technical Report 253*. Center for Stochastic Processes, University of North Carolina, Chapel Hill.

Longin, F. M. (2000) From value at risk to stress testing: the extreme value approach. *J. Banking Finance*, **24**, 1097–1130.

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

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

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

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