Research Report

Estimating the Extremal Index and Its Uncertainty

Jolan Bonny

Supervised by Prof. Anthony Davison
EPFL, Chair of Statistics

October 2025

Abstract

This report explores how to improve the way we measure confidence in estimates related to extreme events, such as floods or financial crashes. These events tend to occur in bursts, and a key step in their analysis is estimating how clustered they are. But just producing an estimate isn't enough, we also need to know how uncertain that estimate is.

Traditionally, this uncertainty is assessed using a simulation-based method called the bootstrap. While effective, it can be slow and computationally demanding. The goal of this project is to explore a faster alternative, based on a mathematical formula rather than repeated simulations.

Both approaches, the standard method and the new one, are implemented and tested on artificial data. The results show that the new approach is much faster, while giving similar results. This makes it a promising tool when speed and scalability are important, such as in large studies or real-time systems.

Contents

1 Introduction

Extreme events such as floods, heatwaves, or financial crashes are rare but can have severe consequences. Understanding how often they occur and how intense they can be is essential in many fields, from hydrology and climate science to finance and engineering. Traditional statistical methods focus on the average behavior of a system, which provides little information about the most extreme cases. Extreme Value Theory (EVT) was developed to address this limitation. It offers mathematical tools to describe and predict the behavior of rare events that lie in the tails of probability distributions.

The central idea of EVT is that even though extreme events are rare, their statistical behavior follows simple and universal patterns. By focusing only on the largest observations in a dataset, we can describe how maxima behave and extrapolate the probability of more extreme values that may not yet have been observed. This makes EVT particularly valuable for estimating quantities such as the maximum expected rainfall over a decade or the worst annual financial loss.

In its classical form, EVT assumes that observations are independent and identically distributed. This assumption simplifies the mathematics and leads to elegant results, but it rarely holds in practice. In most natural and economic systems, large values tend to occur in bursts rather than in isolation: heatwaves consist of several consecutive hot days, and market crashes often unfold over multiple trading sessions. This dependence creates clusters of extremes and violates the assumption of independence.

Modeling such dependence is crucial for realistic risk assessment. Ignoring it can lead to serious underestimation of the probability of severe events. To address this, EVT introduces an additional parameter called the extremal index, which measures how strongly extremes tend to cluster in time. Estimating this parameter accurately, and understanding the uncertainty of that estimate, is essential for reliable predictions.

The following chapters introduce the foundations of EVT, explain how dependence modifies the classical results, and present modern methods to estimate and quantify the uncertainty of the extremal index.

2 Extreme Value Theory

Extreme Value Theory provides a framework to describe the statistical behavior of the most extreme observations in a dataset. Rather than modeling the entire distribution, EVT focuses only on the largest values.

2.1 From random variables to maxima

A random variable X represents the outcome of a random experiment, and its distribution function

$$F(x) = \mathbb{P}(X \leq x)$$

gives the probability that X takes a value below x. If we collect independent samples $X_1, X_2, \ldots, X_n$, the largest value

$$M_n = \max(X_1, X_2, \ldots, X_n)$$

represents the most extreme event among them. As n increases, M_n tends to grow, since larger samples are more likely to contain rare high values.

However, to describe the distribution of M_n as n grows, we need to rescale it. Without adjustment, its distribution would either diverge to infinity or collapse to a single point. To stabilize it, we use two normalizing constants $a_n > 0$ and $b_n \in \mathbb{R}$, and study

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq z\right) = F^n(a_n z + b_n)$$

The constants a_n and b_n act as scaling and centering factors, allowing the rescaled maxima to converge to a non-trivial limiting law.

2.2 The Generalized Extreme Value distribution

The Fisher–Tippett–Gnedenko theorem states that if this limit exists, it must belong to a unique three-parameter family called the Generalized Extreme Value (GEV) distributions [Coles, 2001]:

$$G(z) = \exp\left[-\left(1 + \xi\frac{z-\mu}{\sigma}\right)^{-1/\xi}\right] \qquad 1 + \xi\frac{z-\mu}{\sigma} > 0$$

Here, μ is a location parameter, $\sigma > 0$ a scale parameter, and ξ a shape parameter controlling the heaviness of the tail. Different values of ξ correspond to distinct types of tail behavior, from bounded to heavy-tailed distributions.

Once fitted to data, the GEV model allows us to estimate return levels, values expected to be exceeded on average once every T years, and to extrapolate beyond observed records. This is what makes EVT such a powerful predictive tool: by learning the parameters (μ, σ, ξ) from the past, we can describe how future extremes are likely to behave.

2.3 Exceedances and the Poisson process

Another way to study extremes is to fix a high threshold u and look at the exceedances, that is, the events $\{X_i > u\}$. When the data are independent, these exceedances occur randomly in time and can be modeled by a Poisson process with rate λ. A Poisson process is a model for random events happening independently over time, where the number of events in any interval of length t follows a Poisson distribution with mean λt.

A useful property of this process is that the waiting times between successive exceedances follow an exponential distribution with mean $1/\lambda$. This connection between Poisson processes and exponential waiting times is central to the construction of extreme value estimators based on inter-exceedance times.

In real data, this idealized independent structure rarely holds. The next chapter introduces how this dependence modifies the classical GEV and Poisson model and how it can be described through the extremal index θ.

3 The Extremal Index and Its Estimation

3.1 Dependence and clustering of extremes

In many real-world time series, extreme values tend to appear in bursts rather than as isolated events. A storm can produce several days of heavy rainfall, and a heatwave can generate several consecutive high temperatures. This temporal dependence changes the way extremes behave and must be considered to avoid underestimating their frequency.

When data are independent, exceedances over a high threshold occur at random times, and the gaps between them follow an exponential distribution. In dependent data, exceedances still occur irregularly, but they often group together rather than appearing separately. Each group, or cluster, may contain several exceedances, and the time intervals between clusters remain approximately random.

This clustering is summarized by the extremal index $\theta \in (0, 1]$, which modifies both the limiting distribution of maxima and the structure of exceedances:

$$\mathbb{P}(M_n \leq z) \;\to\; G(z)^\theta$$

When $\theta = 1$, extremes occur independently; when $\theta < 1$, they occur in clusters, and $1/\theta$ approximates the average cluster size.

Therefore estimating θ accurately is essential for describing how often extreme events occur in practice. The next sections explain how dependence modifies the exponential structure of inter-exceedance times, how this leads to a mixed distribution, and how this property can be used to estimate θ from data.

3.2 From dependence to inter-exceedance times

Let S_j denote the time indices where the variable exceeds a high threshold u, i.e. $X_{S_j} > u$, and define the time gaps between exceedances by

$$T_j = S_{j+1} - S_j$$

Under independence, these gaps follow an exponential distribution with mean $1/\lambda$, where λ is the rate of exceedances. In dependent data, extreme values tend to cluster. As the number of observations increases and the threshold u rises toward the upper endpoint of the distribution, exceedances within each cluster become closer and closer in time, until they effectively merge in the limit. Only the times between clusters remain positive.

At this asymptotic level, the distribution of inter-exceedance times becomes a mixture [Ferro and Segers, 2003]:

$$T \;\sim\; (1-\theta)\,\delta_0 \;+\; \theta\,\mathrm{Exp}(\theta\lambda)$$

where δ_0 represents the probability mass at zero (gaps within clusters), and $\mathrm{Exp}(\theta\lambda)$ describes the random times separating distinct clusters. Intuitively, θ is the proportion of exceedances that start a new cluster, while λ represents the overall exceedance rate.

3.3 Constructing the estimator

The estimator we use is built by just considering the exponential part of the distribution and comparing it to the large inter-cluster gaps observed in the data Holešovský and Fusek [2022]. In practice, exceedances within a cluster are close but not exactly simultaneous. To separate the small within-cluster gaps from the larger inter-cluster ones, we introduce a small cutoff value $D > 0$. Gaps smaller than D are treated as belonging to the same cluster, while gaps larger than D are used to estimate θ.

From the theoretical mixture model, the expectation of the positive (inter-cluster) gaps satisfies

$$\mathbb{E}[T - D \mid T > D] = \frac{1}{\theta\lambda}$$

This relationship is the starting point for building the estimator.

Step 1: Empirical estimation of the mean gap The theoretical expectation is replaced by its empirical version, computed directly from the observed inter-exceedance times:

$$\widehat{\mathbb{E}}[T - D \mid T > D] = \frac{\sum_{j=1}^{n_u-1}(T_j - D)\,\mathbf{1}(T_j > D)}{\sum_{j=1}^{n_u-1}\mathbf{1}(T_j > D)}$$

Here, "empirical" simply means calculated from data rather than from a theoretical distribution.

Step 2: Estimating $\lambda\theta$ Taking the inverse of this sample mean provides an estimate of $\lambda\theta$:

$$\widehat{\lambda\theta} = \frac{\sum_{j=1}^{n_u-1} \mathbf{1}(T_j > D)}{\sum_{j=1}^{n_u-1} (T_j - D)\,\mathbf{1}(T_j > D)}$$

Step 3: Estimating λ The rate λ of exceedances is estimated empirically as the proportion of data points above the threshold:

$$\hat{\lambda} = \frac{n_u}{n}$$

where n_u is the number of exceedances among n total observations.

Step 4: Estimating θ Dividing the previous expression by $\hat{\lambda}$ gives the practical estimator

$$\hat{\theta}_D = \frac{\sum_{j=1}^{n_u-1} \mathbf{1}(T_j > D)}{\hat{\lambda} \sum_{j=1}^{n_u-1} (T_j - D)\,\mathbf{1}(T_j > D)}$$

This estimator uses only the large inter-exceedance gaps, which contain the relevant information about clustering, and is consistent under general dependence conditions.

3.4 Bias corrections and bounded transformation

The estimator $\hat{\theta}_D$ converges to the true value of θ as the sample size increases and the threshold rises, but this convergence can be slow. The theoretical mixture model describes an idealized limit with infinite data and threshold really high, whereas in real data it is not the case. As a result, small systematic biases appear in finite samples.

To correct for this, two analytical bias corrections are applied to $\hat{\theta}_D$. These corrections improve convergence speed and reduce bias when applied to realistic datasets.

Finally, since θ must lie within $(0, 1]$, the corrected estimate is bounded to ensure meaningful results:

$$\hat{\theta}_T = \min\big(1, \max(0, \hat{\theta}_D^{(\text{corrected})})\big)$$

This bounded and bias-corrected version, denoted $\hat{\theta}_T$, provides stable and interpretable estimates of the extremal index across a wide range of thresholds.

Summary

Dependence between observations causes extremes to cluster in time, leading to a mixture of zero and exponential waiting times between exceedances. By isolating the exponential part and estimating its mean empirically, one obtains a consistent estimator of the extremal index. Bias corrections improve its finite-sample performance, and bounding ensures that the final estimates remain within $(0, 1]$.

4 Measuring Uncertainty: Bootstrap and Influence Functions

Estimating the extremal index θ gives valuable insight into how strongly extremes cluster in time, but the estimate alone is not sufficient. We also need to know how uncertain it is. Quantifying this uncertainty is essential in applications where θ is used to predict return levels or evaluate risks. This chapter presents two approaches to estimate this uncertainty: the traditional bootstrap and a new analytical method based on influence functions. Both rely on the same declustering scheme, originally formalized under the dependence conditions of Leadbetter et al. [1983].

4.1 Bootstrap method (Ferro and Segers, 2003)

The bootstrap estimates uncertainty by repeatedly generating synthetic datasets from the original one. For each new dataset, the estimator is recomputed, and the variability of these estimates gives an approximation of the estimator's variance.

Because extremes occur in clusters, we cannot simply resample single data points: this would destroy the dependence structure. Instead, the series is first *declustered* following the approach of Ferro and Segers [2003]. The inter-exceedance times are computed and ordered, and the largest ones are treated as separators between clusters. The remaining smaller gaps are grouped within clusters. The number of clusters is estimated as

$$\hat{n}_c = 1 + \lfloor \hat{\theta}(n_u - 1) \rfloor$$

To perform the bootstrap, the sets of between-cluster intervals (T) and within-cluster groups (C) are resampled separately. New synthetic series are reconstructed by alternating sampled clusters and long gaps. The estimator $\hat{\theta}_D$ is recomputed for each resampled series, and the empirical spread of these values provides an estimate of its variance and confidence intervals. This method is accurate but computationally heavy, especially when repeated over many thresholds.

4.2 Analytical approach: functional reformulation and influence functions

An alternative is to approximate the variance analytically, without resampling. This relies on the *delta method* and on a concept known as the *influence function*, which measures how much an estimator changes when one observation in the data is slightly modified [Davison and Hinkley, 1997].

We start from the estimator formula:

$$\hat{\theta}_D \;=\; \frac{\sum_{j=1}^{n_u-1} \mathbf{1}(T_j > D)}{\hat{\lambda} \sum_{j=1}^{n_u-1} (T_j - D)\, \mathbf{1}(T_j > D)} \qquad \hat{\lambda} = \frac{n_u}{n}$$

After declustering, the data are divided into two sets:

- the between-cluster gaps, denoted $T = \{t_1^T, \ldots, t_{n_T}^T\}$
- the within-cluster gaps, grouped into clusters $C = \{C_1, \ldots, C_{n_C}\}$

Both types of intervals contribute to the estimation of θ. We can express the estimator as two separate sums:

$$\hat{\theta}_D \;=\; \frac{\sum_{t_j \in T} \mathbf{1}(t_j > D) + \sum_{y \in C} \sum_{t \in y} \mathbf{1}(t > D)}{\hat{\lambda}\left(\sum_{t_j \in T} (t_j - D)\,\mathbf{1}(t_j > D) + \sum_{y \in C} \sum_{t \in y} (t - D)\,\mathbf{1}(t > D)\right)}$$

To analyse this expression, we rewrite it using empirical distributions. An *empirical distribution* represents the observed frequencies of data values in a sample; for instance, $\hat{F}_T(x)$ gives the proportion of inter-cluster gaps shorter than or equal to x. Defining

$$a(x) = \mathbf{1}(x > D) \qquad b(x) = (x - D)\,\mathbf{1}(x > D)$$

and analogously $A(y)$ and $B(y)$ for the within-cluster gaps, the estimator becomes

$$\hat{\theta}_D \;=\; \frac{w \int a(x)\, d\hat{F}_T(x) + \int A(y)\, d\hat{F}_C(y)}{\hat{\lambda}\left(w \int b(x)\, d\hat{F}_T(x) + \int B(y)\, d\hat{F}_C(y)\right)} \qquad w = \frac{\hat{n}_c - 1}{\hat{n}_c}$$

This form allows us to apply the delta method, which approximates how an estimator varies when the empirical distributions $\hat{F}_T$ and $\hat{F}_C$ are perturbed.

Influence functions. To measure the impact of a small change in the data, we add a tiny weight ε to one observation and compute how much $\hat{\theta}_D$ changes as ε varies. The derivative of this change at $\varepsilon = 0$ gives the influence function, noted $\mathrm{IF}(z)$. In simple terms, $\mathrm{IF}(z)$ tells us how sensitive the estimator is to the observation z. Each observation, whether in T or C, has its own influence value.

Differentiating the functional above leads to two sets of influence values:

$$\ell_j \;=\; \frac{w}{R}\Big[a(t_j) - \bar{a} - \hat{\lambda}\,\hat{\theta}_D(b(t_j) - \bar{b})\Big] \quad L_j \;=\; \frac{1}{R}\Big[A(C_j) - \bar{A} - \hat{\lambda}\,\hat{\theta}_D(B(C_j) - \bar{B})\Big]$$

where R denotes the denominator of the estimator, and bars (such as $\bar{a}$) represent empirical averages. Each ℓ_j or L_j quantifies the effect of one observation on the estimated value of $\hat{\theta}_D$.

Variance approximation. Summing the squared influence values provides an analytical approximation of the variance:

$$\mathrm{Var}(\hat{\theta}_D) \approx \frac{1}{(|T|-1)(|T|-2)} \sum_{j \in T} \ell_j^2 + \frac{1}{|C|(|C|-1)} \sum_{j \in C} L_j^2$$

This approach is much faster than the bootstrap and should give nearly identical results in practice. It may however be slightly conservative, since it assumes that the number of clusters $\hat{n}_c$ and the rate $\hat{\lambda}$ are fixed.

5 Results and Discussion

This section presents the comparison between the analytical variance derived from influence functions and the empirical variance obtained by the bootstrap. Both methods are applied to simulated data from a dependent process that naturally produces clusters of extremes: the MoveMax process.

5.1 Simulation design

The MoveMax process provides a simple and controlled setting for studying dependence among extremes. Each observation is defined as

$$X_t = \log\left(\frac{\max(w_1 Z_t, w_2 Z_{t+1}, \ldots, w_K Z_{t+K-1})}{\sum_{i=1}^{K} w_i}\right)$$

where the Z_t are independent and identically distributed exponential variables. Because each X_t depends on several consecutive Z_t, large values tend to appear in bursts, forming natural clusters. The degree of clustering is controlled by the weights $w = (w_1, \ldots, w_K)$: larger overlaps between windows lead to stronger dependence.

For the experiments, we used $w = (1.0, 0.7, 0.5)$, corresponding to a theoretical extremal index $\theta_{\text{true}} = 0.476$. Each simulated series contained $n = 1000$ observations. For quantile thresholds between $q = 0.60$ and 0.95, corresponding to values exceeding the 60th to 95th percentiles of the data, the estimator of Holešovský and Fusek (2022) was applied, and its uncertainty was measured using both the analytical variance and a bootstrap with 300 resamples.

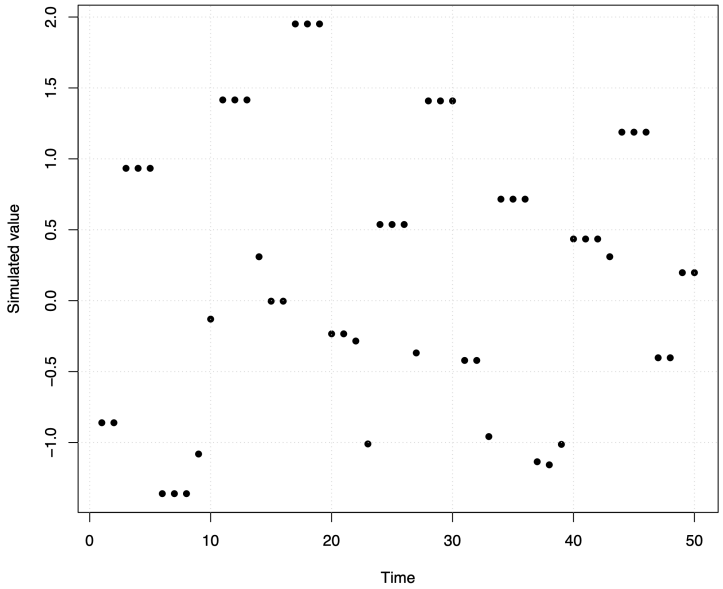

Figure 1: Example of a simulated MoveMax process with $w = (1.00, 1.00, 1.00)$ and $n = 50$. Overlapping windows of maxima generate visible clusters of extremes.

5.2 Results

The estimated values of $\hat{\theta}_T$ across thresholds are displayed in Figure 2, together with 90% confidence intervals computed from the analytical variance. The dashed line represents the theoretical value θ_{true}. The estimates remain stable across thresholds, and the intervals capture the true value for most quantiles, showing that the analytical variance provides a realistic measure of uncertainty.

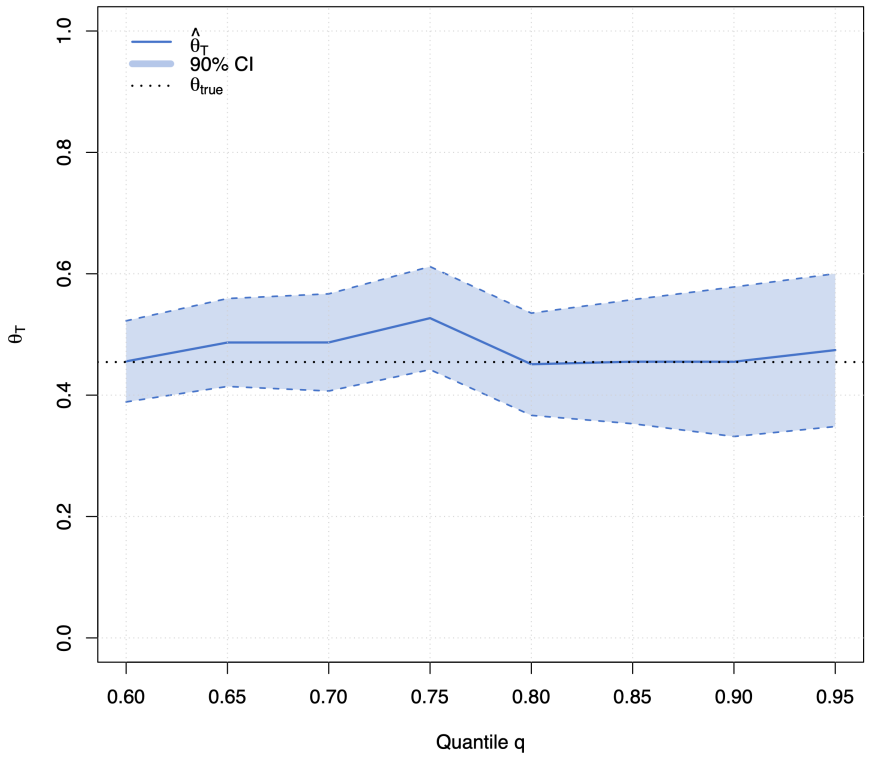

Figure 2: Estimated $\hat{\theta}_T$ with 90% confidence intervals based on the analytical variance. The dashed line shows the theoretical value $\theta_{\text{true}} = 0.476$.

Figure 3 compares the standard deviations obtained from the bootstrap and from the analytical method. The two curves overlap almost perfectly across thresholds, confirming that both approaches yield nearly identical uncertainty estimates. At very high thresholds, where fewer exceedances are available, small differences appear, but the analytical variance remains close to the bootstrap reference.

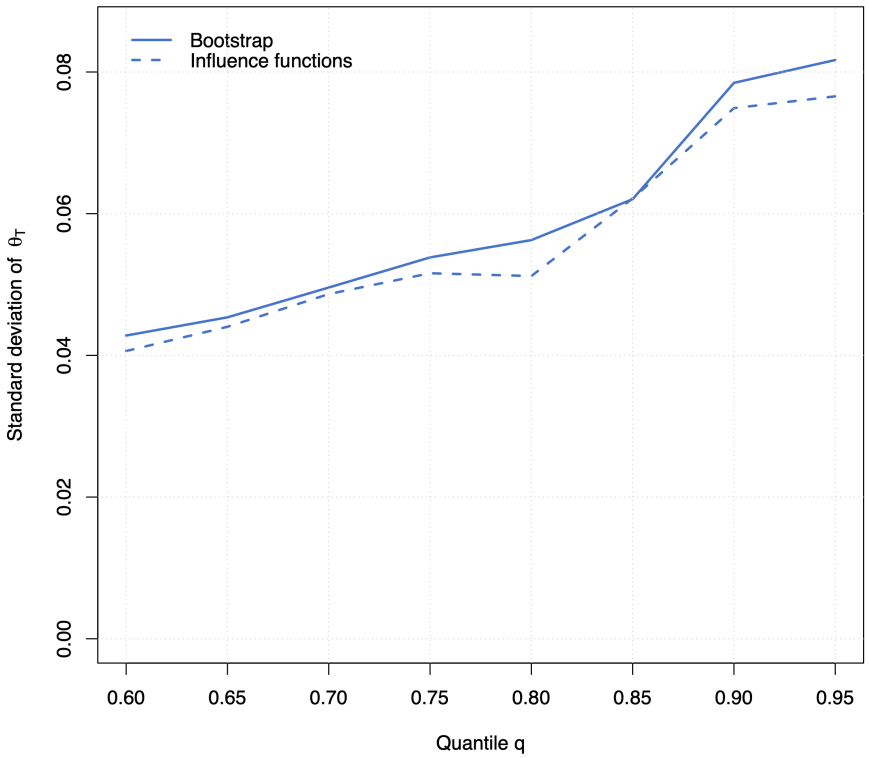

Figure 3: Comparison of standard deviations from the bootstrap (solid line) and from the analytical method (dashed line). The two approaches produce nearly identical estimates of uncertainty.

The slight underestimation observed for the analytical variance is expected: it assumes the number of clusters $\hat{n}_c$ and the exceedance rate $\hat{\lambda}$ are fixed, whereas the bootstrap implicitly includes their randomness through resampling. Nevertheless, the difference is small compared with the overall stability and speed of the analytical approach.

5.3 Interpretation

These results show that the influence-function approach reproduces almost exactly the variability captured by the bootstrap, but at a fraction of its computational cost. The agreement between the two confirms that the analytical approximation captures the dominant sources of variation of $\hat{\theta}_T$. In practice, this makes it possible to obtain reliable confidence intervals without repeated resampling, which is particularly useful when estimating θ for multiple thresholds or large datasets.

Beyond computation, the analytical formulation also provides insight into how individual data points influence the final estimate. This interpretability is one of its strengths: the influence values highlight directly which clusters or intervals contribute most to the estimator's uncertainty.

Overall, the results demonstrate that the analytical variance derived from influence functions offers an efficient and accurate alternative to the bootstrap for assessing uncertainty in extremal-index estimation.

6 Conclusion

This study explored how to quantify the uncertainty of extremal-index estimators more efficiently. Using the bias-corrected estimator of Holešovský and Fusek (2022), we reformulated it as an integral functional and derived an analytical approximation of its variance using influence functions. This approach avoids resampling and provides a direct, interpretable measure of uncertainty.

Through simulations of the MoveMax process, the analytical variance was shown to align closely with the bootstrap reference, confirming both its validity and robustness. Although it slightly underestimates uncertainty due to fixed-sample assumptions, this difference is small and consistent across thresholds.

The influence-function method therefore provides a valuable alternative for large-scale or computationally demanding applications, where repeated bootstraps would be prohibitive. Beyond its efficiency, it offers a better understanding of how different parts of the data affect the estimate, bridging theory and interpretation.

Future work could apply this analytical framework to other extremal-index estimators and use the faster implementation to perform large-scale simulation studies. Such experiments could test the precision of the analytical variance through coverage analyses, investigate in which situations the method performs best, and identify potential weaknesses in the declustering scheme or ways to improve it.

References

Stuart Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer, 2001.

A. C. Davison and D. V. Hinkley. *Bootstrap Methods and Their Application*. Cambridge University Press, 1997.

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

J. Holešovský and M. Fusek. Improved interexceedance-times-based estimator of the extremal index using truncated distribution. *Extremes*, 2022.

M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes*. Springer, 1983.