Trends in Extreme Value Index

Laurens de Haan and Chen Zhou

Dec, 07, 2020

Published on Journal of the American Statistical Association(2020).

Presented by Liujun Chen

Table of Contents

Background

- Classic extreme value analysis assumes that the observations are i.i.d.
- Recent studies aim at dealing with the case when observations are drawn from different distributions.
- This paper considers a continuously changing extreme value index and try to estimate the functional extreme value index accurately.

Model Setting

- Consider a set of distributions $F_s(x)$ for $s \in [0,1]$ and independent random variables $X_i \sim F_{\frac{i}{n}}(x)$ for $i = 1, \ldots, n$.
- Here $F_s(x)$ is assumed to be continuous with respect to s and x. And assume that $F_s \in D_{\gamma(s)}$.
- This article considers the case that the function γ is positive and continuous on $[0,1]$.
- The goal is to estimate the function γ and test the hypothesis that $\gamma = \gamma_0$ for some given function γ_0.

Table of Contents

Methodology

- The idea for estimating $\gamma(s)$ locally is similar to the kernel density estimation.
- More specifically, use only observations X_i in the h-neighborhood of s, that is
$$i \in I_n(s) = \left\{ \left| \frac{i}{n} - s \right| \leq h \right\},$$
where $h := h(n)$ is the bandwidth such that as $n \to \infty$, $h \to \infty$ and $nh \to \infty$.
- Correspondingly, there will be $[2nh]$ observations for $s \in [h, 1-h]$.

Local estimator

- To apply any known estimators for extreme value index, such as Hill estimator, choose am intermediate sequence $k = k(n)$ such that $k \to \infty, k/n \to 0$ as $n \to \infty$.
- Then one can use the top $[2kh]$ order statistics among the $[2nh]$ local observations in the h-neighborhood of s to estimate $\gamma(s)$.
- The local Hill estimator for $\gamma(s)$ is defined as follows. Rank the $[2nh]$ observations X_i with $i \in I_n(s)$ as $X^{(s)}_{1,[2nh]} \leq \cdots \leq X^{(s)}_{[2nh],[2nh]}$. Then

$$\hat{\gamma}_H(s) := \frac{1}{[2kh]} \sum_{i \in I_n(s)} \left(\log X_i - \log X_{[2nh]-[2kh],[2nh]} \right)^+ .$$

Second order condition

To obtain the asymptotic theory, the following conditions are required.

First, the second order condition is assumed. Suppose there exists a continuous negative function $\rho(s)$ on $[0,1]$ and a set of auxiliary function $A_s(t)$ that are continuous with respect to s, such that

$$\lim_{t\to\infty} \frac{\frac{U_s(tx)}{U_s(t)} - x^{\gamma(s)}}{A_s(t)} = x^{\gamma(s)} \frac{x^{\rho(s)} - 1}{\rho(s)}, \tag{3}$$

holds for $x > 1/2$ and uniformly for all $s \in [0,1]$.

Conditions for k and h

Choose the intermediate sequence k and the bandwidth h as follows:

- $$h = h_n \to 0, k = k_n \to \infty, k_n/n \to 0, \frac{k_n h_n}{|\log h_n|} \to \infty, \tag{4}$$
- $$\Delta_{1,n} := \sqrt{k_n} \sup_{0 \le s \le 1} |A_s(\frac{n}{k_n})| \to 0 \tag{5}$$
- $$\Delta_{2,n} := \sqrt{k_n} \log k_n \sup_{|s_1 - s_2| \le 2h_n} |\gamma(s_1) - \gamma(s_2)| \to 0. \tag{6}$$
- $$\Delta_{3,n} := \sup_{|s_1 - s_2| \le h_n} \left| \frac{\frac{U_{s_1}\left(\frac{n}{k_n}\right)}{U_{s_2}\left(\frac{n}{k_n}\right)} - 1}{A_{s_2}\left(\frac{n}{k_n}\right)} \right| \to 0 \tag{7}$$

Theorem 2.1: Asymptotical normality of the local estimator

Theorem (2.1)

Let $X_1, X_2, \ldots, X_n$ be independent random variables with distributions $X_i \sim F_{\frac{i}{n}}(x)$, where $F_s(x)$ is a set of distribution functions depends on $s \in [0, 1]$. Assume that $F_s(x)$ is continuous with respect to x and s and $F_s \in D_{\gamma(s)}$ where $\gamma(s)$ is a positive continuous function on [0,1]. Assume conditions (3)-(7). Then as $n \to \infty$, we have that for all $s \in (0, 1)$,

$$\sqrt{2kh}(\hat{\gamma}_H(s) - \gamma(s)) \xrightarrow{d} N(0, \gamma^2(s)).$$

Sketch of the Proof

Notations:

- $\bar{U}_{s,n} = \max_{\{i:|i/n-s|\leq h\}} U_{i/n}$, $\underline{U}_{s,n} = \min_{\{i:|i/n-s|\leq h\}} U_{i/n}$

Write $X_i = U_{i/n}(Z_i)$, where $\{Z_i\}_{i=1}^{n}$ are iid standard Pareto distributed random variables. Also rank $\{Z_i : i \in I_n(s)\}$ into order statistics.

Since $U_{i/n}$ are different, the order statistics $X^{(s)}_{j,[znh]}$ may not corrspond to the order statistics $Z^{(s)}_{j,[2nh]}$. Nevertheless,

$$\underline{U}_{s,n}(Z^{(s)}_{j,[2nh]}) \leq X^{(s)}_{j,[znh]} \leq \bar{U}_{s,n}(Z^{(s)}_{j,[2nh]}).$$

Then,

$$\hat{\gamma}_H(s) \leq \frac{1}{[2kh]} \sum_{j=1}^{2kh} \left(\log \frac{\bar{U}_{s,n}(Z^{(s)}_{[2nh]-j+1,[2nh]})}{U_s(n/k)} - \log \frac{\underline{U}_{s,n}(Z^{(s)}_{[2nh]-j+1,[2nh]})}{U_s(n/k)} \right)$$

Sketch of the Proof

Some Lemmas(Corollary A.1):

$$\sqrt{k}\sup_{0\le s\le 1,1/2\le x\le q_n}\left|\frac{\bar{U}_{s,n}(\frac{n}{k}x)}{U_s(\frac{n}{k})x^{\gamma(s)}}-1\right|\to 0,$$

$$\sqrt{k}\sup_{0\le s\le 1,1/2\le x\le q_n}\left|\frac{\underline{U}_{s,n}(\frac{n}{k}x)}{U_s(\frac{n}{k})x^{\gamma(s)}}-1\right|\to 0,$$

Apply Corollary A.1 to bound $\hat{\gamma}_H(s)$. Need to verify that

$$Pr\left(\frac{k}{n}Z^{(s)}_{[2nh]-j+1,[2nh]}\in[1/2,q_n],\ \text{for all } j=1,2,\ldots,[2kh]\right)\to 1.$$

The rest of proofs are similar to that of the standard Hill estimator.

Global Estimator

Next, consider testing the hypothesis that $\gamma(s) = \gamma_0(s)$ for all $s \in [0, 1]$.

- Although we are able to estimate the function γ locally, since the local estimators use only local observations, their asymptotic limits are obviously independent. What does this mean?
- In addition, the local estimator converges with a slow speed of convergence $1/\sqrt{2kh}$.
- To achieve the stated goal, we consider the estimation $\Gamma(s) = \int_0^s \gamma(u)du$ and test the equivalent hypothesis that $\Gamma = \Gamma_0$.

Global Estimator

- Consider a discretized version of $\hat{\gamma}_H(s)$: $\hat{\gamma}_H\left((2[\frac{s}{2h}]+1)h\right)$.
- Define the estimator of $\Gamma(s)$ as the integral of the discretized version as follows: for all $0 \le s \le 1$,

$$\hat{\Gamma}_H(s) = \int_0^s \hat{\gamma}_H\left(\left(2\left[\frac{u}{2h}\right]+1\right)h\right)du.$$

- Note that $\hat{\gamma}_H(s)$ is only defined for $s \le 1-h$.
- For $u > 1-h$, we may have $(2[\frac{s}{2h}]+1)h = (2[\frac{1}{2h}]+1)h$. Then we extend the range of the estimator and define $\hat{\gamma}_H((2[\frac{s}{2h}]+1)h) := \hat{\gamma}_H((2[\frac{s}{2h}]-1)h)$

Asymptotic Properties of the Global Estimator

Theorem (2.2)

Assume the same conditions in Theorem 2.1. Then under a Skorokhod construction, there exists a series of Brownian motions $W_n(s)$ such that as $n \to \infty$,

$$\sup_{s \in [0,1]} \left| \sqrt{k} \left(\hat{\Gamma}_H(s) - \Gamma(s) \right) - \int_0^s \gamma(u) dW_n(u) \right| \xrightarrow{P} 0.$$

Sketch of the proofs

Apply the same discretized to the function γ as follows:

$$\tilde{\gamma} = \begin{cases} \gamma\left((2[\frac{s}{2h}]+1)h\right), & \text{for } s \in [0, 2h[1/(2h)]] \\ \gamma\left((2[\frac{s}{2h}]-1)h\right), & \text{for } s \in [2h[1/(2h)], 1]. \end{cases}$$

Main Steps:

- Prove that there exists a series of Brownian motions such that

$$\sup_{s\in[0,1]} \left| \sqrt{k}(\hat{\Gamma}(s) - \tilde{\Gamma}(s)) - \int_0^s \tilde{\gamma}(u) dW_n(u) \right| \xrightarrow{P} 0.$$

- Handle the uniform negligible difference between $\tilde{\gamma}(\cdot)$ and γ.

Sketch of the Proofs

Since the estimator $\hat{\Gamma}(s)$ involves the local estimators $\hat{\gamma}_H(s)$ at $s = (2p-1)h$ for $p = 1, 2, 3 \ldots, 1/[(2h)]$.
Similar to that in Theorem 2.1, need to verify that

$$Pr\left(\frac{k}{n} Z^{((2p-1)h)}_{[2nh]-j+1,[2nh]} \in [\frac{1}{2}, q_n], \forall 1 \le j \le [2kh]+1, 1 \le p \le [\frac{1}{2h}]\right) \to 1.$$

Then we obtian some uniform expansion:

$$\sqrt{k} \sup_{1 \le p \le [1/(2h)]} \left| \hat{\gamma}_H((2p-1)h) - \gamma((2p-1)h) J_n^{(p)} \right| = o_P(1),$$

where $J_n^{(p)} = \frac{1}{[2kh]} \log \frac{Z^{((2p-1)h)}_{[2nh]-j+1,[2nh]}}{Z^{((2p-1)h)}_{[2nh]-[2kh]+1,[2nh]}}$.
The rest of the proof follows by the theory of tail empirical process.(Still difficult.)

Testing Trends in Extreme Value Indices

The null hypothesis is $H_0 : \Gamma(s) = \Gamma_0(s)$ for all s.

Clearly, one may construct the Kolmogrov-Smirnov(KS) type test with the testing statistic defined as

$$T := \sup_{s\in[0,1]} \left| \hat{\Gamma}_H(s) - \Gamma_0(s) \right| .$$

Then by Theorem 2.2, under the null hypothesis H_0,

$$\sqrt{k}T \xrightarrow{d} \sup_{s\in[0,1]} \left| \int_0^s \gamma(u) dW(u) \right| .$$

Testing Trends in Extreme Value Indices

It is often of interest to test whether the extreme value index remains constant over time, that is, $H_0 : \gamma(s) = \gamma$ for all $s \in [0, 1]$ without specifying γ. In this case, one may use $\hat{\Gamma}_H(1)$ as an estimator of the constant extreme value index γ and define the testing statistic as

$$\tilde{T} := \sup_{s \in [0,1]} \left| \frac{\hat{\Gamma}_H(s)}{\hat{\Gamma}_H(1)} - s \right|.$$

It is straightforward to show under H_0, as $n \to \infty$

$$\sqrt{k}\,\tilde{T} \xrightarrow{d} \sup_{s \in [0,1]} |B(s)|,$$

where $B(s)$ is a standard Brownian bridge defined on [0,1].

Table of Contents

Simulation Setting

Run a simulation study to demonstrate the finite sample performance of the testing procedure using $\tilde{T}$.

- $m = 2000$ samples
- $n = 5000$ observations in each sample
- $k = 100, 200$
- $h = 0.025, 0.04$

For each sample, simulate the observations from the following data generating process

$$X_i = Z_i^{1/\gamma(i/n)}, i = 1, 2, \ldots, n,$$

where $\{Z_i\}_{i=1}^n$ are iid observations from the standard Fréchet distribution.

Simulation Setting

For the function $\gamma(s)$ we consider

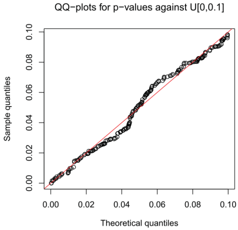
- a linear trend as $\gamma(s) = 1 + bs$
- or a trend following the sin function as $\gamma(s) = 1 + c\sin(2\pi s)$.

- If $b = 0$ or $c = 0$, the two model resemble the iid case.
- Consider four alternative cases $b = 1, b = 2, c = 1/4$ and $c = 1/2$.

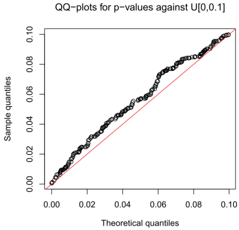
QQ plots for $b = 0$

(i) $k = 200$

$h = 0.025$

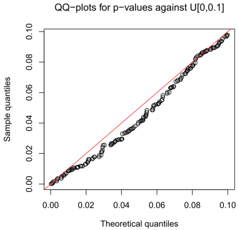


$h = 0.04$

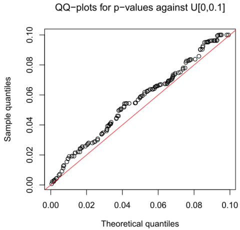


(ii) $k = 100$

$h = 0.025$



$h = 0.04$



Rejection Rate

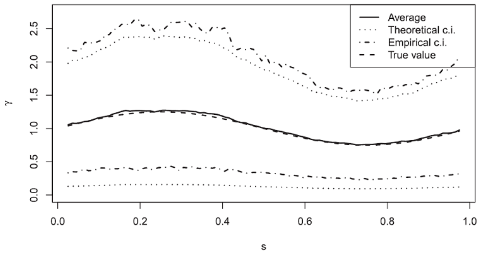
Rejection Rate $: = \#\{j : p_j < \alpha\} / m$

Table 1. Rejection rates in simulations: sample size $n = 5000$.

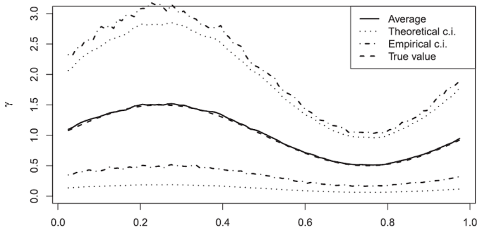
			$k = 200$		$k = 100$	
		α	$h = 0.025$	$h = 0.040$	$h = 0.025$	$h = 0.040$
iid observations		0.1	0.104	0.117	0.092	0.071
		0.05	0.051	0.053	0.052	0.029
		0.01	0.011	0.010	0.013	0.006
Linear trend	$b = 1$	0.1	0.831	0.682	0.539	0.375
		0.05	0.731	0.559	0.407	0.262
		0.01	0.505	0.338	0.207	0.095
	$b = 2$	0.1	0.989	0.960	0.843	0.703
		0.05	0.970	0.917	0.743	0.584
		0.01	0.888	0.740	0.480	0.271
Sin trend	$c = 1/4$	0.1	0.500	0.696	0.254	0.393
		0.05	0.388	0.597	0.165	0.292
		0.01	0.195	0.385	0.064	0.126
	$c = 1/2$	0.1	0.991	0.999	0.850	0.932
		0.05	0.976	0.994	0.770	0.886
		0.01	0.921	0.984	0.533	0.687

Estimated varying extreme value index: sin trends

(i) $c = 1/4$



(ii) $c = 1/2$



QQ plots for the estimates of $\gamma(1/2)$.

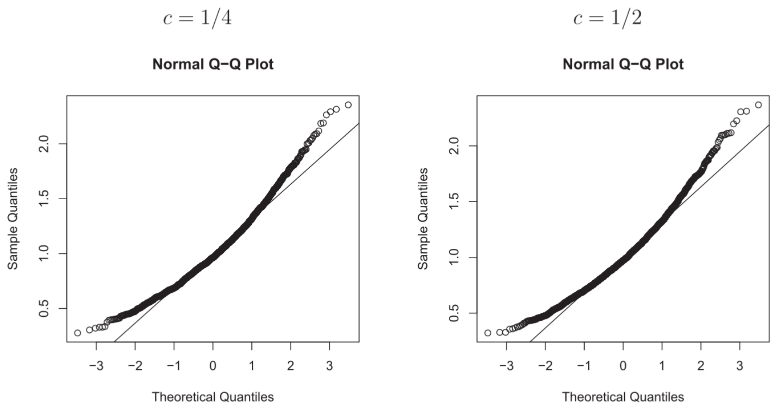


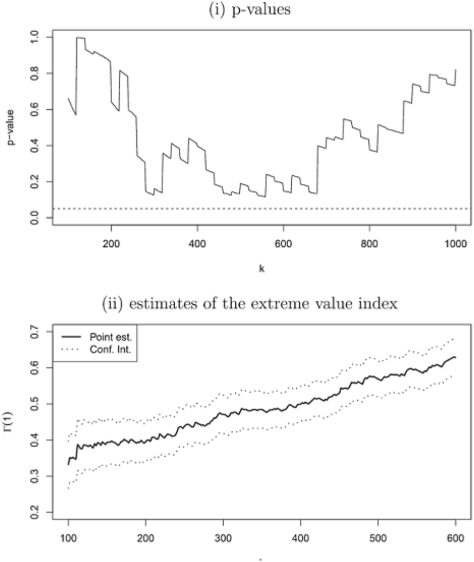
Figure 3. QQ plots for the estimates of $\gamma(1/2)$

Table of Contents

Precipitiattion at Saint-Martin-de-Londres

- A dataset consisting of the precipitation at Saint-Martin-de-Londres from 1976 to 2015, with 14,610 daily observations.
- Test the constancy of the extreme value indices over the entire period.
- Conclusion: do not reject the null hypothesis under the 5% significance level (the dash line).

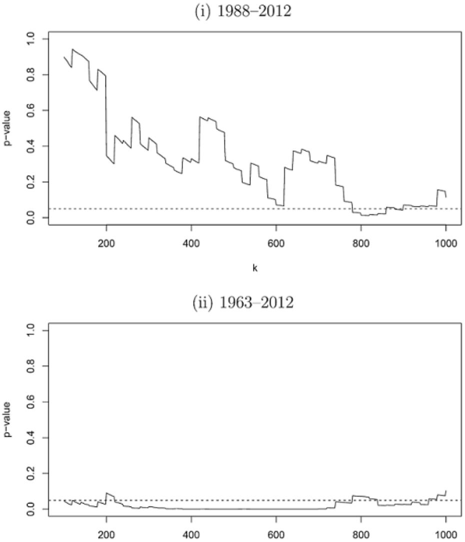
Testing the constancy of EVI:preciptation



Application 2: loss Return of S&P 500

- Datasets: S&P 500 index daily returns: $X_t = \log(P_t/P_{t+1})$
- Goal: test whether the extreme value indices holds constant over the a period.
- Period 1: 1988-2012 ($n = 6302$)
- Period 2: 1963-2012 ($n = 12586$)

Testing the constancy of EVI: S&P500

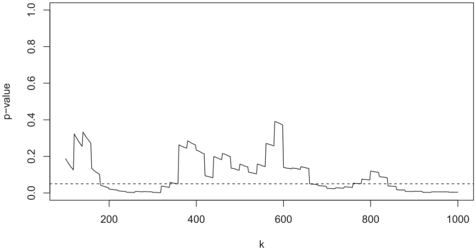


Serial Dependence

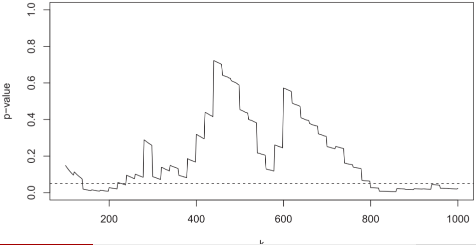
- Financial data such as stock returns exhibits serial dependence. Correspondingly, the critial value of the proposed test should be higher.
- By using the test based on assuming no serial dependence, we tent to over reject the null.
- Given that the analysis using the data from 1988 to 2012 did not reject the null, accounting for serial dependence may not alter the conclusion. However, the rejection result based on the data from 1963 to 2012 may suffer from the serial dependence issue.

Testing the constancy of EVI: S&P500

(i) odd days



(ii) even days



Thanks!