Trends in Extreme Value Index

Introduction

- Classical extreme value analysis assumes that the observations are independent and identically distributed (iid).
- Recent studies aim at dealing with the case when observations are drawn from different distributions.
- In this article, we aim at dealing with non-iid observations: we consider a continuously changing extreme value index and try to estimate the functional extreme value index accurately.

Introduction

- Consider a set of distribution functions $F_s(x)$ 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.
- To perform extreme value analysis, assume that $F_s \in D_{\gamma(s)}$, where D denotes the max-domain of attraction with corresponding extreme value index.
- In this article, we consider 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, based on the observations $X_1, \ldots, X_n$.

Introduction

- The idea for estimating $\gamma(s)$ locally is similar to the kernel density estimation.
- More specifically, we will use only observations X_i in the h-neighborhood of s, that is

$$i \in I_n(s) = \{|\frac{i}{n} - s| \leq h\},$$

  where $h := h(n)$ is the bandwidth such that as $n \to \infty, h \to \infty$ and $nh \to \infty$.
- Correspondingly there will be $[2hn]$ observations for $s \in [h, 1-h]$.
- To apply any known estimators for extreme value index, such as Hill estimator, we choose an intermediate sequence $k := k(n)$ such that $n \to \infty, k \to \infty$ and $k/n \to 0$.

Introduction

- Then we 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

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

- We start with considering the local asymptotic normality.

Conditions

- The second order condition: 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

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

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

Conditions

- $$h = h_n \to 0, k = k_n \to \infty, k_n/n \to 0, \frac{k_n h_n}{|\log h_n|} \to \infty, \tag{4}$$

- $$\Delta_{1,n} := \sqrt{k_n} \sup_{0 \le s \le 1} |A_s(\frac{n}{k_n})| \to 0, \tag{5}$$

- $$\Delta_{2,n} := \sqrt{k_n} \log k_n \sup_{|s_1 - s_2| \le 2h_n} |\gamma(s_1) - \gamma(s_2)| \to 0. \tag{6}$$

- $$\Delta_{3,n} := \sup_{|s_1 - s_2| \le h_n} \left| \frac{\frac{U_{s_1}\left(\frac{n}{k_n}\right)}{U_{s_2}\left(\frac{n}{k_n}\right)} - 1}{A_{s_2}\left(\frac{n}{k_n}\right)} \right| \to 0 \tag{7}$$

Theorem 2.1

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 $F_s \in D_{\gamma(s)}$ 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 $s \in (0, 1)$,

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

A Global Estimator

- Next, we 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.
- That prevents us from constructing a testing procedure.
- To achieve the stated goal, we consider the estimation $\Gamma(s) = \int_0^s \gamma(u)du$ and test the equivalent hypothesis that $\Gamma = \Gamma_0$.

A Global Estimator

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

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

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 $n \to \infty$,

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

Testing Trends in Extreme Value Indices

Similar to testing the specific trend in the "skedasis" function in Einmahl, de Haan, and Zhou (2016), we apply an equivalent test to test $H_0 : \Gamma(s) = \Gamma_0(s)$ for all s.

Clearly, one may construct a KS type test with testing statistic defined as

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

Then, Theorem 2.2 implies that under the hypothesis H_0, as $n \to \infty$,

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

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.

Simulation Study

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

For each sample, we 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 Frechet distribution.

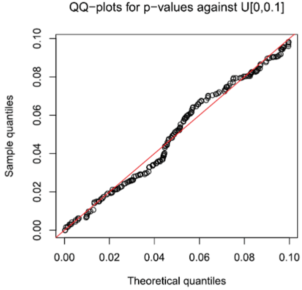
Simulation Study

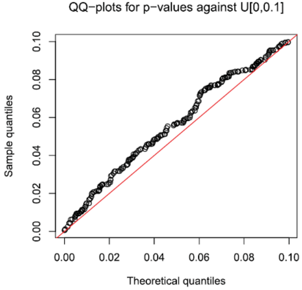
For the function $\gamma(s)$ we consider either a linear trend as $\gamma(s) = 1 + bs$ or a trend following the sin function as $\gamma(s) = 1 + c\sin(2\pi s)$.

If $b = 0$ or $c = 0$, the two model resemble the iid case.

We consider four alternative cases $b = 1, b = 2, c = 1/4$ and $c = 1/2$.

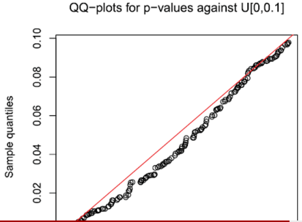
Simulation Study

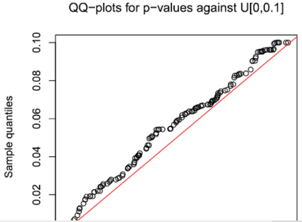




(ii) $k = 100$

$h = 0.025$ $h = 0.04$



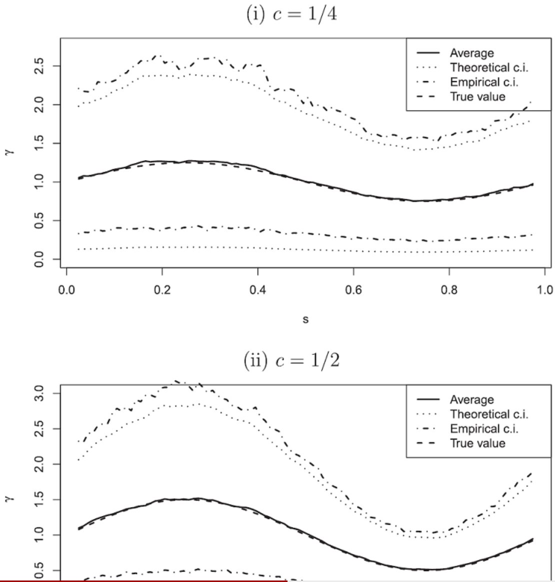


Simulation

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

Simulation

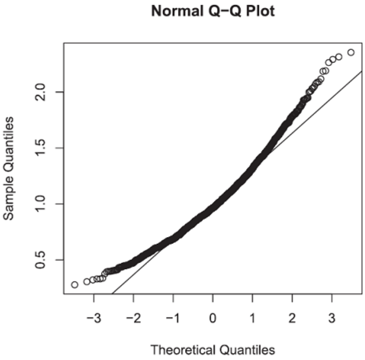


(i) c = 1/4
Average
Theoretical c.i.
Empirical c.i.
True value
γ
s
0.0
0.2
0.4
0.6
0.8
1.0
0.5
1.5
2.0
2.5
(ii) c = 1/2
Average
Theoretical c.i.
Empirical c.i.
True value
γ
3.0

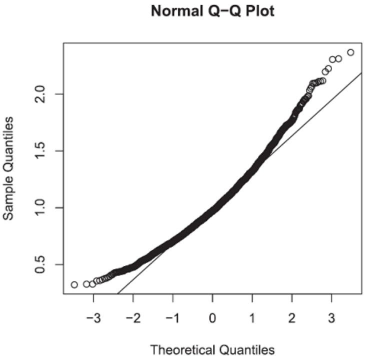
Simulation

We also compare the results for $n = 2000$.

$c = 1/4$



$c = 1/2$



Simulation

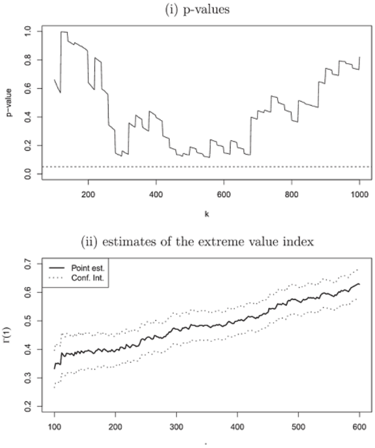
Table 2. Rejection rates in simulations: sample size $n = 2000$.

			$k = 100$		$k = 50$	
		α	$h = 0.025$	$h = 0.040$	$h = 0.025$	$h = 0.040$
iid observations		0.1	0.089	0.079	0.132	0.079
		0.05	0.048	0.035	0.074	0.037
		0.01	0.011	0.004	0.022	0.007
Linear trend	$b = 1$	0.1	0.517	0.397	0.367	0.204
		0.05	0.406	0.285	0.260	0.124
		0.01	0.186	0.114	0.099	0.028
	$b = 2$	0.1	0.837	0.683	0.570	0.375
		0.05	0.760	0.542	0.461	0.266
		0.01	0.490	0.281	0.208	0.106
Sin trend	$c = 1/4$	0.1	0.277	0.438	0.222	0.227
		0.05	0.188	0.334	0.129	0.157
		0.01	0.071	0.138	0.038	0.040
	$c = 1/2$	0.1	0.876	0.945	0.582	0.696
		0.05	0.811	0.908	0.470	0.579
		0.01	0.574	0.765	0.228	0.304

Application 1: Precipitation

- We employ a dataset consisting of the precipitation at Saint-Martin-de-Londres from 1976 to 2015, with 14,610 daily observations.
- We test the constancy of the extreme value indices over the entire period.
- We do not reject the null hypothesis under the 5% significance level (the dash line).
- We then estimate the constant extreme value index by applying the Hill estimator to all observations, that is, estimating $\Gamma(1)$.

Application 1: Precipitation



Application 2: Loss Returns of S&P500