

The paper I am going to present is “Trends in Extreme Value Indices”. This paper is published on Journal of the American Statistical Association in 2020. It is written by Laurens de Haan and Chen Zhou.

This is the outline of My presentation. First, I will give a introduction to this study. And then I will present the main methodology and show the main theorems. And Finally, a simulation study and two real data applications will be shown.

Now, we turn to the first part.

Classic extreme value analysis assumes that the observations are iid. However, as we know, the iid assumption may be too strong and real data may not satisfy the iid assumption. This study aims at dealing with case when observations are drawn from different distribution. They consider a continuously changing extreme value index and try to estimate the functional extreme value index accurately.

This is the model setting. Consider a set of distributions $F_s(x)$ for $s \in [0, 1]$. Now we have independent random variables X_1 to X_n and the distribution function of X_i is $F_{i/n}$.

Here, assume that the function $F_s(x)$ is continuous with respect to both s and x . Also, assume that the function F_s belongs to the maximum domain of attraction with extreme value index $\gamma(s)$. This means that the extreme value index of distribution function of X_i is $\gamma(i/n)$.

This paper considers the case that the function γ is positive. This means that X_1 to X_n are all heavy tailed distributions. Also, the function γ is assumed to be continuous.

The goal of this paper is to estimate the function γ and test the hypothesis that $\gamma = \gamma_0$ for some given function γ_0 . In particular, it can be applied to test whether the extreme value index remains at a constant level across all observations X_1 to X_n .

And this is the introduction part and do you have any questions about the model setting?

We first discuss how to estimate the function γ locally. And then I will present how to obtain a global estimator and how to use this to do the hypothesis testing that $\gamma = \gamma_0$ for some given function γ_0 .

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 . And the mathematical definition of the h -neighborhood is displayed as this, where h is the bandwidth and h satisfies that as n tends to infinity, h will tend to 0 and n times h will tend to infinity.

Then there will be $[2nh]$ observations in the h -neighborhood for each $s \in [h, 1-h]$. To apply any known estimators for extreme value index, such as the Hill estimator,

we choose an intermediate sequence $k = k(n)$ such that as $n \rightarrow \infty, k \rightarrow \infty$ and $k/n \rightarrow 0$. This choice of k is very standard in the extreme value analysis.

Then, we will use the top $[2kh]$ order statistics among the $[2nh]$ local observation in the h -neighborhood of s to estimate $\gamma(s)$. Rank the $[2nh]$ observations in the h -neighborhood of s as this. Then we can apply the Hill estimator to estimate $\gamma(s)$ locally. And the mathematical definition of the local Hill estimator is defined as this.

This is the main idea of how to estimate the function γ locally. To obtain the asymptotic properties of this estimator, we need some conditions.

The first condition is second order condition. Suppose that there exist a continuous negative function $\rho(s)$ and a set of function $A_s(t)$, such that the equation (3) holds. Here U_s is the left inverse function of 1 over $1 - F_s$. Assume a second order condition is quite standard and sometimes necessary for the extreme value analysis. What is different here is that this is a functional form of the second order condition.

Note that the approach can be regarded as a combination of kernel density estimation and extreme value statistics. To prove the asymptotic normality, we need to combine two limiting procedures. First, the observations used are from a h -neighborhood that is shrinking. Second, within each h -neighborhood, we need to apply a threshold to all observations that is increasing toward infinity. If the h -neighborhood shrinks too fast, there will not be sufficient observations in each neighborhood for statistical inference. If it shrinks too slowly, we would have involved too many observations with different extreme value indices such that the Hill estimation will be distorted. Therefore, the two limiting procedures have to be balanced such that the resulting estimators have proper asymptotic behavior. Then, we need some conditions for k and h . They are displayed as condition 4-7. Condition (4) ensures that the number of high order statistics used in each local interval tends to infinity. Recall that the local Hill estimator involves $2kh$ top order observations.

Condition (5) is the one usually required for extreme value analysis to guarantee to have no asymptotic bias in the estimator. In the iid case, we often assume that $\sqrt{k_n}A(n/k_n)$ tends to 0.

Condition (6) indeed puts a restriction on the smooth the function γ . And the k_n should be properly chosen so that condition (6) holds.

Note that $U_s(n/k_n)$ is the $1 - k_n/n$ quantile function. So condition (7) means that $(1 - k_n/n)$ -quantiles of distributions are sufficiently smooth in short interval.

Under the above conditions, the estimation for $\gamma(s)$, the local Hill estimator possess the asymptotic normality. This results is displayed in Theorem 2.1. This result is comparable with the asymptotic normality of Hill estimator in the iid case, but now the estimation is based on the observations with different extreme value indices. The speed of convergence is $\sqrt{2kh}$ because only the top $[2kh]$ order statistics are used in the estimation.

Here, I will give a sketch of the Proof. The proof for the local estimator is not difficult. If you are not interested in the proofs, you can omit this part.

First, we transform X_i to $U_{i/n}(Z_i)$ where Z_i are iid standard Pareto distributed random variables. Indeed, they are equal in distribution. This transformation is quite common in extreme value analysis. And the proof of the asymptotic normality of Hill estimator in iid case is also based on this transformation. Then rank Z_i in the h neighborhood of s into order statistics. In the iid case, the order statistics of X_i correspond to the order statistics of Z_i . Since here $U_{i/n}$ are different, the order statistics of X_i may not correspond to the order statistics of Z_i . Nevertheless, the order statistics of X can be bounded by the minimum and maximum of the U function corresponding to the order statistics of Z . Then by the definition of the local Hill estimator, which is based on the top order statistics of X , then we can obtain the upper bound of the local Hill estimator as this. We can also obtain the lower bound. Although here we use the maximum and minimum of the U function. Recall that the U function is quite smooth. Then by the aforementioned conditions, we have this lemma. We see that even U_s are different, in the short interval, the properties of regular varying still holds. Then we can apply this lemma to bound the local Hill estimator by applying $x = k/n$ times the order statistics of Z . Since this lemma requires that x in the interval $1/2$ to q_n , we need to verify this condition. The rest of the proofs are quite similar to that of the standard Hill estimator.

Now, recall that the goal is to test the hypothesis that $\gamma(s) = \gamma_0(s)$ for all $s \in [0, 1]$. Although we are able to estimate the function γ locally, we cannot construct a hypothesis testing directly. In addition, the local estimators converges with a slow speed of convergence $1/\sqrt{2kh}$. To achieve the hypothesis testing goal, consider the estimation $\Gamma(s)$, which is the integrate of γ function from 0 to s . $\Gamma = \Gamma_0$ is the equivalent hypothesis.

The function Γ is estimated by aggregating the local estimators of $\gamma(s)$ to a global estimator. Consider a discretized version of the local Hill estimator. And then define the estimator of estimator of $\Gamma(s)$ as the integral of the discretized version of local Hill estimator. However, the local Hill estimator is only defined for s less than $1 - h$. So, we need to extend the range of local Hill estimator to accommodate the case when $s > 1 - h$. This is just some accommodations.

Note that $\hat{\Gamma}_H(s)$ is a stochastic process defined on $[0, 1]$. Then, assume the same conditions as in Theorem 2.1, There exist a series of Brownian motions $W_n(s)$ such that $\hat{\Gamma}_H(s)$ can be approximated by this process.

There are two features to note here. Firstly, the convergence is uniformly for all $s \in [0, 1]$. Secondly, the speed of convergence for the global estimators $\hat{\Gamma}_H(s)$ is $1/\sqrt{k}$. And recall that the speed of convergence for the local estimator is $1/\sqrt{2kh}$.

The proofs for the above theorem is much more difficult than the proof of Theorem 2.1. The difficulties lie in the two parts. First, the global estimator is the integrate of the local estimator, so we need to show that the convergence for

the local Hill estimator is uniform for s . Second, The convergence in Theorem 2.2 is convergence in process. So some techniques related to tail empirical process should be used.

The full proof is quite difficult and I just give a brief sketch. Recall that we applied the discretization to the estimation of function γ . Here, apply the same discretization to the function γ . The main steps can be summarized as this. First, prove the results for the discretized version of γ . Then to handle the difference between γ and its discretized version. Since the global estimator involves the local estimators at $s = (2p - 1)h$ for $p = 1$ to $1/[2h]$. Recall that one important setp of proof of Theorem 2.1 is to show that this type of inequality hold. Since now we are dealing with different s , we need to show the inequality holds uniformly for $s = (2p - 1)h$, with $p = 1$ to $1/[2h]$. Proving this inequality is not easy , it requires a stronger version of Hoeffding's inequality.

As long as this inequality holds, we can obtain this uniform expansion by the uniform inequality of the regular varying function. It is obvious that $J_n^{(p)}$ is asymptotic normal. So we have proved the uniform convergence of the local Hill estimator. And for different p , the interval used for the local estimators are disjoint. So for different p , $J_n^{(p)}$ are independent. I think this is why the authors use this type of discretization. The rest of the proof follows by the theory of empirical process. It is still very difficult. If you have interest, we can discuss after the group meet.

Now, the above Theorem, Theorem 2.2 provides the possibility to test if the extreme value indicies follow a specific trend, that is $\gamma(s) = \gamma_0(s)$ for all $s \in [0, 1]$. Clearly, one may construct a Kolmogorov-Smirnov (KS) type test with the test statistic defined as this. Then, under the null hypothesis, this relation holds.

It is often of interest to test whether the extreme value index remains constant over time, without prior knowledge on the constant extreme value index, that is, $H_0 : \gamma(s) = \gamma$ 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} . It is straightforward to show that under H_0 , $\sqrt{k}\tilde{T}$ convergences in distribution to a supreme of a standard Brownian bridges defined on $[0, 1]$. Note that the limit distribution is identical to that in the classic KS test.

Now, we turn to the simulation study part. This simulation study is used to demonstrate the finite sample performance of the testing procedure using \tilde{T} . That is to test the hypotheis that the extreme value indices remain constant over time.

This is the simulation setting. This paper generate $m = 2000$ samples with $n = 5000$ observations in each sample. For the two parameters k and h , choose several combinations between $k = 100, 200$ and $h = 0.025, 0.04$. For each sample, simulate the observations from this data generating process $X_i = Z_i^{1/\gamma(i/n)}$, where Z_i are iid observations from the standard Frechet distribution.

For the function $\gamma(s)$, we consider either a linear trend as $\gamma(s) = 1 + bs$ or a

trend as $1 + c \sin(2\pi s)$. If $b = 0$ or $c = 0$, the two model resemble the iid case, that is, the null hypothesis that extreme value indices remain constant holds. We consider four alternative cases: $b = 1, b = 2, c = 1/4$ and $c = 1/2$. In total, we have 20 sets of simulations due to various choices of k, h and the model of $\gamma(s)$. For each simulated sample, we apply the test statistic \tilde{T} to test whether the extreme value indices remain constant and obtain the corresponding p -value.

For the simulations based $b = 0$ (or $c=0$), that is, when the null hypothesis holds. We know that under unull hypothesis, the P-value follows a uniform distribution. And then we can make a QQ plot bettween the simulated p-values aganist a uniform distribution. We see that the dots in QQ plots line up on the 45 degree in line. The plot confirm the validity of the test under null hypothesis.

Next, for all sets of simulations, calculate the rejection rate based on each significance level α as the percentage of the sample whose p value is less than α . α is chosen to be 0.01, 0.05 and 0.1. The rejection rates are reported in Table 1.

In the first panel, we observe that under the null hypothesis, the rejection rates, that is the type1 error, are close to the significance level. The difference between the two choice of k is very samll when $h = 0.025$.

In the next two panels, the linear and sin trend panel, the rejections rate can be read as the power of the test. Between the two choice of h , $h = 0.025$ leads to a slightly higher power for rejecting the linear trend while $h = 0.04$ leads to sightly higher power for the sin trend. Between the two choices of k , $k = 200$ leads to a much higher power in all the models. Therefore, choosing a higher k is preferred as long as the bias not an issue, whereas the choice of h depends on the shape of trend.

When comparing across models, the power is higher for $b = 2$ than for $b = 1$. And higher for $c = 1/2$ than for $c = 1/4$. This is in line with our intuition, The test is more powerfull to detect large deviation from the null hypothesis.

For the two sin trends. The authors plot the estimated $\gamma(s)$ and the 95% confidence interval for any given s . There are two ways to construct the confidence interval. The first is based on Theorem 2.1, that is Theoretical Confidence interval. Second, we can use the empirical confidence interval based on m samples. And this figure, $k = 200$ and $h = 0.025$.

From this figure, the average of estimates of trend are very close to the true value. And the empirical confidence interval is also close to the theoretical one. However, the empirical confidence interval is shifted slighted upware compared to the theoretical one. One explation is the estimation bias. However, since the average estimates is close to the true value. So, this explanation is not reasonable. The authors give another explanation. Note that the asymptotic normality requires a large value of $2kh$. However, $2kh = 10$ in this exmaple, which is relatively low. And the QQ plot suggests that this explanation may be reasonable.

Now, we turn to the Application part. The authors conduct two applications

to test whether the extreme value indices remain constant over time. The first application is about the precipitation at Saint-Martin-de-Londres, a small town in France, from 1976 to 2015, with 14610 daily observation. The obtained p-value against various levels of k are plotted in this figure. We see that the p-value is always greater than 0.05. So the conclusion is that do not reject the null hypothesis under the 5% significance level. The obtained estimates against various levels of k are shown in the lower panel. By choosing $k = 200$, the authors obtain that the estimate of the constant extreme value index is 0.395.

The second application is the daily loss return of the Standard and Poor 500 index. The goal is same as before. They want to test whether the extreme value indices remain constant over the whole period. They consider two periods. The first period is from 1988 to 2012, which consists of $n=6302$ observations. Another period is longer, from 1963 to 2012, which consists of $n=12586$ observations.

The obtained p-values against various level of k are shown in this figure. The upper one is for period 1. The lower one is for period 2. For the period 1, from 1988 to 2012, the null hypothesis is not rejected for k up to 750 under the 0.05 significance level. And we can think the extreme value indices remain constant over this period. And we reject the constant extreme value index hypothesis for period 2. And we conclude that there is a change of extreme value index during the period from 1963 to 2012.

One concern in the aforementioned analysis is that financial data such stock return exhibits serial dependence. The presence of serial dependence would in general enlarge the asymptotic variance of the local estimators for $\gamma(s)$. Corresponding, the critical value of the proposed test should be higher. By using the test based on assuming no serial dependence, we tend to over reject the null. Given that the analysis using data in period 1, from 1988 to 2012, did not reject null, accounting for serial dependence may not change the conclusion. However, the rejection result based on data from period 2, 1963 to 2012, may suffer from serial dependence issue.

The authors split the dataset into two subsets that consist of daily returns on the even and odd days, respectively. The split of the full data helps to reduce the serial dependence and data from each subset is more close to the iid assumption.

We observe that the null hypothesis is not rejected at 5% significance level, for the dataset containing daily returns on the even days only. For the daily returns on the odd days, the result is not conclusive either: for k ranging from 400 to 600, the null hypothesis is not rejected at 5% significance level. However for a lower choice of k , such as $k = 200$ or $k = 300$, the null is rejected under 5% significance level. Overall, the additional analysis reveals that the rejection result for the full dataset might be affected by the serial dependence. With accounting for serial dependence, there is no conclusive evidence that the extreme value index changes over this period.

And this is all for my presentation. Thanks for listening.