EXTREME VALUE ANALYSIS OF FINANCIAL MARKET

Author
Zijun Liu

Supervisor
Professor Wael Bahsoun

Module Code
22MAP301

Institution
Loughborough University

September 18, 2023

Contents

1 Introduction

This paper presents the analysis of the leading stock indexes over the world based on the extreme value theory. Block Maxima method, Peak Over Threshold method and generalised distributions are introduced to measure quantiles such as r-period repeat level, Value at Risk and Expected Shortfall of S&P 500, CAC 40 and FTSE 100 stock index. We focus on the daily log loss and define the extreme event on loss for the analysis based on the recorded economic recessions since 21st century. Such analysis will give the predictions about the expected loss within a period of time and will enable us to find the different attributes behind these stock indexes.

2 Review of Literature and Elementary Examples

Extreme value theory is a branch of statistics dealing with the extreme deviations from the median of probability distributions, which attempts to predict the occurrence of events that have never (or rarely) been observed from a given ordered sample of random variables.

The analysis established on such theory, named extreme value analysis (EVA), is widely used in many fields of study. For example, EVA is frequently used in the study of hydrology to estimate the probability of an unexpectedly large flooding event, such as the 100-year flood [1]. However, such application in the real world always requires a critical vision. In this chapter, we will have an overview of the underlying hypotheses and theories.

2.1 Univariate Extreme Value Theory

Let $X_1, \ldots, X_n$ be a be a sequence of independent identically distributed random variables with cumulative distribution function F and let $M_n = \max(X_1, \ldots, X_n)$ denotes the maximum. Then, the distribution of the maximum can be derived as follow:

$$\begin{aligned} \Pr(M_n \leq z) &= \Pr(X_1 \leq z, \ldots, X_n \leq z) \\ &= \Pr(X_1 \leq z) \cdots \Pr(X_n \leq z) \\ &= (F(z))^n. \end{aligned} \tag{1.1}$$

Furthermore, it shall be noted that dealing with minima follows the same approaches and in applications, all that needs to be done is reversing the signs of the observations by letting $\min(X_i) = -\max(-X_i)$ and apply the same procedures for maxima.

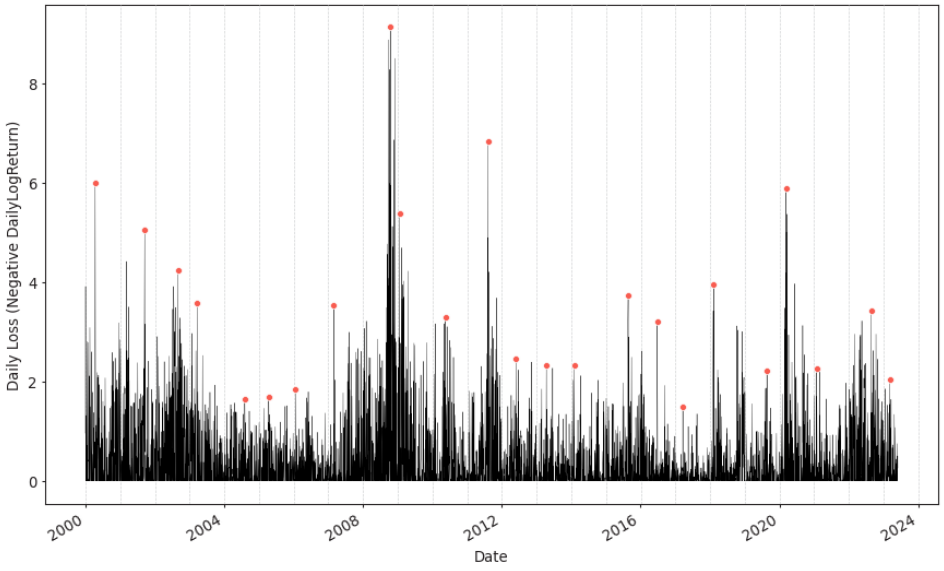
In practice, the distribution function F is unknown. To do the further analysis, there are two fundamental approaches, the Block Maxima method and the peak over threshold method, respectively.

2.2 Block Maxima Method

Block Maxima (BM) method relies on deriving maxima series in each block. Dividing the data into non-overlapping blocks of a fixed size (e.g., monthly, quarterly, or annually), there is the maximum within each block. Therefore, for each block, you will have a sample of the

extreme values. (See Figure 1. Each red point is the the maximum in the block and the grey vertical dotted lines are the separations between two blocks.)

Figure 1: Block Maxima Method.



We know F cannot be obtained in reality. Instead, we have an approach based on asymptotic arguments, the limiting distributions for $(F(z))^n$ as $n \to \infty$. The related fundamental theorem is named the extremal types theorem, presented by Fisher and Tippett [2] in 1928. According to this theorem, as the sample size n gets large, the distribution of extremes denoted M_n converges at the generalised extreme value (GEV) distribution [3].

2.2.1 GEV Distribution

Theorem 2.1. Extremal Types Theorem
Let (X_n) be independent with distribution function F and let $M_n = \max_{1 \le i \le n} X_{(i)}$. If there exist constants $a_n > 0$, b_n and a non-degenerate distribution function G such that as $n \to \infty$,

$$\Pr\left(\frac{M_n - b_n}{a_n} \le z\right) = (F(z))^n \to G(z), \tag{1.2}$$

then G must be of the same type as one of the three extreme value classes below.

Type I (Gumbel): $G(z) = \exp\left\{-\exp\left(-\left(\frac{z-b}{a}\right)\right)\right\}$ for $z \in \mathbb{R}$;

Type II (Fréchet): $G(z) = \begin{cases} \exp\left\{-\left(\frac{z-b}{a}\right)^{-\alpha}\right\} & z > b \\ 0 & z \le b \end{cases}$ for some $\alpha > 0$;

Type III (Weibull): $G(z) = \begin{cases} \exp\left\{-\left(-\left(\frac{z-b}{a}\right)\right)^{\alpha}\right\} & z < b \\ 1 & z \ge b \end{cases}$ for some $\alpha > 0$.

The generalised extreme value (GEV) distribution is derived by combining these 3 classes, whose cumulative distribution function (CDF) is

$$G(z;\mu,\sigma,\xi) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}, \tag{1.3}$$

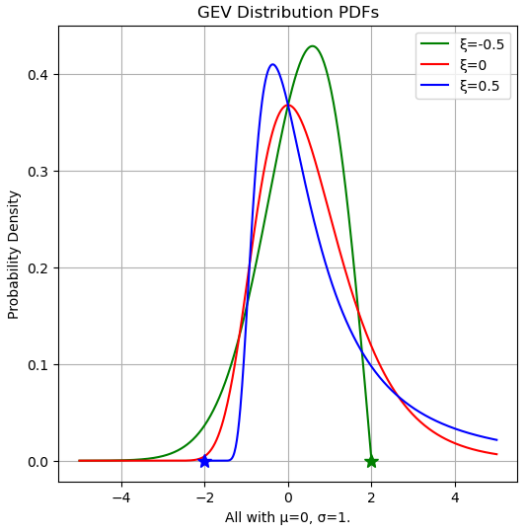
where $a_+ = \max(0,a)$ and $-\infty < \mu < \infty, \sigma > 0, \xi < \infty$ are location, scale and shape parameters, respectively. The situation where $\xi = 0$ is given by, as $\xi \to 0$,

$$G(z;\mu,\sigma) = \exp\left\{-\exp\left(\frac{z-\mu}{\sigma}\right)\right\}. \tag{1.4}$$

The probability density function (PDF) of GEV distribution can be found by differentiating (1.3) and (1.4), which is

$$\begin{aligned} &\frac{1}{\sigma}\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]_+^{-(1/\xi+1)} \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\} && \text{if } \xi \neq 0, \\ &\frac{1}{\sigma}\exp\left\{-\frac{z-\mu}{\sigma}\right\}\exp\left\{-\exp\left(-\frac{z-\mu}{\sigma}\right)\right\} && \text{if } \xi = 0. \end{aligned} \tag{1.5}$$

Figure 2: PDFs of GEV.



As seen in Figure 2, the shape parameter ξ governs the distributions, where the red light-tailed curve is the PDF of the Gumbel distribution, the blue heavy-tailed curve is the PDF of the Fréchet distribution, and the green bounded curve is the PDF of the Weibull distribution.

For both the Gumbel and Fréchet distributions, the limiting distribution G is unbounded, which means their upper–endpoint tends to ∞. Of these two, the Fréchet distribution has a heavier tail. For the Weibull distribution, the limiting distribution is bounded.

It should be noted that the Extremal Types Theorem does not ensure the existence of a non–degenerate limit for M_n. However, when such a distribution does exist, the limiting

distribution of sample maximums will follow one of the distributions in GEV, no matter what its parent distribution F is.

2.2.2 Elementary Examples of GEV

Example 1
Suppose $X_1, X_2, \ldots, X_n$ is a sequence of independent Fréchet variables, that is

$$F(x) = e^{-1/x}, \quad x > 0.$$

Show the limit distribution of $(M_n - b_n)/a_n$ is one of extreme value types if $a_n = n$ and $b_n = 0$, and identify its distribution.

Before we move forward, it should be noticed that if as $n \to \infty$, $\frac{M_n - b_n}{a_n} \to \mathcal{G}(\mu, \sigma, \xi)$, then $M_n \to \mathcal{G}(\mu^*, \sigma^*, \xi)$ where the a_n and b_n is absorbed into μ^* and σ^*.

Proof:
We want the distribution of $(M_n - b_n)/a_n$, i.e.

$$\begin{aligned}
\Pr\left\{\frac{M_n - b_n}{a_n} \le z\right\} &= \Pr\left\{\frac{\max(X_1, \ldots, X_n) - b_n}{a_n} \le z\right\} \\
&= \Pr\left\{\frac{X_1 - b_n}{a_n} \le z, \ldots, \frac{X_n - b_n}{a_n} \le z\right\} \\
&= \Pr\{X_1 \le a_n z + b_n\} \times \cdots \\
&= \left[\exp\left\{-\frac{1}{a_n z + b_n}\right\}\right]^n.
\end{aligned}$$

Letting $a_n = n$ and $b_n = 0$ gives

$$\left[\exp\left\{-\frac{1}{nz}\right\}\right]^n = e^{-1/z},$$

which is the Frechet distribution with $\alpha = 1$, i.e. a unit Fréchet distribution.

Example 2
For the standardized Fréchet distribution with $\xi = 0.3$, calculate its 5% quantile, 95% quantile, and the 99% quantile.

Note that the quantiles associated with the GEV distribution can be obtain by inverting (1.3) and (1.4) to be

$$\begin{aligned}
\mathrm{z} &= \mu - \frac{\sigma}{\xi}\left[1 - (-\ln(\mathrm{p}))\right]^{-\xi} \qquad \text{(Fréchet, } \xi > 0\text{)}, \\
\mathrm{z} &= \mu - \sigma \ln(-\ln(\mathrm{p})) \qquad \text{(Gumbel, } \xi = 0\text{)},
\end{aligned}$$

where p is the chosen probability, and z is the observed value associated with the pth quantile.

Thus,

The 5% quanile is : $-\frac{1}{0.3}\left[1-(-\ln(0.05))^{-0.3}\right]=-0.9348$,
The 95% quantile is : $-\frac{1}{0.3}\left[1-(-\ln(0.95))^{-0.3}\right]=4.7924$,
The 99% quantile is : $-\frac{1}{0.3}\left[1-(-\ln(0.99))^{-0.3}\right]=9.9169$.

2.2.3 Parameter Estimate

Maximum likelihood estimation (MLE) [4] is the most common approach to fitting the GEV to the sample of block maximums M_n. MLE is a statistical method used to estimate the parameters of a probability distribution that best explain the observed data. The fundamental idea behind MLE is to find the values of the model's parameters that maximise its likelihood function, which measures how likely the observed data is under the assumed model.

Suppose that $X=\{x_1,\ldots,x_n\}$ is an independent random sample from the GEV distribution, then the log-likelihood is given by

$$L(\mu,\sigma,\xi;X)=\prod_{i=1}^{m} g\left(x_i;\mu,\sigma,\xi\right), \tag{1.6}$$

where g is the PDF of GEV and can be found, with (1.5), to be

$$\frac{1}{\sigma}\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-(1/\xi+1)}\exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}. \tag{1.7}$$

With (1.6) and (1.7), the likelihood is given as follow:

$$L(\mu,\sigma,\xi;X)=\prod_{i=1}^{m}\frac{1}{\sigma}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]_+^{-(1/\xi+1)}\times\exp\left\{-\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}, \tag{1.8}$$

and the log–likelihood is given as follow:

$$\begin{aligned}\ell(\mu,\sigma,\xi;X)=&\sum_{i=1}^{m}\log\sigma^{-1}+\sum_{i=1}^{m}\log\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]_+^{-(1/\xi+1)}\\&-\sum_{i=1}^{m}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]_+^{-1/\xi}\\=&-m\log\sigma-(1/\xi+1)\sum_{i=1}^{m}\log\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]_+\\&-\sum_{i=1}^{m}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]_+^{-1/\xi}.\end{aligned} \tag{1.9}$$

We want to find the values at $\frac{\partial\ell}{\partial\mu}=\frac{\partial\ell}{\partial\sigma}=\frac{\partial\ell}{\partial\xi}=0$. For example,

$$\frac{\partial\ell}{\partial\mu}=\frac{\xi+1}{\sigma}\sum_{i=1}^{m}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-1}+\frac{1}{\sigma}\sum_{i=1}^{m}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-(1/\xi+1)}=0.$$

The other two log-likelihood equations can be obtained by the same way and are omitted here.

However, in practice, the log-likelihood equations cannot be solved analytically, so there are no closed–form solutions for $\hat{\mu}, \hat{\sigma}, \hat{\xi}$ for the GEV. Instead, we can get around this by using a numerical method to obtain (approximate) solutions to the log-likelihood equations and this step will be done using Python.

2.2.4 Quantiles in GEV on BM Method

After we have the parameters, we are able to calculate some quantiles, such as r-period repeat level [5], value at risk (VaR) [6] and expected shortfall (ES) [7]. These 3 quantiles are very useful in risk management and we will use them to explain the EVA of corresponding applications in Chapter 4.

The r-period repeat level of a particular event is the inverse of the probability that the event will be exceeded in any given period. Let r-period repeat level z_r for any period in GEV distribution be $1 - \frac{1}{r}$ and we have

$$\begin{aligned}
\text{GEV}\,(z_r) &= 1 - \frac{1}{r} \\
z_r &= \text{GEV}^{-1}(1 - \frac{1}{r}) \\
z_r &= \mu - \frac{\sigma}{\xi}\left[1 - (-\ln(1 - \frac{1}{r}))^{-\xi}\right] \quad \text{if } \xi \neq 0, \\
&= \mu - \sigma \ln(-\ln(1 - \frac{1}{r})) \quad \text{if } \xi = 0.
\end{aligned}$$

VaR estimates how much a set of investments might lose with a given probability in a set time period, such as a day. It is typically used by researchers in the financial industry to gauge the amount of assets needed to cover possible losses. If the sample of block maximum loss X follows the GEV and each block has k observations, then the VaR at the different confidence level CI is equal to

$$\begin{aligned}
\text{VaR}_{CI}(X) &= \mu - \frac{\sigma}{\xi}\left[1 - (-k \ln(CI))^{-\xi}\right] \quad \text{if } \xi \neq 0, \\
&= \mu - \sigma \ln(-k \ln(CI)) \quad \text{if } \xi = 0,
\end{aligned}$$

(Andersson) [8]. ES is another useful measure in risk management. As the loss exceeds the VaR, ES is used to determine the maximum loss within a given time period. The ES at the different confidence level CI is equal to

$$\begin{aligned}
\text{ES}_{CI}(X) &= \mu + \frac{a}{(1 - CI^k)\xi}[\gamma(1 - \xi, -k \ln(CI)) - 1 + CI^k)] \quad \text{if } \xi \neq 0, \\
&= \mu + \frac{\sigma}{1 - CI^k}[y - \text{li}(CI^k) + CI^k \ln(-k \ln(CI))] \quad \text{if } \xi = 0,
\end{aligned}$$

where $\gamma(s, x)$ is the lower incomplete gamma function, y is the Euler-Mascheroni constant and $\text{li}(x)$ is the logarithmic integral function [7].

Example 3

Suppose we fit the GEV to the set of given sample of annual maximums and obtain esti-

mates of the location, scale and shape as $\hat{\mu}, \hat{\sigma}$ and $\hat{\xi}$, respectively. Suppose we require an estimate of z_{100}, the given maximum we might expect to be exceeded once in a hundred years.

Then we can write down the following statement.

$$\begin{aligned}
\Pr\left(\text{annual maximum } > z_{100}\right) &= \frac{1}{100}, \\
\implies 1 - \Pr\left(\text{annual maximum } \leq z_{100}\right) &= \frac{1}{100}, \\
\implies 1 - G\left(\hat{z}_{100}; \hat{\mu}, \hat{\sigma}, \hat{\xi}\right) &= \frac{1}{100}, \\
\implies 1 - \exp\left\{-\left[1 + \hat{\xi}\left(\frac{\hat{\boldsymbol{z}}_{100} - \hat{\mu}}{\hat{\sigma}}\right)\right]^{-1/\hat{\xi}}\right\} &= \frac{1}{100}, \\
\implies \exp\left\{-\left[1 + \hat{\xi}\left(\frac{\hat{\boldsymbol{z}}_{100} - \hat{\mu}}{\hat{\sigma}}\right)\right]^{-1/\hat{\xi}}\right\} &= \frac{99}{100}.
\end{aligned}$$

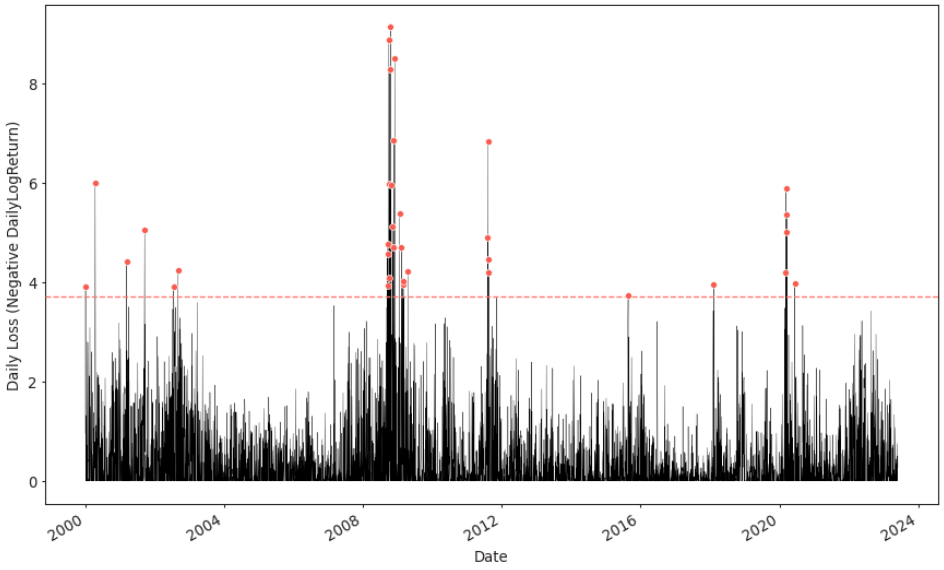
Finally, we have

$$z_{100} = \hat{\mu} + \frac{\hat{o}}{\hat{\xi}}\left[(-\ln(0.99))^{-\hat{\xi}} - 1\right].$$

2.3 Peak over Threshold Method

Peak over threshold (POT) method relies on extracting peak values above (below) a certain threshold from a continuous record. Unlike the BM method, the POT method does not require non-overlapping blocks. However, choosing an appropriate threshold value should be considered deliberately. The threshold should be high enough to ensure that the data over this threshold are rare events and can be approximated by the general Pareto distribution (GPD) [9]. (See Figure 3. The red horizontal line is the threshold and the red points above such line are considered to be extreme values.)

Figure 3: POT Method.



The true distribution is unknown, however, the Pickands–Balkema–De Haan theorem [10] which will be demonstrated below gives the asymptotic tail distribution of these random variables.

2.3.1 Generalized Pareto Distribution

For an unknown distribution function F of a random variable X, the conditional distribution function F_u of the variable X above a certain threshold u is defined as

$$F_u(z) = \Pr\left(X - u \leq z \mid X > u\right) = \frac{F(z+u) - F(u)}{1 - F(u)}, \tag{2.1}$$

for $0 < z \leq x_F - u$, where x_F is either the finite or infinite right endpoint of the underlying distribution F. The function F_u describes the distribution of the excess value over a threshold u, given that the threshold is exceeded. The proof is given in Appendix I.

Theorem 2.2. Pickands–Balkema–De Haan Theorem
Let $X_1, \ldots, X_n$ be a be a sequence of independent and identically distributed random variables and let F_u be their conditional exceedance distribution function, then for a large class of underlying distribution functions F and large u, F_u is well approximated by the GPD, that is

$$F_u(z) \to G_{\xi,\sigma}(z) \quad \text{, as } u \to \infty.$$

Its CDF is defined as

$$G_{\xi,\sigma}(z) = \begin{cases} 1 - (1 + \frac{\xi z}{\sigma})^{-1/\xi} & \text{for } \xi \neq 0, \\ 1 - e^{\frac{-z}{\sigma}} & \text{for } \xi = 0, \end{cases} \tag{2.2}$$

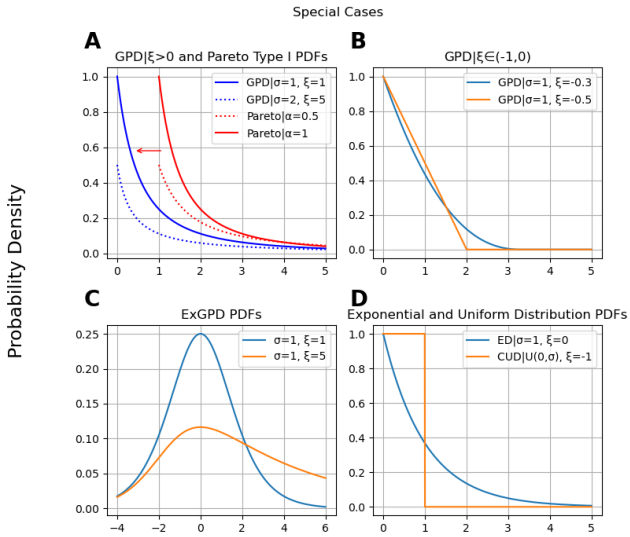
where $\sigma > 0$, $z \geq 0$ for $\xi \geq 0$ and $0 \leq z \leq -\sigma/\xi$ for $\xi < 0$. The corresponding PDF is defined as

$$f_{\xi,\sigma}(z) = \begin{cases} (1 + \frac{\xi z}{\sigma})^{-\frac{\xi+1}{\xi}} & \text{for } \xi \neq 0, \\ e^{\frac{-z}{\sigma}} & \text{for } \xi = 0. \end{cases}$$

As seen above, such approximated GPD has only two parameters: a positive scale parameter, σ; a shape parameter, ξ that can be positive, zero, or negative. The class of GPD with shape parameter less than 1 is omitted here since such distribution is so unnatural and is meaningless under most circumstances.

Then there are special cases depending on ξ, such as: exponential distribution with mean σ, if $\xi = 0$ (see Figure 4 D); uniform distribution on $[0, \sigma]$, if $\xi = -1$ (see Figure 4 D); lomax distribution if $\xi \in (-1, 0) \cup (0, \infty)$, which is a shift of Pareto Type I distribution and so that its support is shifted from 1 to 0 (see Figure 4 A) for $\xi > 0$ and is bounded (see Figure 4 B) for $\xi \in (-1, 0)$. Another special case is that if $X \sim GPD(\mu = 0, \sigma, \xi)$, then $Y = \log(X) \sim \text{exGPD}(\sigma, \xi)$, which is named the exponentiated generalized pareto distribution [11] (see Figure 4 C).

Figure 4: A. GPD cases and Pareto cases; B. Exponentiated GPD cases; C. Uniform cases; D. Exponential cases.



As seen in Figure 4, the GPD in sub-figure A and sub-figure B seems to be applicable and in practice, the shape parameter of GPD is most likely to lie in $(-1, 1]$. This reason will be explained in detail in the section 1.3.3 Parameter Estimation below in this paper. For ExGPD, it is an interesting type of distribution and will be mentioned further in the part of applications in this paper. However, the uniform distribution is not useful and will not be discussed further.

Let us go back to the parameters of GPD. It should be noticed that the threshold u, not the scale parameter μ, can be shown in the CDF with (2.2) by letting $z = x_F - u$, then

$$G_{\xi,\sigma}(z) = 1 - (1 + \frac{\xi(x_F - u)}{\sigma})^{-1/\xi} \quad \text{for } \xi \neq 0,$$

where $u < x_F$, given by the initial condition in (2.1) where $0 < z = x_F - u$. An interesting fact here is that in a true GPD distribution, rather than only approaching the approximation, the scale parameter σ is dependent on the threshold u, but the shape parameter ξ is not.

The reason for the dependence of the scale parameter on the threshold lies in the behaviour of extreme events and the underlying tail behaviour of the distribution. When the exceedances in POT method follow a true GPD, the scale parameter represents the dispersion or variability of these extreme events beyond the threshold. As the threshold is set higher, only more extreme events will be considered, which means consequently, their variability tends to increase.

However, the shape parameter in a true GPD determines the tail behaviour of the distribution beyond the threshold. In the POT method, since the shape parameter reflects the tail behaviour of the distribution and is determined by the extreme values themselves, it remains constant regardless of the threshold chosen.

In summary, if the distribution of the exceedances has an exact GPD, then increasing the threshold from u to u^* will result in another GPD with the same shape parameter ξ. The formula for the adjusted scale parameter σ^* can be derived from the proof in Appendix I as follow:

$$\begin{aligned} \tilde{\sigma} &= \sigma + \xi(u - \mu), \\ \implies \sigma^* &= \sigma + \xi\,(u^* - u)\,, \end{aligned}$$

for a new threshold u^* where $u^* > u$ and ideally, $u = \mu$. This formula implies that the scale parameter σ^* increases when we move to a higher threshold ($u^* > u$) if $\xi > 0$ and decreases when we move to a higher threshold ($u^* > u$) if $\xi < 0$. Furthermore, there is no change in the scale parameter as we move to a higher threshold if the distribution is exponential ($\xi = 0$).

2.3.2 Elementary Examples of GPD

Example 1

Suppose $X_1, X_2, \ldots, X_n$ is a sequence of independent exp(1) random variables. Show that the limiting distribution of threshold exceedances belongs to GPD.

Proof:

We know if $X_i \sim \exp(1)$, then $F(x) = 1 - e^{-x}$ for $x > 0$. By (2.1), we have

$$\begin{aligned} \Pr\left(X - u > z \mid X > u\right) &= \frac{1 - F(z+u)}{1 - F(u)}, \\ &= \frac{1 - \left(1 - e^{-(z+u)}\right)}{1 - (1 - e^{-u})}, \\ &= \frac{e^{-(z+u)}}{e^{-u}}, \\ &= e^{-z}, \quad z > 0. \end{aligned}$$

Then, we have

$$\implies F_u(z) = 1 - e^{-z}, \quad z > 0.$$

Thus, by (2.2), such limit distribution is an exponential distribution, i.e. GPD with $\sigma = 1$ and $\xi = 0$. Furthermore, the limit distribution here is GPD$(1, 0)$.

Example 2

We consider the case when $\xi > 0$ and $z \in \left(0, \frac{\sigma}{\xi}\right)$. Let $Y \sim \exp(1)$, then $X = \frac{\sigma}{\xi}\left(1 - e^{-\xi Y}\right)$ is a GPD(σ, ξ) random variable.

Proof:

$$
\begin{aligned}
F(z) = \Pr(X < z) &= \Pr\left(\frac{\sigma}{\xi}\left(1 - e^{-\xi Y}\right) < z\right) \\
&= \Pr\left(1 - e^{-\xi Y} < \frac{\xi z}{\sigma}\right) \\
&= \Pr\left(1 - \frac{\xi z}{\sigma} < e^{-\xi Y}\right) \\
&= \Pr\left(\ln\left(1 - \frac{\xi z}{\sigma}\right) < -\xi Y\right) \\
&= \Pr\left(Y < -\frac{1}{\xi}\ln\left(1 - \frac{\xi z}{\sigma}\right)\right) \\
&= \Pr\left(Y < -\ln\left(1 - \frac{\xi z}{\sigma}\right)^{\frac{1}{\xi}}\right) \\
\overset{Y\sim\exp(1)}{\Longrightarrow} &= 1 - e^{\ln\left(1-\frac{\xi z}{\sigma}\right)^{\frac{1}{\xi}}} \\
&= 1 - \left(1 - \frac{\xi z}{\sigma}\right)^{\frac{1}{\xi}} \quad \text{for } z \in \left(0, \frac{\sigma}{\xi}\right) \text{ and } \xi > 0
\end{aligned}
$$

Example 3

In this example, we will learn how the threshold u affects the fit of GPD model by reading some visualizations. Meanwhile, some researchers usually use AIC to test models, but AIC is not a good test in the EVA and we will explain the reasons in this example.

In practice, the exceedances over the threshold may not have a true GPD and thus the theory about the adjusted scale parameter mentioned in section 1.3.1 will fail. However, the potential pattern still works, which means $\sigma^* > \sigma$ if $u^* > u$ for $\xi > 0$ and so on (see Figure 5. B).

Figure 5: A. GPD quantile plot on different thresholds;
B. Parameters and AIC value

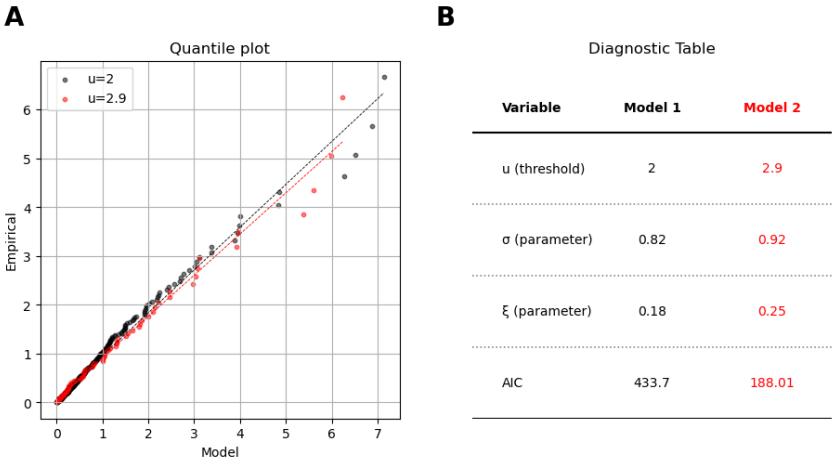


Figure 5. A shows two samples of exceedances and regression lines over different thresholds in a quantile plot; Figure 5. B shows the values of thresholds, 2 parameters and Akaike

Information Criterion (AIC) [12]. Here, AIC $= 2k - 2\ln(L)$ where k denotes the number of parameters and L denotes the maximized value of the likelihood function. By definition, the smaller the AIC value is, the better the fit of the model.

The sample of black points is more close to 45 degree line than the sample of red points is. Theoretically, the exceedances over 2 fits the GPD model better than exceedances over 2.9 does. However, the AIC value gives an contradiction because AIC of model 1 is much higher than which of model 2.

It should be noticed that in POT method, each time the threshold is higher, the AIC is lower; in BM method, each time the block size is larger, the AIC is lower. One of the main causes here is that a higher threshold in POT or a larger block size in BM leads to more extreme values which are fewer in number, resulting in a simpler model with fewer data points to estimate parameters for. Therefore, we will exclude AIC test in this paper. Instead, the diagnostic plot will be introduced in chapter 3.

This example simply explains the importance of threshold in POT method. We will talk further about specific ways of choosing the threshold values concerned with bias, efficiency and etc. in the chapter 4 while dealing with the applications.

2.3.3 Parameter Estimate

In this section, we will talk about how to solve (σ, ξ) of GPD. It starts with the traditional MLE method published by Grimshaw (1993) [13] but the numerical approach will be the new method proposed by Zhang (2010) [14].

Suppose that $X = \{X_1, \ldots, X_n\}$ is a random sample from the GPD with scale σ and shape ξ, and let $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$, then the log-likelihood is given by

$$\ell(\sigma, \xi; X) = \begin{cases} -n\ln\sigma + \left(\frac{1}{\xi} - 1\right)\sum_{i=1}^{n} \ln\left(1 - \frac{\xi X_i}{\sigma}\right), & \xi \neq 0, \\ -n\ln\sigma - \frac{1}{\sigma}\sum_{i=1}^{n} X_i, & \xi = 0, \end{cases} \tag{2.3}$$

where $\sigma > 0$ for $\xi \leq 0$ and $\sigma > \xi X_{(n)}$ for $\xi > 0$. If $\xi > 1$, there is no maximum likelihood estimate since, for any $\xi > 1$,

$$\lim_{\sigma/\xi \to X_{(n)}^{+}} \ell(\sigma, \xi; X) = \infty.$$

Therefore, to obtain a finite maximum of the GPD log-likelihood, the constraint $\xi \leq 1$ must be imposed. Next, we think about the special case of $\xi = 0$. For $\xi = 0$, we have

$$\begin{aligned} \lim_{\xi \to 0} \frac{\partial \ell(\sigma, \xi; X)}{\partial \xi} &= \sum_{i=1}^{n} \frac{X_i^2}{2\sigma^2} - \sum_{i=1}^{n} \frac{X_i}{\sigma} \\ \lim_{\xi \to 0} \frac{\partial \ell(\sigma, \xi; X)}{\partial \sigma} &= \frac{1}{\sigma}\left(\sum_{i=1}^{n} \frac{X_i}{\sigma} - n\right). \end{aligned} \tag{2.4}$$

Then (2.4) =0 $\iff$ $\sum_{i=1}^{n} X_i^2 = (2/n)(\sum_{i=1}^{n} X_i)^2$. Generally, such condition will not be satisfied, then the case $\xi = 0$ can be eliminated from consideration. Hence, the maximum likelihood estimates now becomes an optimization on the constrained space $\mathscr{A} = \{\xi < 0, \sigma > 0\} \cup \{0 < \xi \leq 1, \sigma/\xi > X_n\}$. However, the boundary condition of $\mathscr{A}$, where $\xi = 1$, will not be investigated in this paper because the given data in applications below is never going to have shape parameter of 1. Now the new constrained space is $\mathscr{A}^* = \{\xi < 0, \sigma > 0\} \cup \{0 < \xi < 1, \sigma/\xi > X_n\}$.

For local maximum on $\mathscr{A}^*$, we have

$$\begin{aligned} \frac{\partial \ell(\sigma, \xi; X)}{\partial \xi} &= \frac{n}{\xi}\left(\frac{1}{\xi} - 1\right) - \frac{1}{\xi^2}\sum_{i=1}^{n} \ln\left(1 - \frac{\xi X_i}{\sigma}\right) - \frac{1}{\xi}\left(\frac{1}{\xi} - 1\right)\sum_{i=1}^{n}\left(1 - \frac{\xi X_i}{\sigma}\right)^{-1}, \\ \frac{\partial \ell(\sigma, \xi; X)}{\partial \sigma} &= -\frac{n}{\xi\sigma} + \frac{1}{\sigma}\left(\frac{1}{\xi} - 1\right)\sum_{i=1}^{n}\left(1 - \frac{\xi X_i}{\sigma}\right)^{-1}. \end{aligned} \tag{2.5}$$

Then let (2.5)=0, we have

$$\left\{ \begin{array}{l} \frac{\partial \ell}{\partial \xi} = 0 \\ \frac{\partial \ell}{\partial \sigma} = 0 \end{array} \right\} \Rightarrow \left\{ \begin{array}{c} \left[1 + (1/n)\sum_{i=1}^{n} \ln\left(1 - \frac{\xi X_i}{\sigma}\right)\right] \cdot \left[(1/n)\sum_{i=1}^{n}\left(1 + \frac{\xi X_i}{\sigma}\right)^{-1}\right] = 1 \\ \xi = -1/n\sum_{i=1}^{n} \ln\left(1 - \frac{\xi X_i}{\sigma}\right) \end{array} \right\}. \tag{2.6}$$

To solve (2.6), we consider the re-parameterization from (σ, ξ) to (θ, ξ) with $\theta = \xi/\sigma$ and by(2.3), we now have

$$\ell(\theta, \xi; X) = n\ln(\theta/\xi) + \left(\xi^{-1} - 1\right)\sum_{i=1}^{n} \ln\left(1 - \theta X_i\right),$$

and (2.6) turns into

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}\left(1 - \theta X_i\right)^{-1} - \left(1 + n^{-1}\sum_{i=1}^{n} \ln\left(1 - \theta X_i\right)\right)^{-1} &= 0, \\ \xi = -n^{-1}\sum_{i=1}^{n} \ln\left(1 - \theta X_i\right). \end{aligned} \tag{2.7}$$

As seen from (2.7), we now have the formula with the only variable θ. After having the value of θ, we can calculate (σ, ξ) by

$$\begin{aligned} \xi &= -n^{-1}\sum_{i=1}^{n} \ln\left(1 - \theta X_i\right), \\ \sigma &= \xi/\theta. \end{aligned}$$

Finally, we need to use the new numerical method proposed by Zhang (2010) to find θ. This new method has been verified to have the better performance than other existing methods have if $\xi \in [-1, 1/2]$ by Pham, M. H., Tsokos, C., & Choi, B.-J. (2018) [15]. The shape parameters of applications below are confirmed to be in this range. This new method will be done by using R.

2.3.4 Quantiles in GPD on POT Method

The deduction of quantiles in GPD are similar to which in GEV.

The r-period repeat level here is defined as

$$\begin{aligned} \mathrm{GPD}\,(z_r) &= 1 - \frac{1}{r} \\ \implies z_r &= u + \frac{\sigma}{\xi}\left((\frac{1}{r})^{-\xi} - 1\right) \quad \text{if } \xi \neq 0, \\ &= u - \sigma \ln(1-\alpha) \qquad\quad \text{if } \xi = 0. \end{aligned}$$

The VaR at different CI is defined as

$$\begin{aligned} \mathrm{VaR}_{CI} &= u + \frac{\sigma}{\xi}\left(\left(\frac{N}{N_u}(1-CI)\right)^{-\xi} - 1\right) \quad \text{if } \xi \neq 0, \\ &= u + \sigma \ln\left(\frac{N}{N_u}(1-CI)\right) \qquad\quad \text{if } \xi = 0, \end{aligned}$$

where N is the number of data in the dataset and N_u is the number of exceedances over the threshold.

The formula for ES at different CI is

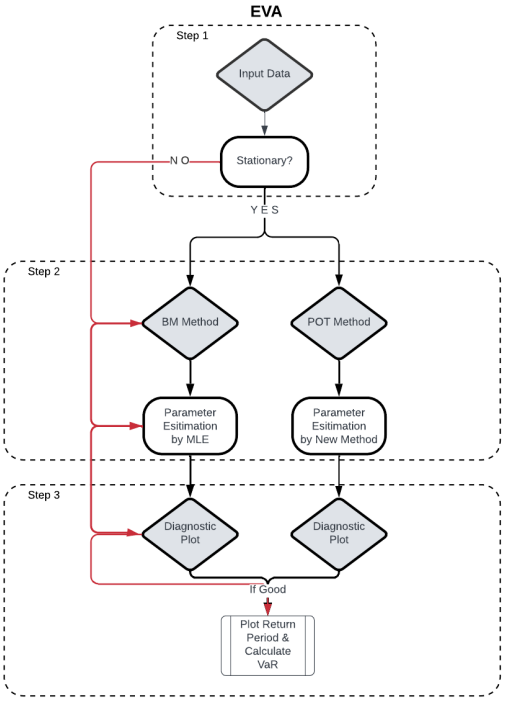
$$\begin{aligned} \mathrm{ES}_{CI} &= \frac{\mathrm{VaR}_{CI} + \sigma - \xi u}{1-\xi} \quad \text{if } \xi \neq 0, \\ &= \mathrm{VaR}_{\alpha} + \sigma \qquad\quad \text{if } \xi = 0, \end{aligned}$$

(Dowd, 2005) [16].

3 Methodology

In this chapter, we talk about the complete process (see Figure 6) while dealing with the real applications. This process will be done using Python and the link to the Python code is attached in Appendix II. The specific type of input data will be discussed in Chapter 4.

Figure 6: Structure.



In step 1, the input dataset is tested for its stationarity. For example, a time series dataset is stationary if it has constant variance but no trend or periodic fluctuations over time, which is also named a stationary process [17]. The stationarity tests used here are Rolling Statistics test [18] and Augmented Dickey–Fuller test [19]. To conclude the series is stationary, the rolling statistics test which is a visualisation should show the mean and standard deviation of such series do not vary with time and Dickey-Fuller test should have a test statistic value way less than the critical values for different significance levels at $(1\%, 5\%, 10\%)$. The reason to do such tests is that the theories discussed in Chapter 2 are under the assumption of stationarity. If the dataset is non-stationary, then one should use Non-Stationary Extreme Value Analysis (NEVA) [20] instead. Different from EVA, NEVA has to deal with the covariate which is the time here and can be expressed as a function of one or multiple parameters. For instance, by BM method, one should have (μ, σ, ξ) from a stationary dataset, however, one will have $(\mu(t), \sigma(t), \xi(t))$ from a non-stationary dataset where $\mu(t) = \mu_0 + \mu_1 t$. In this paper, all the datasets in chapter 4 are stationary so the NEVA is omitted. If one still wants to use EVA to analyze non-stationary datasets, then BM method is better than POT method because the blocks separate the covariate so minimize the impact from the covariate but the threshold can not (see the red path in Figure 6).

In step 2, BM method and POT method is applied on the dataset. In this process, the optimal block size and the optimal threshold value will be discussed. As the sample of extreme values is determined, the (μ, σ, ξ) is estimated by MLE in BM method and (σ, ξ) is estimated by Zhang Estimation in POT method.

In step 3, with the parameters calculated in last step, the diagnostic plot will be displayed. The diagnostic plot contains 3 plots which are probability density plot, quantile-quantile plot (Q-Q plot) [21] and probability–probability plot (P-P plot) [22] respectively. These 3 plots can confirm the good fit of the model which is GEV model in BM method and GPD in POT method if the line matches the histogram in probability density plot and in both Q-Q and P-P plot, R^2 value is close to 1. After making sure the good fit, the return period will be plotted and VaR will be calculated. Finally, we can make conclusions by analyzing these outputs.

4 Modelling

4.1 Application I: Daily Log Loss of S&P-500 Index

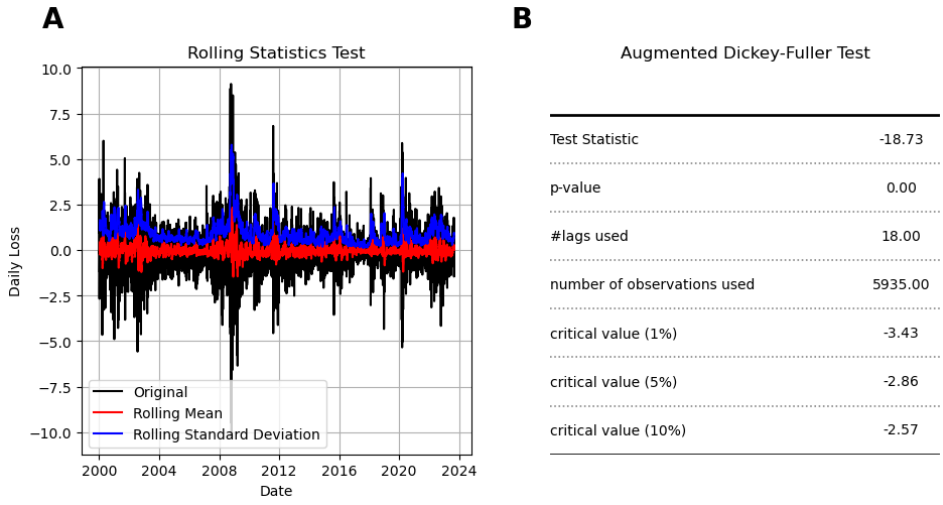
In this application, the dataset is the historical data of S&P-500 Index. We will calculate its daily log loss, denoted $DL(t)$, of 5954 joint transaction days from the first transaction day in 2000 till August 31, 2023 is extracted from yahoo finance using Python API. It should be noticed that each data is observed on transaction day, not every day. We define $DL(t)$ as

$$DL(t) = \begin{cases} [\ln(p(t)) - \ln(p(t-1))] \times 100 & \text{for } t > 1, \\ 0 & \text{for } t = 1, \end{cases}$$

where $p(t)$ is the close price and $p(t-1)$ is the open price on the same transaction day.

We do the stationarity test on $DL(t)$ (see Figure 7). The sub-figure A shows the constant rolling mean and constant rolling standard deviation. In sub-figure B, the test statistic is -18.73 which is way less than the critical values (-3.44 , -2.86 , -2.57) at the different levels. Hence, we can conclude $DL(t)$ is stationary.

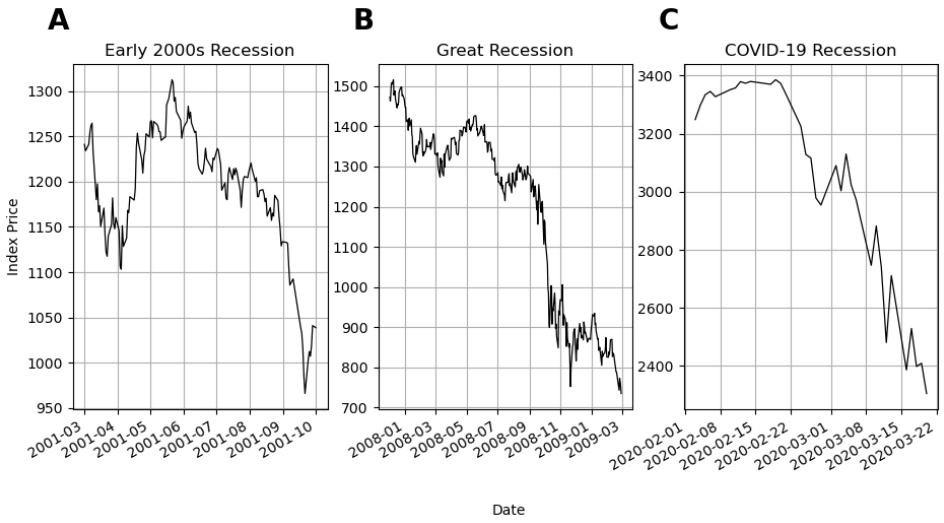
Figure 7: Stationary Test.



After confirming its stationarity, we can start the step 2. Before we discuss the BM and

POT method in depth, we shall look at the recorded economic Recessions [23] (see Figure 8).

Figure 8: Recessions.



Since 21st century, there were 3 recessions recorded on Wikipedia. Distinctly, the price of S&P-500 Index has the downgrade trend during the recessions on the whole. However, the fluctuation still happens very frequently in short periods. To make the good EVA of this stock index, we include monthly blocks instead of using only yearly blocks on BM method and we use the data during recessions to find the proper threshold on POT method.

BM Method

It should be noticed that the number of observed data within each block is less than its block size. For example, if the fixed block size is "1M", there might be only 20 observed data in each block because only 20 transaction days in a month. We look at the MLE of parameters in GEV models of different block sizes (see Table 1).

Table 1: MLE of Parameters.

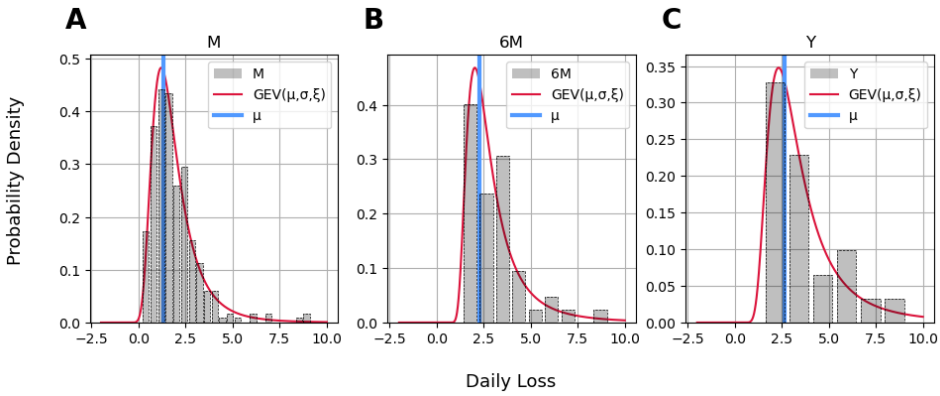
Block Size	μ	σ	ξ
1M	1.320	0.773	0.172
2M	1.710	0.794	0.208
3M	1.921	0.827	0.210
6M	2.250	0.825	0.326
1Y	2.620	1.107	0.315

As seen in Table 1, the larger block size leads to larger location parameter because as the block size becomes larger, the sample of block maximums is getting less in number and is more extreme in value. The scale parameter varies because of the different samples of

block maximums. The shape parameters also become larger. As block size becomes larger, the more extreme values in corresponding right tail weighs more in its probability density. Hence, this indicates a heavier tail as compared to the prior one (see Figure 9).

The shape parameters are all positive so by Figure 2, these samples of block maximums will approach to a Fréchet distribution which has a heavy tail. This is a valid finding because in most applications related to finance, the shape parameter of GEV distribution on BM method is always non-negative, which means there will be a tail in its distribution that implies the future returns or loss.

Figure 9: Probability Density Plot.



As seen in Figure 9, the red curve is the Fréchet PDF model based on the MLE in Table 1. We can see that as the red curve reaches the blue vertical line which is the value of location parameter, it decreases less sharply moving forward along the x-axis if its block size is bigger.

To find the optimal block size for the further analysis, we check the similarity between the observed sample and model-predicted sample of the block maximums by the mean of absolute difference (MoAD) (See Tabel 2).

The MoAD is calculated by the formula $\frac{1}{n}\sum_{i=1}^{n}|x_i - y_i|$ where n is the number of the sample and x_i, y_i is the data in the respective dataset. The smaller the MoAD value is, the better the fit of the model.

Table 2: MoAD.

Block Size	..	ξ	$MoAD$
1M	..	0.172	0.077
2M	..	0.208	0.108
3M	..	0.210	0.12
6M	..	0.326	0.189
1Y	..	0.315	0.415

As seen in Table 2, the optimal block size is "1M".

However, one may question the validation of such test because the shape parameter is close to 0, which raises a conjecture that the sample of block maximums should fit a Gumbel model ($\xi = 0$) instead of a Fréchet model. We recall that in (1.5), the PDF is different between $\xi = 0$ and $\xi \neq 0$. To deal with such concern, we introduce the hypothesis test here.

$$\textbf{Null hypothesis:} \quad \xi = 0,$$
$$\textbf{Alternative hypothesis:} \quad \xi \neq 0.$$

We calculate the likelihood ratio statistic (LRS) [24] of two competing models, which follows a chi-squared distribution [25] under the null hypothesis. The p-value is calculated by the formula $p = 1 - CDF_{X^2}(LRS)$ and the the chosen significance level α is 5%. To reject the null hypothesis, p-value should be less than α, then we say the shape parameter is not equal to 0 in this hypothesis test with a significance of 5% (Fréchet distribution in this example). Otherwise, we fail to reject the null hypothesis and so the shape parameter is accepted to be 0 (Gumbel distribution).

The new table includes parameters, MoAD and hypothesis test of respective Weibull model and Gumbel model (see Table 3).

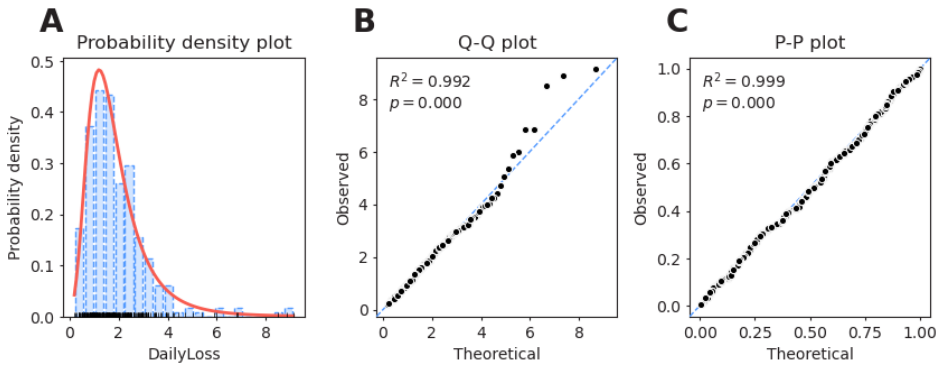
Table 3: MLE.

	Block Size	μ	σ	ξ	$MoAD$	Null Hypothesis (ξ=0)
Fréchet	**1M**	1.320	0.773	0.172	0.077	**Reject**
	2M	1.710	0.794	0.208	0.108	**Reject**
	3M	1.921	0.827	0.210	0.12	**Reject**
	6M	2.250	0.825	0.326	0.189	**Reject**
	1Y	2.620	1.107	0.315	0.415	**Accept**
Gumbel	**1M**	1.397	0.841	0	0.122	
	2M	1.807	0.884	0	0.193	
	3M	2.022	0.920	0	0.188	
	6M	2.410	0.990	0	0.28	
	1Y	2.826	1.310	0	0.432	

As seen in Table 3, the hypothesis test reject $\xi = 0$ if block size is "1M", "2M", "3M" and "6M" (red text) but accept $\xi = 0$ if the other. Although such hypothesis test seems not useful because the MoAD of Weibull model outperforms the MoAD of Gumbel model in each corresponding block size, it is meticulous to do such test in case of missing the better fit.

Based on the optimal block size "1M" and the corresponding parameters, we can make its diagnostic plot (see Figure 10).

Figure 10: Diagnostic Plot.



As seen in Figure 10, we conclude the sample of monthly maximums fits the GEV model very well. By using its parameters, we can calculate the r-period repeat interval and the Var & ES at respective 95% and 99% CI (see Table 4).

We recall 3 economic recessions since 21st century mentioned before, whose time since previous recession is 120 months, 73 months, and 128 months respectively. We only choose its mean which is 107 months as the period of repeat interval in case of the duplication.

Therefore, we assume the recession is expected to happen once every 107 months based on the record. Although such choice is disputable, defining an economic recession even at an early stage is valid because economic recession always takes a long time to build and emerge in the market.

The number of observations in each block is 21 so k in VaR and ES is 21.

Table 4: Quantiles.

r-107 repeat interval	VaR (95%)	VaR (99%)	ES (95%)	ES (99%)
6.856%	1.263%	2.699%	2.472%	3.99%

Given the quantiles in Table 4, the r-107 repeat interval indicates that S&P 500 stock index is expected to have a 6.856% maximum log loss as the recession happens in the next 107 months.

The VaR with 95% CI is 1.263% means there is a 5% probability that tomorrow's potential loss will exceed 1.263%. This implies that in most months (95% of the given time), we expect the losses to be less than or equal to 1.263% of the stock index's price within a day. However, in approximately 5% of months, we could experience losses exceeding 1.263% within a day.

The ES in the worst 5% of cases is 2.472% means if losses exceed the VaR (95%) which is 1.263% then the expected magnitude of those excess losses is 2.472%. This implies that in the rare months when losses exceed the VaR threshold 1.263% (approximately 5% of the time), we can expect the conditional loss in those months to be 2.472% of the stock index's price within a day.

These quantiles are reasonable based on our experience in the stock market, but we still need to check the output on POT method to make a conclusion.

POT Method

From example 3 in the section 2.3.2, we have an image of how the threshold affects the goodness of the fit. To have a suitable threshold, firstly, we have to define the extreme event during given recessions. We need to create a new dataset which includes only positive daily log loss data during recessions shown in Figure 8. After that, we calculate the mean and the median of this dataset, then we have the average of the mean and the median, which is 1.4% (see Table 5).

Table 5: Threshold.

Mean	Median	Threshold	Number of Exceedance (Original Dataset)
1.63%	1.16%	1.40%	445

The main reason to have 1.40% as our initial guess of threshold is that such value confirms all exceedances over 1.40% can be defined as the extreme events in this application. Then, the number of determined extreme events in the original dataset is 445, which is about 7.5% of the number of total data and this percentage also validates the reason mentioned before.

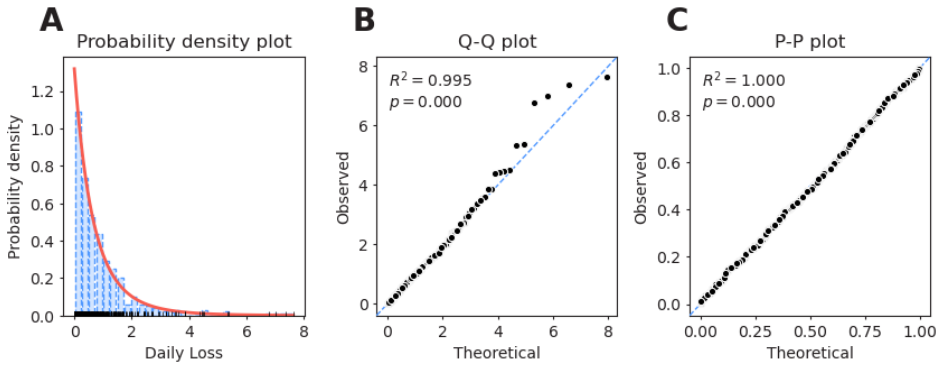
With this initial guess, we make few more possible thresholds (see Table 6). We omit the hypothesis test here because the Lomax model outperforms Exponential model at all thresholds. One can read this output in Appendix II.

Table 6: Zhang Estimation and MoAD.

Threshold (%)	σ	ξ	$MoAD$	Number of Exceedance
1.40	0.792	0.139	0.045	445
1.45	0.762	0.163	0.04	426
1.50	0.753	0.177	0.04	402
1.55	0.779	0.163	0.046	374
1.60	0.809	0.149	0.054	346

We stop at 1.6% because we do not want the number of exceedances gets too small and the MoAD also confirms not to try the threshold over 1.6%. From Table 6, we can see that they are very similar and should all have a good fit of the mode. Finally, we choose 1.5% as the threshold and we need to make the diagnostic plots (see Figure 11).

Figure 11: POT Diagnostic Plot.



As seen in Figure 11, the exceedances over 1.5% fits its GPD model very well. We use its parameters to calculate the quantiles (see Table 7).

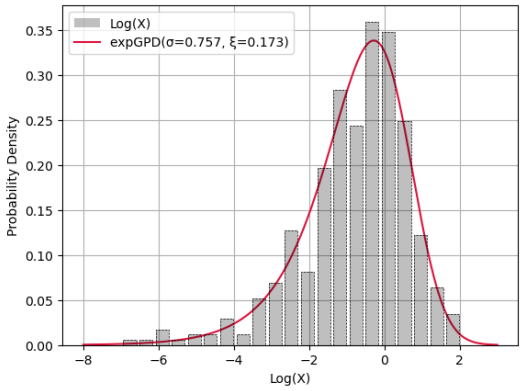
Table 7: Quantiles.

	r-107 repeat interval	VaR (95%)	VaR (99%)	ES (95%)	ES (99%)
BM	6.856%	1.263%	2.699%	2.472%	3.99%
POT	6.97%	1.732%	3.211%	2.697%	4.493%

As seen in Table 7, the quantiles in respective distribution models on the corresponding methods are very similar. This confirms the validation of our EVA on both methods of this stock index. Although many studies conclude the POT method outperforms the BM method while dealing with the applications related to finance, we suggest form the quantiles in a range based on the quantiles on both methods.

One may recall the exGPD mentioned in Figure 4 C, which is if $X \sim GPD(\mu = 0, \sigma, \xi)$, then $Y = \log(X) \sim \text{exGPD}(\sigma, \xi)$. Let us take a look at the exGPD PDF plot (see Figure 12) of the sample of block maximums with block size of "1M".

Figure 12: Exponentiated Generalized Pareto Distribution PD Plot.

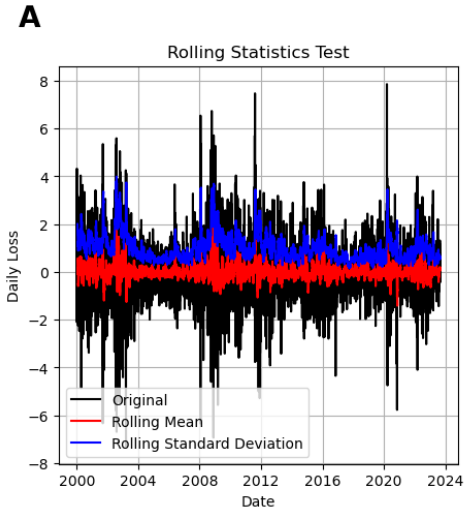


4.2 Application II: Daily Log Loss of CAC 40 Index

Since 21st century, the recessions always started in U.S. and spread around the world. The connection between global markets is explicit but the extent of damage in different markets is still worthy of being analyzed. In this application, with this idea, we do the EVA of CAC 40 Index.

We have made the complete EVA of application I, so in the following applications, we omit the unnecessary process. For the complete analysis, one shall read the code in Appendix II.

Figure 13: Stationary Test.



The stationarity test confirms the daily log loss data of CAC 40 is stationary. Compared to the Figure 7 A, the Figure 13 A shows the daily log loss of CAC 40 is more frequent in fluctuation over 2% and meanwhile, it is less extreme on the upper tail of its distribution. To validate this interpretation, we move on to the parameter estimation.

BM Method

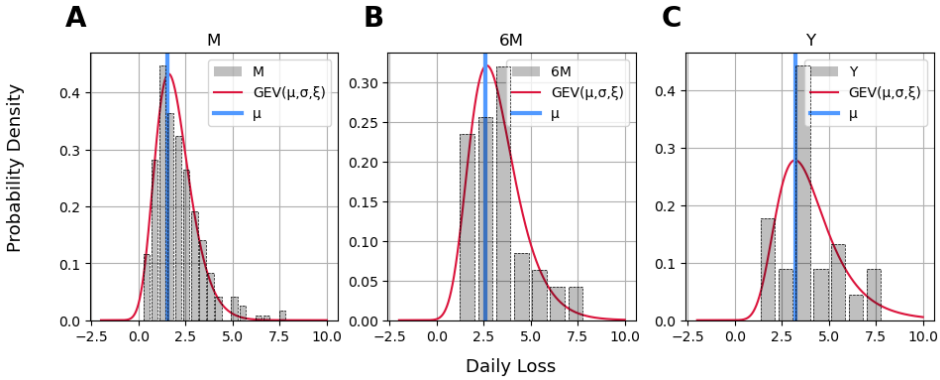
Table 8: MLE.

	Block Size	μ	σ	ξ	$MoAD$	**Null Hypothesis (ξ=0)**
Fréchet	**1M**	1.513	0.854	0.092	0.065	**Reject**
	2M	1.914	0.894	0.122	0.117	**Accept**
	3M	2.210	0.955	0.087	0.135	**Accept**
	6M	2.597	1.148	0.053	0.221	**Accept**
Weibull	**1Y**	3.241	1.322	-0.024	0.355	**Accept**
Gumbel	**1M**	1.557	0.887	0	0.085	
	2M	1.975	0.944	0	0.129	
	3M	2.257	0.988	0	0.145	
	6M	2.631	1.171	0	0.229	
	1Y	3.224	1.312	0	0.351	

As seen in Table 8, the shape parameters are more close to 0 as compared to which in Table 3. This discovery proves our interpretation before. We recall that in Figure 9 in Application I, the probability density of the daily log loss on the right side of its location parameter decreases slowly, which leads to greater positive shape parameters. In this application,

however, the more frequent daily loss over its location parameter and less extreme block maximums on the upper tail lead to smaller shape parameters and a bounded curve on the upper tail (see Figure 14).

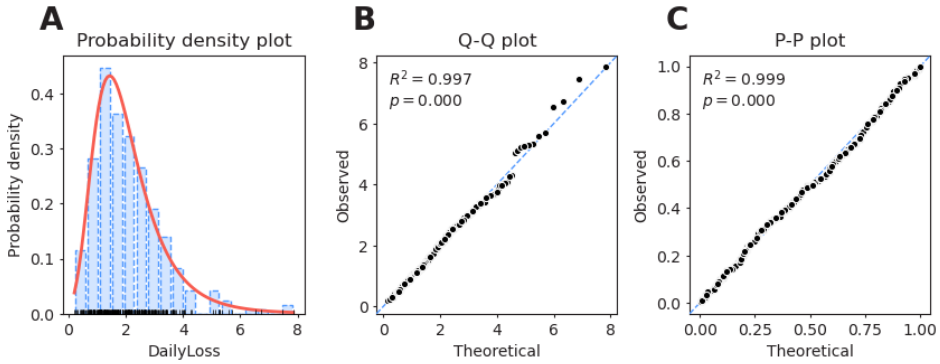
Figure 14: Probability Density Plot.



In Table 8, another interesting finding is the sample of "1Y" maximums approaches to a Weibull distribution but its Gumbel approximation shows a better fit hinged on the lower MoAD (see blue text in Table 8. So for the sample of "1Y" maximums, the Gumbel model is slightly better.

The optimal block size in this application is still "1M" and let us read its diagnostic plot.

Figure 15: Diagnostic Plot.



We continue on the EVA on POT method directly and we make conclusions then.

POT Method

Table 9: Threshold.

Mean	Median	Threshold	Number of Exceedance (Original Dataset)
1.413%	1.017%	1.215%	585

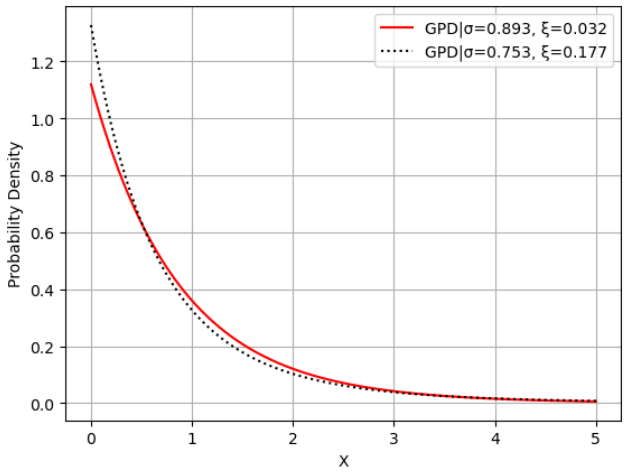
Table 10: Zhang Estimation and MoAD.

Threshold (%)	σ	ξ	$MoAD$	Number of Exceedance
1.215	0.921	0.017	0.031	585
1.30	0.942	0.006	0.04	522
1.40	0.944	0.004	0.038	476
1.50	0.893	0.032	0.031	438
1.60	0.883	0.039	0.033	396

As seen in Table 10, the shape parameters are more close to 0 and the scale parameters are larger compared to which in Table 6 in Application I. To make an interpretation, we look at two different PDFs of GPD both at the threshold 1.5% (see Figure 16).

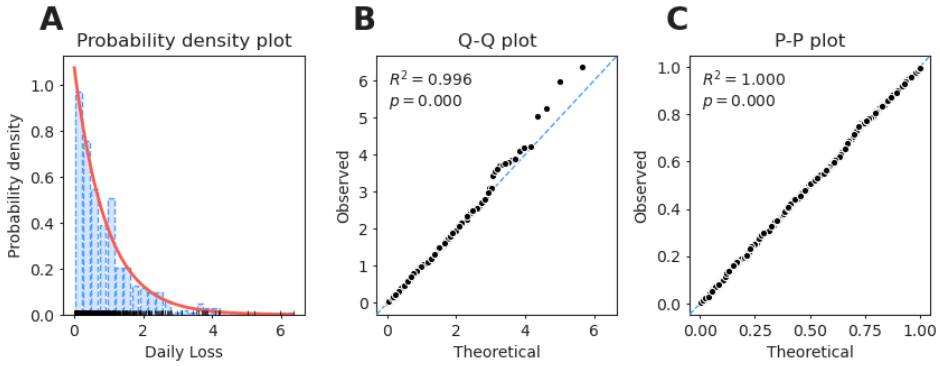
Clearly, the daily log loss of CAC 40 has a flatter curve but a less heavier tail compared to the daily log loss of S&P 500, which indicates the daily log loss of CAC 40 is more dense over the threshold 1.5% and less extreme in the most harsh situations (on the tail). This also confirms the interpretation we made on BM method.

Figure 16: Different Cases.



For both applications, we choose 1.5% as the optimal threshold. We make the diagnostic plot (see Figure 17).

Figure 17: POT Diagnostic Plot.



We conclude the threshold 1.5% is valid in this application. With all the information, we make the quantiles table (See Table 11).

Table 11: Quantiles.

	r-107 repeat interval	VaR (95%)	VaR (99%)	ES (95%)	ES (99%)
BM	6.491%	1.45%	2.941%	2.661%	4.087%
POT	6.003%	1.833%	3.326%	2.767%	4.309%

One can make similar interpretations of the quantiles according to the interpretations we discussed in Application I. We will make the conclusion on a whole based on the given information in chapter 5.

4.3 Application III: Daily Log Loss of FTSE 100 Index

We will do the EVA of this application very quickly. Our main purpose is to discuss the different attributes hidden in these 3 stock indexes. The stationary test confirms the daily log loss of FTSE 100 Index is stationary and shows the daily log loss of FTSE 100 Index has a more fluctuating and a more extreme pattern than S&P 500 does.

The optimal block size on BM method is still "1M" and the Fréchet outperforms Gumbel. The optimal threshold determined by MoAD is 1.274% but we still choose 1.5% as input in order to compare its quantiles with which in previous 2 applications. The parameters is in Table 12.

Table 12: Parameters on Two Methods.

		μ	σ	ξ
BM	**1M**	1.406	0.764	0.21
POT	**1.5%**	0	0.866	0.142

The diagnostic plot on respective method is confirmed to be good and we omit it here. We calculate the quantiles (see Table 13).

Table 13: Quantiles.

	r-107 repeat interval	VaR (95%)	VaR (99%)	ES (95%)	ES (99%)
BM	7.462%	1.35%	2.812%	2.606%	4.232%
POT	7.248%	1.751%	3.382%	2.803%	4.705%

5 Conclusion

To have a good understanding of the following conclusion, we suggest one should finish reading the materials above. We discussed a lot of interpretations and primary inferences in Chapter 4 and we omit those here in case of duplication. We will only make the conclusion on the given information in Table 14.

Although 3 stock indexes are different in many ways, the daily log loss data minimizes such influence from different total market capitalization and market rules. Meanwhile, even if we only focus on the most extreme events in EVA, one shall be able to find the similar pattern while dealing with other kinds of analysis such as quantities including std, variance and etc. These premises indicate the validation of our following conclusion as well.

Table 14: Quantiles.

		r-107	VaR (95%)	VaR (99%)	ES (95%)	ES (99%)
S&P 500	**BM**	6.856%	1.263%	2.699%	2.472%	3.99%
	POT	6.97%	1.732%	3.211%	2.697%	4.493%
CAC 40	**BM**	6.491%	1.45%	2.941%	2.661%	4.087%
	POT	6.003%	1.833%	3.326%	2.767%	4.309%
FTSE 100	**BM**	7.462%	1.35%	2.812%	2.606%	4.232%
	POT	7.248%	1.751%	3.382%	2.803%	4.705%

As seen in Table 14, CAC 40 has a smaller r-107 repeat level but larger VaR and ES at different CIs except the slightly lower ES at 99% CI. Lower repeat level indicates if recession happens in next 107 months, CAC 40 is expected to experience a lower log loss than S&P 500 does within a given time; larger VaR and ES indicates in the next day, CAC 40 is expected to experience a higher log loss than S&P 500 does at a given CI. However, only the ES of CAC 40 on POT method at 99% is lower than which of S&P 500 because the daily log loss of CAC 40 is less extreme in value on the most upper 1% tail. VaR at 99% CI is less affected by this because the shape parameter is divided once in the formula of VaR but $1 - \xi$ is divided in the formula of ES. These interpretations validate what we mainly

discussed in Application II, that is CAC 40 is more fluctuating but less extreme, especially over the threshold 1.5% or we say among the defined extreme events.

Following the above deduction, we conclude the daily log loss of CAC 40 is more fluctuating and less extreme than which of S&P 500 among the respective extreme values; the daily log loss of FTSE 100 is more fluctuating and more extreme than which of S&P 500 among the respective extreme values.

Given the information, we are unable to find a consistent and different attribute between CAC 40 and FTSE 100 on a whole but the daily log loss of FTSE 100 on the upper 1% tail is definitely much more extreme than which of CAC 40. This also shows that the daily log loss of S&P 500 always affects which of CAC 40 and which of FTSE 100 but the connection between CAC 40 and FTSE 100 is less visible.

6 Acknowledgements

I thank Professor Wael Bahsoun for being my supervisor and offering me this project opportunity, which was beneficial to reinforce my quantitative background. The significant insights and comments from Professor Wael led this to a much improved article. My work was supported by the Department of Mathematical Sciences while I was a postgraduate student at Loughborough University.

7 Appendix

7.1 Appendix I

To have (2.1), we consider

$$\begin{aligned}\Pr\{X > z+u \mid X > u\} &= \frac{\Pr(X > z+u, X > u)}{\Pr(X > u)},\\ &= \frac{\Pr(X > z+u)}{\Pr(X > u)},\\ &= \frac{1-F(z+u)}{1-F(u)}.\end{aligned}$$

From (1.1) and (1.2), we know that for $\xi \neq 0$,

$$\Pr\{M_n \leq z\} = (F(z))^n \approx G(z) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}.$$

Then, we have

$$n \ln F(z) \approx -\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]_+^{-1/\xi}.$$

By Taylor series expansion, for a large value of z, we know that

$$\ln F(z) \approx -\left[1-F(z)\right].$$

Plug in, we have

$$
\begin{aligned}
n(-[1-F(z)]) &\approx -\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]_+^{-1/\xi}, \\
\implies 1-F(u) &\approx \frac{1}{n}\left[1+\xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi} \quad \text{for large } u, \\
\overset{z>0}{\implies} 1-F(z+u) &\approx \frac{1}{n}\left[1+\xi\left(\frac{z+u-\mu}{\sigma}\right)\right]_+^{-1/\xi} \quad \text{for large } u.
\end{aligned}
$$

Substitution into above, then

$$
\begin{aligned}
\Pr\{x > z+u \mid x > u\} &= \frac{1-F(z+u)}{1-F(u)}, \\
&\approx \frac{\frac{1}{n}\left[1+\xi\left(\frac{z+u-\mu}{\sigma}\right)\right]_+^{-1/\xi}}{\frac{1}{n}\left[1+\xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi}}, \\
&= \left[\frac{1+\xi\left(\frac{u-\mu}{\sigma}\right)+\xi\frac{z}{\sigma}}{1+\xi\left(\frac{u-\mu}{\sigma}\right)}\right]_+^{-1/\xi}, \\
&= \left[\frac{1+\xi\left(\frac{u-\mu}{\sigma}\right)}{1+\xi\left(\frac{u-\mu}{\sigma}\right)}+\frac{\xi\frac{z}{\sigma}}{1+\xi\left(\frac{u-\mu}{\sigma}\right)}\right]_+^{-1/\xi}, \\
&= \left[1+\frac{\xi z}{\sigma+\xi(u-\mu)}\right]_+^{-1/\xi}, \\
&= \left[1+\frac{\xi z}{\tilde{\sigma}}\right]_+^{-1/\xi} \quad \text{where } \tilde{\sigma} = \sigma+\xi(u-\mu).
\end{aligned}
$$

Ideally, in a true GPD, $u = \mu \implies \tilde{\sigma} = \sigma$. Finally, we have

$$
\Pr\{X - u \le z \mid X > u\} = 1 - \left[1+\frac{\xi z}{\sigma}\right]_+^{-1/\xi}.
$$

7.2 Appendix II

The Python code is saved in a .ipynb file and the link is followed.

`https://github.com/DreenL/EVA-/blob/main/EVA.ipynb`

8 Reference

[1] Wikipedia contributors. 100-year flood — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=100-year_flood&oldid=1162077242, 2023. [Online; accessed 14-August-2023].

[2] R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2):180–190, 1928.

[3] Weisstein, Eric W. "extreme value distribution." from mathworld–a wolfram web resource. https://mathworld.wolfram.com/ExtremeValueDistribution.html, 2023.

[4] Wikipedia contributors. Maximum likelihood estimation — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Maximum_likelihood_estimation&oldid=1169528282, 2023. [Online; accessed 14-August-2023].

[5] Wikipedia contributors. Return period — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Return_period&oldid=1174340497, 2023. [Online; accessed 8-September-2023].

[6] Wikipedia contributors. Value at risk — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Value_at_risk&oldid=1164229035, 2023. [Online; accessed 29-August-2023].

[7] Wikipedia contributors. Expected shortfall — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Expected_shortfall&oldid=1166991371, 2023. [Online; accessed 8-September-2023].

[8] Klara Andersson. A comparative study of var and es using extreme value theory. 2020.

[9] Wikipedia contributors. Generalized pareto distribution — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Generalized_Pareto_distribution&oldid=1162971563, 2023. [Online; accessed 14-August-2023].

[10] Wikipedia contributors. Pickands–balkema–de haan theorem — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Pickands%E2%80%93Balkema%E2%80%93De_Haan_theorem&oldid=1169982116, 2023. [Online; accessed 14-August-2023].

[11] Seyoon Lee and Joseph HT Kim. Exponentiated generalized pareto distribution: Properties and applications towards extreme value theory. *Communications in Statistics-Theory and Methods*, 48(8):2014–2038, 2019.

[12] Wikipedia contributors. Akaike information criterion — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Akaike_information_criterion&oldid=1169352239, 2023. [Online; accessed 14-August-2023].

[13] Scott D Grimshaw. Computing maximum likelihood estimates for the generalized pareto distribution. *Technometrics*, 35(2):185–191, 1993.

[14] Jin Zhang. Improving on estimation for the generalized pareto distribution. *Technometrics*, 52(3):335–339, 2010.

[15] Minh H Pham, Chris Tsokos, and Bong-Jin Choi. Maximum likelihood estimation for the generalized pareto distribution and goodness-of-fit test with censored data. *Journal of Modern Applied Statistical Methods*, 17(2):11, 2019.

[16] Kevin Dowd. *Measuring market risk*. John Wiley & Sons, 2007.

[17] Wikipedia contributors. Stationary process — Wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=Stationary_process&oldid=1166251052`, 2023. [Online; accessed 29-August-2023].

[18] Wikipedia contributors. Moving average — Wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=Moving_average&oldid=1171747945`, 2023. [Online; accessed 29-August-2023].

[19] Wikipedia contributors. Augmented dickey–fuller test — Wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=Augmented_Dickey%E2%80%93Fuller_test&oldid=1124807741`, 2022. [Online; accessed 29-August-2023].

[20] Elisa Ragno, Amir AghaKouchak, Linyin Cheng, and Mojtaba Sadegh. A generalized framework for process-informed nonstationary extreme value analysis. *Advances in Water Resources*, 130:270–282, 2019.

[21] Wikipedia contributors. Q–q plot — Wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=Q%E2%80%93Q_plot&oldid=1151403653`, 2023. [Online; accessed 29-August-2023].

[22] Wikipedia contributors. P–p plot — Wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=P%E2%80%93P_plot&oldid=1170611246`, 2023. [Online; accessed 29-August-2023].

[23] Wikipedia contributors. List of recessions in the united states — Wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=List_of_recessions_in_the_United_States&oldid=1170452884`, 2023. [Online; accessed 19-August-2023].

[24] Wikipedia contributors. Likelihood-ratio test — Wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=Likelihood-ratio_test&oldid=1151611188`, 2023. [Online; accessed 1-September-2023].

[25] Wikipedia contributors. Chi-squared distribution — Wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=Chi-squared_distribution&oldid=1172005223`, 2023. [Online; accessed 1-September-2023].