ORIE 4656 Final Project - Jasmine Samadi, Emily Zhang

Description:

For this report we chose historical data (from April 23rd 2019 to April 23rd 2024) from Novo Nordisk (the company that makes Ozempic) and NVIDIA. Our data came from Yahoo Finance.

Calculating returns:

```r
nvda <- read.csv("NVDA.csv")
novo <- read.csv("NOVO-B.CO.csv")
returns <- function(data) {
  return (tail(data, -1) - head(data, -1))/ head(data, -1) * 100
}
nvda_returns <- returns(nvda["Close"])
novo_returns <- returns(novo["Close"])
```

First we uploaded the CSV files and calculated the daily returns based on the closing daily closing prices.

- The net or simple return at time t is

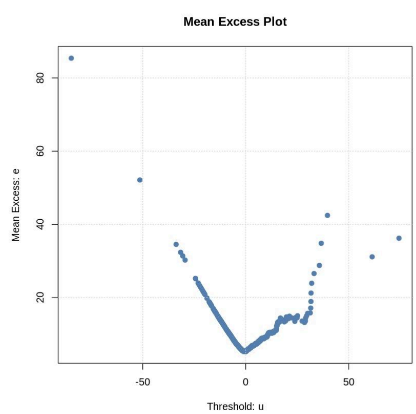
$$R_t := \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1.$$

We used this formula on the left hand side to calculate the daily returns. tail(data, -1) was our daily return at time t because it returns everything other than the first element. We need the vectors to be comparable in length in order to calculate the returns. Therefore, we take out the oldest value and start at the i+1 index. head(data, -1) is our Pt-1 because it returns all elements except for the last one, or the most recent one.

Mean Excess Plot:

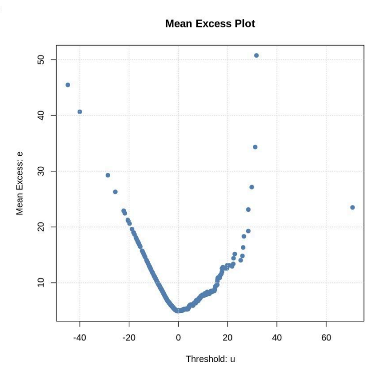
Mean Excess Plot NVIDIA returns

```r
mePlot(as.numeric(nvda_returns$Close), doplot = TRUE, labels = TRUE)
```



Mean Excess Plot Novo Nordisk Returns

```r
mePlot(as.numeric(novo_returns$Close), doplot = TRUE, labels = TRUE)
```



The empirical mean excess plot can help us detect tail behavior. Based on the documentation of the mePlot function, if the plot is a straight line, that indicates that GPD is a good fit. If the plot is flat, that typically means that the data is exponential. If the plot is curved it means that the data might follow a Weibull or Gamma distribution. Based on (Ghosh and Resnick 2010) the data supports a GPD model if the mean excess plot has a linear slope. Furthermore, if a GPD model is a good fit for the data, a positive slope indicates a positive shape parameter, a negative slope indicates a negative shape parameter and a zero slope indicates a zero slope parameter.

Both of the mean excess plots for NVIDIA and Novo Nordisk returns have a V shape. The negative thresholds of the plots show a decreasing straight line up to the threshold 0. For values after the threshold, the mean excess function increases but there is no clear trend. Because the plots are not consistently linear, we can say that GPD is not a good fit for the NVIDIA or Novo Nordisk returns.

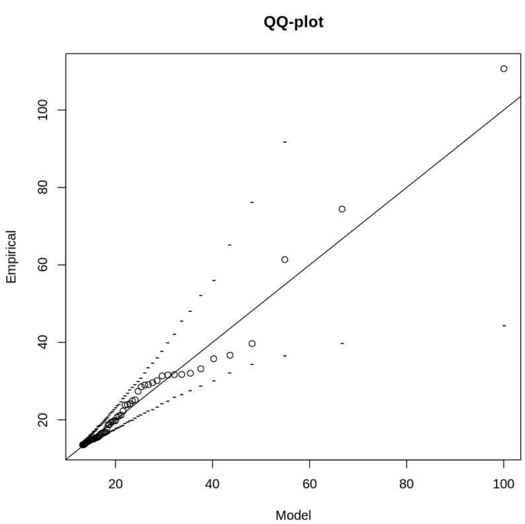
We can further analyze this by fitting GPD to the data and creating QQ plots.

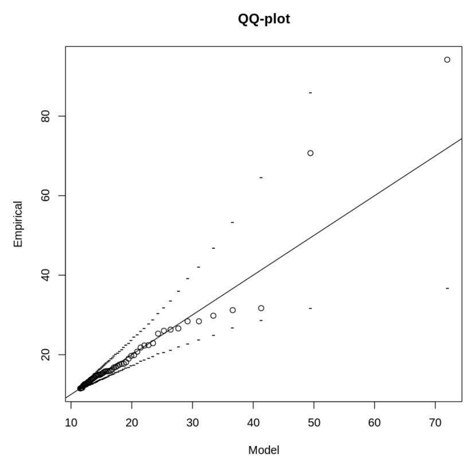
```r
threshold_nvda <- quantile(nvda_returns$Close, 0.95)

# Fit the GPD above the threshold
fit_GPD_nvda <- fitgpd(nvda_returns$Close, threshold_nvda)
plot(fit_GPD_nvda)
```

```r
threshold_novo <- quantile(novo_returns$Close, 0.95)

# Fit the GPD above the threshold
fit_GPD_novo <- fitgpd(novo_returns$Close, threshold_novo)
plot(fit_GPD_novo)
```





Here, we look at the 95% quantile because we are looking at 5% of data for extreme values. This method will use the mean likelihood estimator (MLE) for our data to a GPD model. Once we plot the QQ plots we can see that there are similar results for the NVIDIA and Novo Nordisk returns. Since the QQ plots do not follow a straight line and more so look exponential, we can see that GPD is not a good fit for our data. This validates the results we have from interpreting the mean excess plots. Therefore, we can conclude that both the NVIDIA and Novo Nordisk returns are not a good fit for the General Pareto Distribution.

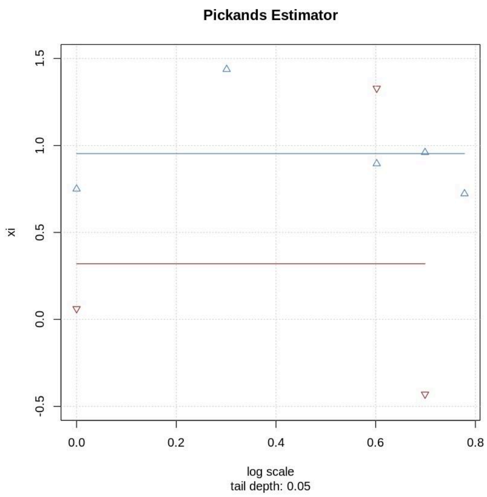
3 Ways to Estimate Shape Parameter:

Pickands Estimator with NVIDIA returns:

Upper and Lower Tails Method 1:

```
shaparmPickands(as.numeric(nvda_returns$Close), p = 0.05, xiRange = NULL,
        doplot = TRUE, plottype = c("both"), labels = TRUE)
```

$xi =
0.953078980183663 · 0.288316532779257 · 0.31996413746016 · 0.90736058693029

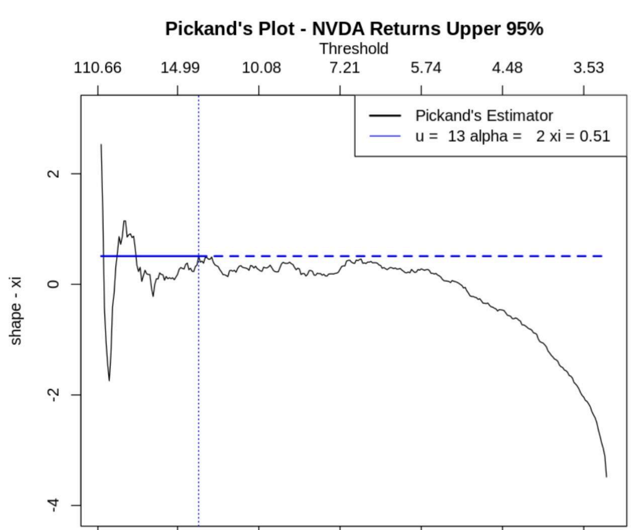


The first way we analyzed Pickands estimator for the shape was to use sharpmPickands. We set p = 0.05 because we are interested in values that happen 5% of the time or less. We plot both the upper and lower tails. This tells us that the upper tail has an estimated shape parameter of <u>0.953078980183663</u> with a standard error of <u>0.288316532779257</u>. The lower tail has an estimated shape parameter of<u> 0.31996413746016</u> with a standard error of <u>0.90736058693029</u>.

In class we also used another function for the Pickands estimator: pickandsplot. In order to see how accurate our shape parameters are we decided to conduct both methods and compare their results.

Upper Tail Method 2:

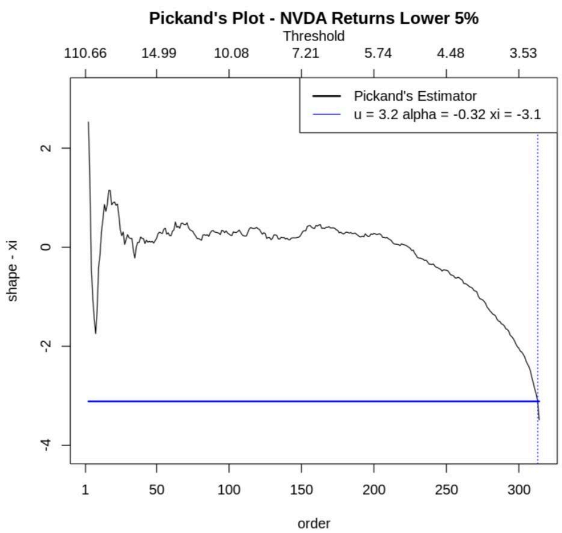
```
pickandsplot(nvda_returns$Close,
             orderlim = NULL,  #sets vector of lower/upper limits of order statistics to default values
             tlim = NULL,      #sets vector of lower/upper limits of range of threshold to default values
             y.alpha = FALSE,  #setting it to false lets us plot the shape values
             alpha = NULL,     #set to null because we are not plotting the CI
             ylim = NULL,      #lets plot come up with automatic range of y values
             legend.loc = "topright",  #sets location of legend
             try.thresh = quantile(nvda_returns$Close, 0.95, na.rm = TRUE),   #creates vector of thresholds
             main = "Pickand's Plot - NVDA Returns Upper 95%",  #title of plot
             xlab = "order", #x-axis title
             ylab = "shape - xi")  #y-axis title
```



By using the function pickandsplot, we set y.alpha = FALSE so that it plots shape values. We also set try.thresh = quantile(nvda_returns$Close, 0.95, na.rm =

TRUE) so that we look at the top 95% of the data. When plotted we can see that there is an estimated shape of xi = 0.51.

Lower Tail Method 2:




```
pickandsplot(nvda_returns$Close,
             orderlim = NULL,  #sets vector of lower/upper limits of order statistics to
             tlim = NULL,      #sets vector of lower/upper limits of range of threshold
             y.alpha = FALSE,  #setting it to false lets us plot the shape values
             alpha = NULL,     #set to null because we are not plotting the CI
             ylim = NULL,      #lets plot come up with automatic range of y values
             legend.loc = "topright",  #sets location of legend
             try.thresh = quantile(nvda_returns$Close, 0.5, na.rm = TRUE),   #creates vec
             main = "Pickand's Plot - NVDA Returns Lower 5%",  #title of plot
             xlab = "order", #x-axis title
             ylab = "shape - xi")  #y-axis title
```

For the lower tail we conduct the same process as the upper tail but for the threshold instead of using 0.95 we use 0.5 because we want to look at the bottom 5% of the data. Here we get an estimated shape parameter xi = -3.1.

Comparison:

To summarize, the first method gave us an upper tail xi = 0.953078980183663 and the second method gave us an upper tail xi = 0.51. These two methods have vastly different outputs even though they are both using Pickands estimator. The standard error for the first method is 0.90736058693029, showing us that the estimate is not as accurate as it could be. The second method does not have a concrete way of finding the standard error so we do not know if 0.51 is a better estimator or not. What we can conclude from these results is that the shape parameter is probably positive, which means that the shape parameter of the upper tails for the NVIDIA returns is heavy tailed and can be modeled with a Frechet distribution.

The first method gives us a shape parameter of 0.31996413746016 for the lower tail with an standard error of 0.90736058693029. The second method gives us a shape parameter of -3.1. This method does not give us an accurate way of calculating standard error for the whole estimator, so we do not have that data point to compare the two methodologies. We can however discredit the negative shape parameter. A negative shape parameter indicates a bounded Weibull distribution on the extremes. Having a bounded tail means that it is not likely for extreme events to occur. We know from other methods like Hills estimator, Moment estimator, and the extremal index that extreme events do have a probability of occurring. Since the extremal index for the NVIDIA returns is 0.39, we know that there is a moderate probability of extremal events occurring. In addition, since these are two ways of calculating the Pickands estimator and we are getting very different results, it is safe to say that Pickands does not estimate the shape parameter of this data well. We should use other estimators that yield a smaller standard error.

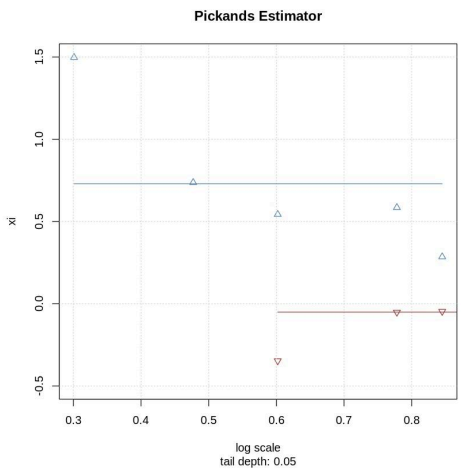
Pickands Estimator with Novo Nordisk returns:

Upper and Lower Tails Method 1:

```
shaparmPickands(as.numeric(novo_returns$Close), p = 0.05, xiRange = NULL,
        doplot = TRUE, plottype = c("both"), labels = TRUE)
```

$xi =

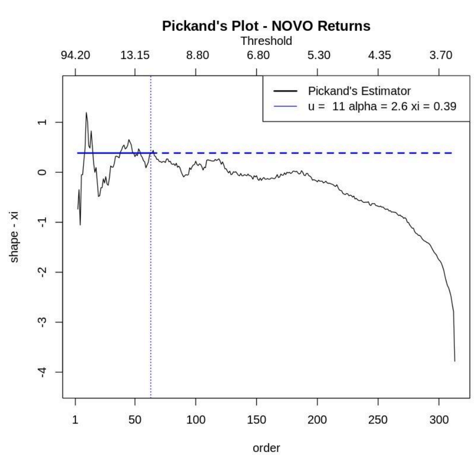
0.729586714463351 · 0.459011860328266 · -0.0509603102132268 · 0.239984399126291



Here we used the shaparmPickands function again to calculate Pickands estimator for Novo Nordisk using the same logic and reasoning as the NVIDIA returns. This gives us an upper tail shape parameter of <u>0.729586714463351</u> with a standard error of <u>0.459011860328266</u>. The lower tail outputs a shape parameter of <u>-0.0509603102132268</u> with a standard error of <u>0.239984399126291</u>.

Upper and Lower Tails Method 2

```
pickandsplot(novo_returns$Close,
             orderlim = NULL,  #sets vector of lower/upper limits of order statist
             tlim = NULL,      #sets vector of lower/upper limits of range of thre
             y.alpha = FALSE,  #setting it to false lets us plot the shape values
             alpha = NULL,     #set to null because we are not plotting the CI
             ylim = NULL,      #lets plot come up with automatic range of y values
             legend.loc = "topright",  #sets location of legend
             try.thresh = quantile(novo_returns$Close, 0.95, na.rm = TRUE),   #cre
             main = "Pickand's Plot - NOVO Returns",  #title of plot
             xlab = "order", #x-axis title
             ylab = "shape - xi")  #y-axis title
```



Here we used the same approach and reasoning as the second method in the NVIDIA section. This methodology gave us a shape parameter for the upper tail of <u>0.39</u>. Once again, a standard error for this approach was not able to be reasonably done. The lower tail shape parameter was not able to be calculated with this method either. Because the threshold for the lower tail was negative using quantile(novo_returns$close, 0.05, na.ram=TRUE), the try.thresh parameter in the pickandsplot function would not calculate. The try.thresh parameter is not allowed to be negative. Therefore, we did not have anything to compare the shape parameter from the first method of Pickands estimator to.

Comparison:

Given that both methods used Pickands estimator and have very different results, Pickands would not be the best way to estimate the shape parameter. The standard error for the first method was about 0.45 and the range of the estimated shapes was around 0.339. We can conclude that other estimation approaches should be used to better estimate the shape. Since both estimates are positive for the upper tail, we can likely say that the data is heavy tailed and the shape parameter follows a Frechet distribution.

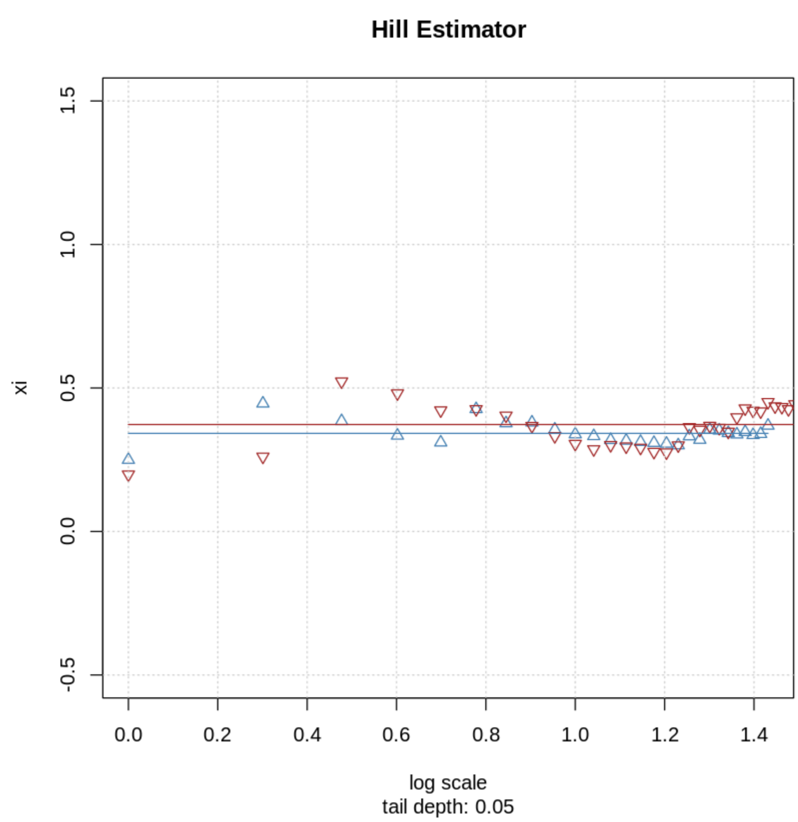
Hills Estimator

Novo Nordisk:

```
shaparmHill(as.numeric(novo_returns$Close), p = 0.05, xiRange = NULL,
    doplot = TRUE, plottype = c("both"), labels = TRUE)
```

$xi =

0.341777829325254
0.0393304220575824
0.372676008748336
0.0756241751080773

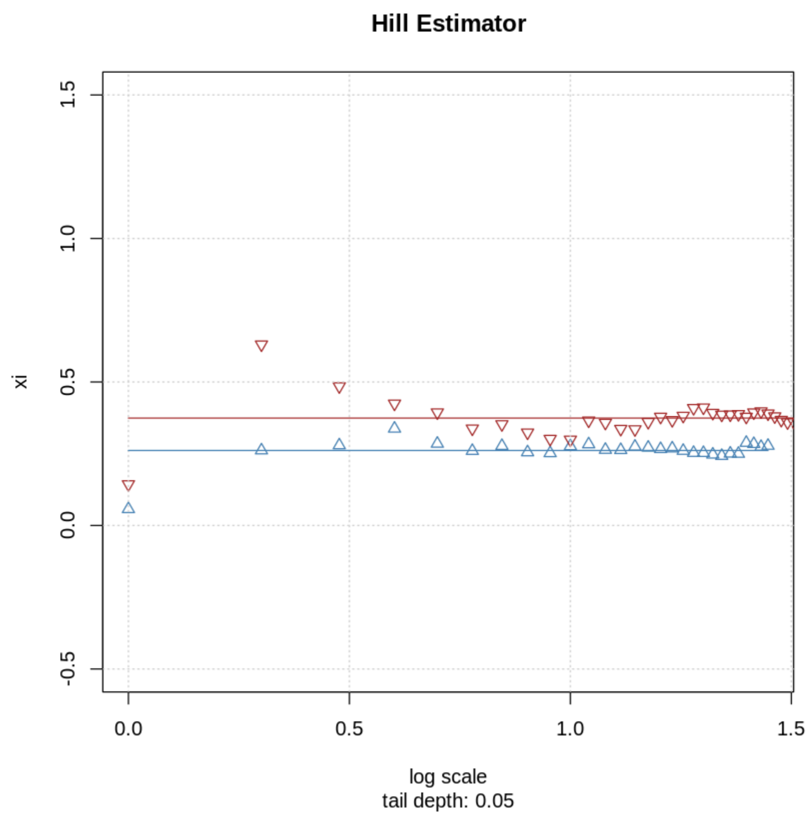


Nvidia:

```r
shaparmHill(as.numeric(nvda_returns$Close), p = 0.05, xiRange = NULL,
    doplot = TRUE, plottype = c("both"), labels = TRUE)
```

$xi =

0.261193454313381
0.0440571018736759
0.374172334977069
0.0713874852891768

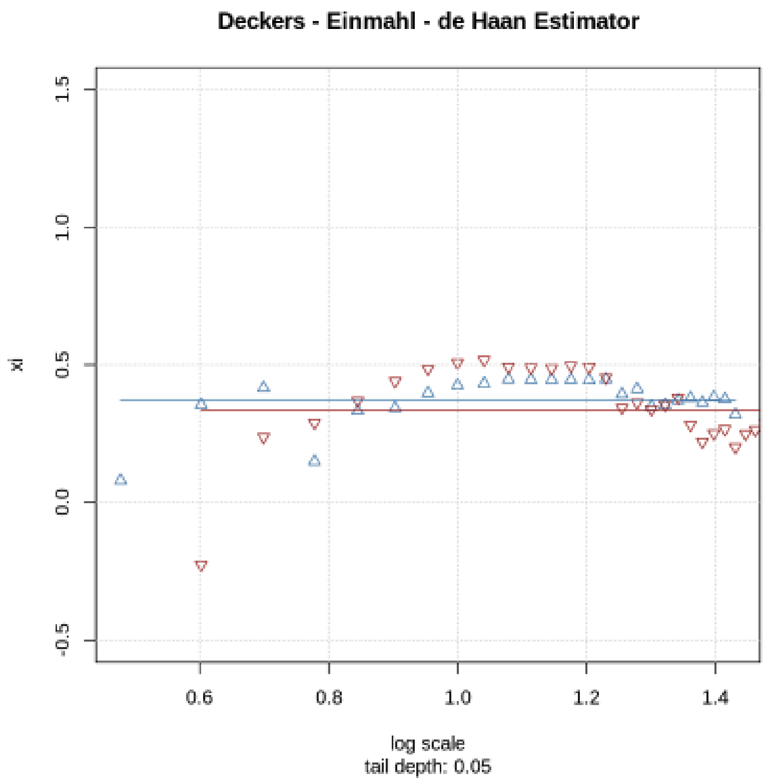


Utilizing the same package used for Pickands with equivalent parameters, we computed the Hill estimator for both Nvidia and Novo Nordisk returns. Evidently, our estimates for lower and upper tail are quite close, 0.261193454313381, 0.374172334977069 respectively for Nvidia, and 0.341777829325254, 0.372676008748336 for Novo Nordisk. For both distributions, the standard error of the estimator's is less than 0.1, which suggests that the hill estimator works quite well to estimate the shape. As the estimators are positive for both, we can conclude that the Frechet distribution models the extremes for both datasets.

Moment Estimator

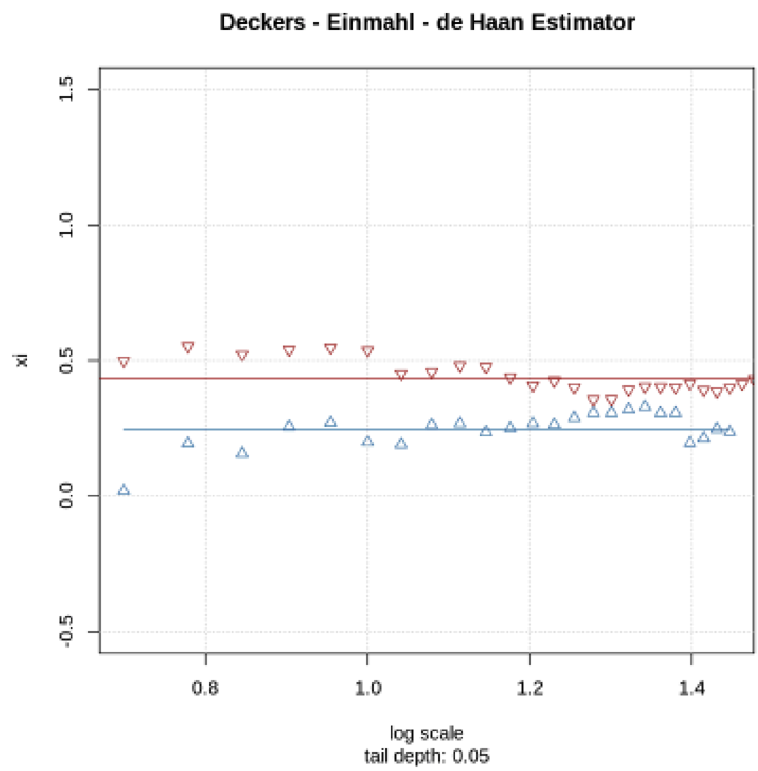
```r
shaparmDEHaan(as.numeric(nvda_returns$Close), p = 0.05, xiRange = NULL,
        doplot = TRUE, plottype = c("both"), labels = TRUE)
```

$xi =
0.372522886502923 · 0.0880195747862163 · 0.336559633279503 · 0.149686343363446



```r
shaparmDEHaan(as.numeric(novo_returns$Close), p = 0.05, xiRange = NULL,
        doplot = TRUE, plottype = c("both"), labels = TRUE)
```

$xi =
0.245520694566929 · 0.0665123683564489 · 0.432409259719913 · 0.07507461465771



Similar to the Hill estimator, the resulting moment estimators happen to relatively agree with each other for upper and lower tail shapes. For Nvidia, the estimators for lower and upper tails respectively are 0.27 and 0.34. For Novo Nordisk, the respective lower and upper tail estimators are 0.25 and 0.43. The associated standard errors are also miniscule, similar to Hill's estimator in that they are also less than 0.1. Thus, the moment estimator's for the shape parameters also agree with Hill's in that the shape parameter is positive, indicating a Frechet distribution for both distributions again.

Conclusions about shape parameters:

For the NVIDIA returns, the Hills estimator gives an upper tail shape with a standard error of: 0.0393304220575824. This is similar to the Moment estimator's standard error of 0.0880195747862163. The Pickands estimator is clearly not as accurate, as its standard error is 0.288316532779257. Hill's estimator outputs a shape parameter of 0.341777829325254, and Moments estimator outputs 0.372522886502923. Therefore, we can conclude that for the upper tail of NVIDIA returns, the data is heavy tailed and fits a Frechet distribution well. The shape parameter is likely to be close to 0.34 and 0.37.

For the NVIDIA lower tail, Hills, Moment, and Pickands have standard errors 0.0756241751080773, 0.149686343363446, and 0.90736058693029. They have estimated shape parameters of 0.372676008748336, 0.336559633279503, and 0.31996413746016. In method 2 for Pickands estimator, it outputs an estimated shape parameter of -3.1. Based on other estimators, we can say that this is not an accurate representation of the shape. Therefore the data is heavy tailed and fits a Frechet distribution well. Even though the standard error for the Moment Estimator and Pickands Estimator is much larger than Hills, all of the shape parameters are relatively similar. The standard error for Hill's estimator is the smallest, so we can conclude that the shape parameter of the lower NVIDIA returns tails is close to 0.37.

Now we will analyze the Novo Nordisk returns, starting with the upper tail shape parameter estimates. The Hills, Moment, and Pickands estimators have standard errors 0.0440571018736759, 0.0665123683564489, 0.459011860328266. The estimated shapes are: 0.261193454313381, 0.245520694566929, 0.72958671446335. Like the previous results, Hill's estimator has the smallest standard error with the Moment Estimators standard error being very close but slightly larger. Pickands estimator has a very large standard error. Pickands estimator also has a much larger estimated shape parameter. Therefore, we can conclude that the estimated shape parameter for the upper tail of Novo Nordisk returns is close to 0.26. This means that the data fits a Frechet distribution well and is heavy tailed.

Lastly, we will analyze the lower tail of the Novo Nordisk returns. The standard errors for Hills, Moment, and Pickands estimators are: 0.0713874852891768, 0.07507461465771, 0.239984399126291. The estimated shape parameters are: 0.374172334977069, 0.432409259719913, -0.0509603102132268. The standard error for the Hills and Moment estimators are the same up to the 2nd decimal point. The standard error for the Pickands estimator is much larger. Furthermore, the shape estimator for Pickands is a negative number, which would mean that the data would best be fit for a Weibull distribution. We can discredit this as the standard errors of the Hills and Moment estimators are much smaller. We can conclude that the data does indeed follow a Frechet distribution and the estimated shape parameter is closer to 0.37.

Extremal Index

```
threshold_nvda <- quantile(nvda_returns$Close, 0.95)
threshold_novo <- quantile(novo_returns$Close, 0.95)

# Calculate extremal indices
extremal_index_nvda <- extremalindex(nvda_returns$Close, threshold_nvda)
extremal_index_novo <- extremalindex(novo_returns$Close, threshold_novo)

# Display results
extremal_index_nvda
extremal_index_novo
```

In order to calculate the extremal index, we first created thresholds to get only extreme values. We used a threshold of 0.95 so that we would be analyzing the 95th percentile of returns for both NVIDIA and Novo Nordisk. Then, we used the extremalindex() function with our set thresholds to calculate the extremal index.

<u>Output NVIDIA:</u>

```
 Intervals Method Estimator for the Extremal Index
NULL

 theta.tilde used because there exist inter-exceedance times > 2.
   extremal.index number.of.clusters         run.length
        0.3940195         22.0000000          6.0000000
```

The extremal index for NVIDIA returns is equal to 0.3940195. There were 22 clusters of extreme values.

<u>Output Novo Nordisk:</u>

```
 Intervals Method Estimator for the Extremal Index
NULL

 theta.tilde used because there exist inter-exceedance times > 2.
   extremal.index number.of.clusters         run.length
        0.4026378         25.0000000          9.0000000
```

The extremal index for Novo Nordisk returns is equal to 0.4026378. There were 25 clusters of extreme values.

<u>Conclusion:</u>

The extremal index tells us the degree of which extremes cluster. An extremal index closer to 0 indicates that extreme values are more likely to come in clusters. An extremal index closer to 1 indicates that the extremes are more likely to be independent. For both the NVIDIA and Novo Nordisk returns the extremal indexes are close to 0.4 which means that about 40% of the extreme data presents itself in clusters. Both stocks also experience clusters of extremes at a similar frequency as the number of clusters is comparable.

Citations:

Ghosh, Souvik, and Sidney Resnick. "A discussion on mean excess plots." *Stochastic Processes and Their Applications*, vol. 120, no. 8, Aug. 2010, pp. 1492–1517, https://doi.org/10.1016/j.spa.2010.04.002.